

UNIVERSIDADE DE LISBOA
FACULDADE DE PSICOLOGIA



Deteção de simulação: Validação do *Coin in Hand-Extended Version* com adultos idosos

Inês Margarida Bernardes Ferreira

MESTRADO INTEGRADO EM PSICOLOGIA

Área de Especialização em Psicologia Clínica e da Saúde

Psicologia da Saúde e da Doença

2019

UNIVERSIDADE DE LISBOA
FACULDADE DE PSICOLOGIA



Deteção de Simulação: Validação do *Coin in Hand-Extended Version* com adultos idosos

Inês Margarida Bernardes Ferreira

Dissertação orientada pela Professora Doutora Sandra Fernandes

MESTRADO INTEGRADO EM PSICOLOGIA

Área de Especialização em Psicologia Clínica e da Saúde

Psicologia da Saúde e da Doença

2019

Without validity test results, providers may reach incorrect conclusions regarding the diagnosis and severity of a patient's difficulties or symptoms, and may subsequently initiate a number of inappropriate referrals, treatments, and/or accommodations, increasing iatrogenic effects for the patient, wasting money, and limiting resources for other patients.

(Lippa, 2018)

For effort test interpretation to be meaningful with a potentially demented patient, data are needed regarding which effort tests provide the lowest rate of false positive error, the relationship between severity of dementia and false positive rates, and the extent to which effort cut-offs need to be adjusted in dementia groups.

(Dean et al., 2009)

Agradecimentos

Nos próximos parágrafos agradeço a todas as pessoas que tornaram este percurso numa experiência bonita, significativa e *nossa*. Assim, devo um agradecimento especial,

À Professora Doutora Sandra Fernandes, por ter aceitado fazer esta viagem comigo, pela orientação e disponibilidades constantes, por me fornecer discernimento, segurança e confiança nos tempos mais atribulados, pelo apoio incansável e inigualável, pelo exemplo extraordinário de profissional que é. Obrigada por tudo. Esta viagem não teria feito sentido se não fosse consigo.

Ao Doutor Luís Querido, pelo auxílio na análise estatística dos resultados e, em especial, por ter sido uma fonte de tranquilidade, objetividade e segurança quando o *stress* teria sido o caminho mais fácil. Por me ajudar a fazer frente aos obstáculos e por me ter ensinado a olhar para os resultados com uma curiosidade insaciável por todos os detalhes.

À Dra. Julia Daugherty e ao Professor Doutor Miguel Pérez, pelo auxílio na análise estatística dos resultados e pelo entusiasmo demonstrado por este projeto.

Aos meus pais, Elsa Bernardes e Paulo Ferreira, por terem sido o pilar central deste caminho, pelos sacrifícios que, ao longo destes anos, fizeram, para que hoje este pudesse ser um desfecho realizador para os três. Pelo amor infinito que me dão e me faz querer ser cada vez melhor por vocês. Pelo apoio que me deram, desde sempre, para seguir aquilo que me fizesse feliz. Por acreditarem em mim quando eu não consegui e por me darem forças quando as minhas se esgotaram. Quando as dificuldades se enfatizaram, foram vocês que me fizeram acreditar e lutar, porque vocês mereciam isso. Obrigada a ambos! Começa aqui a minha recompensa por todos os sacrifícios que fizeram até hoje por mim.

Ao Daniel, pelo amor, a segurança e a felicidade que me dá. Pelo exemplo de perseverança e determinação que é. Por me mostrar que é possível atingirmos os objetivos a que nos propomos e que temos, dentro de nós e nos outros, forças que nem conhecíamos.

Obrigada por acreditares em mim até ao último segundo, em todas as circunstâncias e desafios.

Obrigada pelo porto seguro que és e pelo amor que me dás.

À Ana, à Mariana e à Marta, por terem sido pilares importantes e forças revigorantes neste processo. Obrigada pelas palavras sempre certas nos momentos certos. Obrigada pelo apoio desmedido que me deram sempre, em todos os momentos da minha vida. Obrigada por esta amizade tão bonita que construímos e que levamos para o resto da nossa vida.

Aos meus avós e restantes familiares, pelo carinho e o amor que me dão, pelas palavras de esperança e confiança que me transmitem sempre e por me mostrarem que a família é o lugar mais bonito do mundo.

Aos professores da Faculdade de Psicologia da Universidade de Lisboa, que me proporcionaram um percurso académico pautado pelo rigor, a ética e espírito crítico. Por me ensinarem e me ajudarem a desenvolver competências profissionais e pessoais fundamentais e por me mostrarem a importância da paixão pela profissão que teremos.

Aos participantes do presente estudo, pela disponibilidade e interesse que demonstraram em contribuir para este projeto.

Ao Centro Psicogeriátrico de Nossa Senhora de Fátima das Irmãs Hospitaleiras do Sangrado Coração de Jesus e, em especial, à Dr^a Paula Agostinho. À Cruz Vermelha Portuguesa, particularmente à Residência Sénior da Parede e, sobretudo, à Dr^a Mónica Bissaya Barreto. À Junta de Freguesia de Alvalade, em particular ao Espaço Sénior de Entrecampos e à Dr^a Ana Almeida, e aos Briosos, em especial, ao Engenheiro Luís Aires e à Dr^a Helena. À Universidade Sénior da Ajuda, em particular ao Dr. Tiago Lopes e à Dr^a Rute Júlio. À Universidade Sénior de Alcântara, especialmente à Dr^a Patrícia Cara D'Anjo. A todas as Instituições e profissionais envolvidos, o meu muito obrigada por toda a vossa disponibilidade e apoio na agilização do processo de recolha de dados.

Resumo

O presente estudo teve como objetivo validar o Teste de Validade de Desempenho *Coin in Hand-Extended Version* (CIH-EV; Daugherty et al., no prelo) com grupos de adultos idosos saudáveis e com demência, segundo um paradigma de simulação. Um dos grupos, de idosos saudáveis não-simuladores (n=30) foi instruído a ter o melhor desempenho possível, e outro grupo, de idosos saudáveis simuladores (n=29) foi instruído a simular um déficit mnésico típico de uma demência, com o objetivo de obter algum tipo de pensão, subsídio ou reforma antecipada. Além destes, incluiu-se um grupo clínico (n=17), com diagnóstico de demência. Os resultados mostraram que os grupos não-simulador e clínico tiveram um desempenho mais preciso no CIH-EV do que o grupo simulador, ainda que o grupo clínico tenha apresentado tempos de resposta superiores. A precisão e o tempo de resposta não diferiram entre os três níveis de dificuldade do teste nos três grupos. Através da análise das curvas ROC, determinou-se o ponto de corte (≤ 26) para adultos idosos saudáveis a partir da comparação entre os grupos não-simulador e simulador, traduzindo 100% de sensibilidade e 97% de especificidade. A comparação entre os grupos simulador e clínico sugeriu o ponto de corte ≤ 17 para idosos com demência, representando 94% de sensibilidade e 97% de especificidade. O CIH-EV apresentou boa validade convergente com as diferentes medidas de validade de desempenho utilizadas (TOMM, Rey-15IMT e RDS) e mostrou-se insensível às variáveis sociodemográficas e ao funcionamento neurocognitivo em todos os grupos. Além disso, apresentou uma reduzida taxa de falsos positivos relativamente aos outros testes e, em geral, foi o instrumento com maior precisão de diagnóstico nos três grupos. Assim, o presente estudo corrobora a utilidade do CIH-EV como um instrumento eficaz na detecção de simulação de déficit cognitivo em adultos idosos saudáveis e em idosos com demência.

Palavras-chave: Validade de desempenho; simulação; CIH-EV; idosos; demência.

Abstract

The aim of the current study was to validate the Performance Validity Test *Coin in Hand-Extended Version* (CIH-EV; Daugherty et al., in press) in groups of healthy older adults and older adults with dementia, according to a simulation paradigm. One of the groups, the non-simulating healthy older adults (n=30) was instructed to have their best performance, and the other group, the simulating older adults (n=29) was instructed to simulate a typical dementia memory deficit, with the purpose of obtaining allowance, financial aid or early retirement. Furthermore, a clinical group (n=17), diagnosed with dementia, was included. The results showed that the non-simulator and clinical groups have performed more accurately than the simulator group, although the clinic group have shown superior response times. There were no significant differences in accuracy and response time between the three test difficulty levels, in all groups. Through ROC curves analysis, the cut-off point (≤ 26) for healthy older adults was determined from the comparison between non-simulator and simulator groups, showing 100% of sensibility and 97% of specificity. The comparison between the simulator and clinical groups suggested a cut-off point ≤ 17 for older adults with dementia, showing 94% of sensibility and 97% of specificity. The CIH-EV demonstrated good convergent validity with all performance validity measures used (TOMM, Rey-15IMT and RDS) and was insensitive to sociodemographic variables and neurocognitive functioning in all groups. Besides, it showed reduced rate of false positives when compared to the other tests and, generally, it was the most accurate diagnostic test in the three groups. Thus, the present study corroborates the CIH-EV utility as an effective instrument in detecting the simulation of cognitive deficits in healthy older adults and older adults with dementia.

Keywords: Performance validity; simulation; CIH-EV; older adults; dementia

Índice

Introdução	1
I. Enquadramento Teórico	2
1.1. Simulação: definição e critérios de diagnóstico	2
1.2. Conceito de Simulação e suas implicações	6
1.3. Detecção da Simulação: Metodologias	8
1.3.1. Testes (neuro)psicológicos tradicionais.....	10
1.3.1.1. Subteste de Memória de Dígitos da WAIS-III-R	11
1.3.2. Testes de Validade de Sintomas/Testes de Validade de Desempenho.....	16
1.3.2.1. <i>Test of Memory Malingering</i> (TOMM)	20
1.3.2.2. <i>Rey-15 Item Memory Test</i> (Rey-15IMT)	25
1.3.2.3. <i>Coin in Hand-Extended Version</i> (CIH-EV)	31
1.3.2.4. Testes de Validade de Desempenho e Demência	35
1.3.3. A dicotomia entre medidas <i>embedded</i> (integradas) <i>versus stand-alone</i> (isoladas).....	37
1.4. O presente estudo	39
1.4.1. Pertinência do Teste de Validade de Desempenho CIH-EV.....	39
1.4.2. Pertinência das amostras selecionadas: adultos idosos saudáveis e com demência.....	40

1.4.3. Objetivos	42
II. Método	44
2.1. Participantes.....	44
2.2. Instrumentos	45
2.2.1. Questionário de Dados Sociodemográficos	45
2.2.2. <i>Mini-Mental State Examination</i> (MMSE)	45
2.2.3. <i>Montreal Cognitive Assessment</i> (MoCA)	46
2.2.4. Subteste de Memória de Dígitos da WAIS-III-R	47
2.2.5. <i>Rey-15 Item Memory Test</i> (Rey-15IMT)	47
2.2.6. <i>Test of Memory Malingering</i> (TOMM)	48
2.2.7. <i>Coin in Hand – Extended Version</i> (CIH-EV)	49
2.3. Procedimento	49
III. Resultados	52
3.1. Análise das diferenças entre grupos nas variáveis sociodemográficas e no desempenho cognitivo	52
3.2. Análise da relação entre as variáveis sociodemográficas, o desempenho cognitivo e o desempenho nos testes	55
3.3. Análise da precisão e do tempo de resposta dos três grupos para os níveis de dificuldade e para o resultado total do CIH-EV.....	55
3.4. Análise e determinação dos pontos de corte do CIH-EV para idosos saudáveis.....	59

3.5. Análise e determinação dos pontos de corte do CIH-EV para idosos com diagnóstico de demência.....	60
3.6. Análise da validade convergente	61
IV. Discussão e conclusões	66

Introdução

Embora a problemática da simulação e do exagero de sintomas acompanhe as sociedades desde sempre (Wessely, 2003), a sua investigação e consideração clínica é relativamente recente, contando com um trabalho intensivo de aproximadamente duas décadas (Heilbronner et al., 2009). O interesse crescente por esta área justifica-se, por um lado, pela observação de que os resultados dos testes não são transparentes e podem ser afetados por variáveis tais como a motivação, o esforço e a cooperação no momento da avaliação (Larrabee, 1992; Merten et al., 2013) e não apenas por um défice cognitivo ou uma lesão ou disfunção cerebral orgânica (Simões et al, 2010; Vickery, Berry, Inman, Harris & Orey, 2001). O alerta para a necessidade de considerar este tipo de variáveis (e.g., esforço, simulação, exagero de sintomas) decorreu da constatação de que os profissionais não eram capazes de identificar os desempenhos inválidos de forma eficaz, baseando-se apenas na sua intuição clínica (e.g., Faust, Hart & Guilmette, 1988; Faust, Hart, Guilmette & Arkes, 1988; Heaton, Smith, Lehman, & Vogt, 1978), pelo que passou a ser necessário procurar medidas e estratégias que permitissem maximizar a confiança nos resultados dos testes e nos diagnósticos e tratamentos deles decorrentes (Bush et al., 2005; Heilbronner et al., 2009). Embora se observe uma evolução significativa relativamente a tais medidas, esta necessidade permanece atual, já que uma percentagem considerável de clínicos continua sem incluir este tipo de medidas nas suas avaliações (Dandachi-FitzGerald, Ponds & Merten, 2013; O'Bryant, Engel, Kleiner, Vasterling & Black, 2007), o que leva a predições incorretas em cerca de 24% dos casos (Dandachi-FitzGerald, Merckelbach, & Ponds, 2017).

Por outro lado, o investimento nesta área é também justificado pelo facto de existirem cada vez mais questões e decisões de contexto forense e médico-legal que estão dependentes dos resultados da avaliação (neuro)psicológica (Vickery et al., 2001). Desta forma, podem existir casos em que se observa quer o exagero ou fabricação de sintomas para obter ganhos

secundários e pessoais, quer a sua minimização e recusa para evitar consequências indesejáveis (Cima et al., 2003). Tais circunstâncias tornam premente a avaliação sistemática da credibilidade dos desempenhos associados a incentivos externos, uma vez que estão em causa decisões com impactos clínicos, financeiros e sociais, que limitam os recursos disponíveis para os indivíduos com perturbações reais (Lippa, 2018). Adicionalmente, tem sido recomendado que a avaliação da credibilidade dos desempenhos esteja presente na prática clínica quotidiana, deixando de ser exclusiva aos contextos forenses e/ou médico-legais (e.g., Bush et al., 2005; Heilbronner et al., 2009).

Neste sentido, o exagero de sintomas, o esforço insuficiente e a simulação são conceitos que se tornam cada vez mais afirmativos na prática clínica e na investigação (e.g., Boone, 2007; Larrabee, 2007; Morgan & Sweet, 2009), ao nível internacional e nacional (Simões et al., 2010), pelo que é necessário esclarecê-los.

I. Enquadramento Teórico

1.1. Simulação: definição e critérios de diagnóstico

Em 1990, a discussão do conceito de simulação ganhou relevo quando Rogers (1990b) documentou a existência de três principais modelos explicativos deste comportamento, que derivaram de definições distintas. O *modelo patogénico* assenta no pressuposto de que a simulação representa uma tentativa mal-sucedida de controlar processos psicóticos e neuróticos inconscientes, que se expressam através de psicopatologia (Rogers, 1990b); o *modelo criminológico*, associado à proposta do DSM-III-R, pressupõe que o comportamento de simulação acontece em indivíduos com perturbação de personalidade anti-social, que fabricam doenças em contextos médico-legais e se mostram pouco cooperativos em avaliações e tratamentos (Rogers, 1990a; 1990b). Apontando diversas limitações a ambos os modelos, o autor propõe um modelo alternativo, o *modelo adaptativo*. Este baseia-se na Teoria da Decisão,

segundo a qual uma decisão é tomada em função da sua utilidade e probabilidade, pelo que Rogers (1990a; 1990b) sugere que de entre uma diversidade de escolhas possíveis, os indivíduos escolhem simular quando esta opção apresenta maior probabilidade de sucesso. É na medida em que o comportamento de simulação permite alcançar os objetivos pessoais que o autor reconhecer o seu carácter adaptativo. Rogers (1990a) propôs ainda critérios de diagnóstico para a simulação de perturbação psiquiátrica, que integram dados de diferentes domínios. Os critérios incluem manifestação de um número invulgarmente elevado de sintomas raros, informação colateral contraditória e evidência de exagero ou fabricação de sintomas a partir dos testes, devendo ser especificados para um diagnóstico mais preciso.

Outros autores sugeriram definições e critérios distintos, como foi o caso de Pankrantz e Erickson (1999). Uma vez que a intenção não pode ser avaliada com precisão, os autores defendem que o diagnóstico não deve incluir julgamentos sobre o estado interno do indivíduo e, por isso, numa perspetiva exclusivamente comportamental, propõem um critério baseado na desproporção entre a perturbação percebida e as características do défice (Pankrantz, 1988). No mesmo sentido, Brandt (1988) recomendou a utilização da inconsistência entre as queixas sintomáticas e os princípios da neuropsicologia como critério de diagnóstico. Ainda segundo a mesma perspetiva, Faust e Ackley (1998) propuseram uma lista de seis manifestações comportamentais, porém, considerando que estas refletiam a intenção do indivíduo. No entanto, até então, nenhum dos critérios mostrara a sua eficácia em termos empíricos. Desta forma, Greiffenstein, Baker e Gola (1994) validaram um conjunto de quatro critérios de diagnóstico para simulação de perturbação mnésica numa amostra em contexto de litígio, de forma a fornecerem dados empíricos que os sustentassem. Estes incluem 1) desempenho baixo improvável em duas ou mais medidas de validade; 2) incapacidade total para desempenhar um papel social relevante; 3) contradição entre informações colaterais e história de sintomas; e 4) perda de memória remota. No mesmo ano, a definição de simulação foi atualizada na quarta

edição do DSM (DSM-IV; American Psychological Association [APA], 1994), em geral pressupondo a existência de fabricação ou exagero de sintomas, motivação ou intenção consciente para simular e presença de incentivos externos. Sendo classificada como um código-V e não uma perturbação mental, a simulação pode e deve ser alvo de atenção, mas não tem critérios de diagnóstico associados (APA, 1994).

Embora diversas propostas fossem lançadas, os critérios continuavam a não ser satisfatórios para alguns autores, entre os quais Slick, Sherman e Iverson (1999), que apontaram diversas limitações. Relativamente aos critérios sugeridos por Rogers (1990a), estes não eram sustentados por dados estatísticos e, em geral, não consideravam evidência psicométrica para a avaliação neuropsicológica de déficit cognitivo, uma vez que foram propostos no âmbito psiquiátrico (Slick et al., 1999). Embora na direção certa, Greiffenstein e colaboradores (1994) não incluíam uma definição explícita de simulação e os seus critérios não incluíam observações comportamentais nem critérios de exclusão ou de diagnóstico diferencial, alguns careciam de especificidade e, em geral, estes só podiam ser aplicados à simulação de déficit mnésico, excluindo outros domínios neurocognitivos (Slick et al., 1999). Relativamente ao DSM-IV (APA, 1994), segundo Slick e colaboradores (1999), este apresentava uma definição geral, não fornecia critérios formais, mas antes índices de suspeita clínica, fazia uma referência muito breve ao diagnóstico diferencial, baseava-se nas dicotomias incentivo externo *versus* interno e comportamento volitivo *versus* não consciente que, na prática, não permitiam uma discriminação clara e, finalmente, especificava outras condições que envolviam o exagero de sintomas, mas que não eram consideradas simulação. Em relação à atualização dos critérios de Pankratz e Binder (1997), estes mantinham-se exclusivamente comportamentais, eliminando qualquer tipo de avaliação de intenção ou estado interno. Porém, esta avaliação integra igualmente o diagnóstico de perturbações mentais diversas, sobre as quais já não se levanta esta problemática (Slick et al., 1999).

Face a estas limitações, Slick e colaboradores (1999) atualizaram a definição de simulação para *o exagero ou a fabricação intencional de perturbação cognitiva, com o objetivo de obter ganho material substancial ou evitar ou escapar de dever ou responsabilidade formal* (Slick et al., 1999; p.552). Simultaneamente, propuseram um conjunto de critérios de diagnóstico para Simulação de Perturbação Neurocognitiva (SPN), com o objetivo de integrar, pela primeira vez, informações psicométricas, comportamentais e colaterais no diagnóstico. Estes incluem A) presença de, pelo menos, um incentivo externo substancial para o exagero ou a fabricação de sintomas no momento da avaliação; B) evidências destes comportamentos a partir da avaliação neuropsicológica, que podem incluir, um ou mais resultados abaixo do acaso em medidas de escolha forçada e diferentes padrões de discrepância; C) evidências a partir do autorrelato, com base em padrões de inconsistência distintos; e D) os comportamentos descritos nos critérios B e C não são melhor explicados por doença psiquiátrica, neurológica ou perturbação do desenvolvimento. Dentro de cada critério, a especificação dos comportamentos, à semelhança do que acontece no DSM-IV (APA, 1994), permite um diagnóstico de SPN com um nível de confiança possível, provável ou definitiva. Ainda que estes sejam os critérios mais utilizados até aos dias de hoje, não deixam de ter algumas críticas associadas (e.g., Boone, 2007; Larrabee, 2007), das quais se destaca a necessidade de determinar a intencionalidade da simulação. De facto, Larrabee (2003) tem mostrado empiricamente que é a improbabilidade dos resultados, na presença de um incentivo externo e na ausência de explicações alternativas válidas que constitui, por si só, a intenção do indivíduo para simular (Larrabee, Greiffenstein, Greve & Bianchini, 2007).

Não obstante, os diferentes níveis de confiança propostos por Slick e colaboradores (1999) refletem necessariamente a ausência de um padrão único e exclusivo de simulação, o que levou diversos autores a rejeitar a perspetiva da simulação como um fenómeno tudo-ou-nada, no qual os simuladores representam uma categoria de pessoas qualitativa e

quantitativamente distinta dos não-simuladores (Ferguson, 2004). Em contrapartida, defendem a existência de um *continuum* de sintomas reportados e de desempenhos nas medidas de esforço (Larrabee, 2005). Foi neste sentido que Ferguson (2004) propôs a existência de quatro níveis contínuos de simulação: 1) atribuição incorreta de sintomas, em que existem sintomas que, embora não se relacionem com o défice, são intencionalmente atribuídos à mesma; 2) manutenção ou agravamento de sintomas, em que o indivíduo já recuperou de défices outrora existentes, mas insiste na sua permanência ou agravamento; 3) simulação parcial ou exagero de sintomas, na qual estão presentes dificuldades genuínas, porém, são maximizadas; e 4) simulação pura, na qual não existe défice, mas os indivíduos simulam a sua presença. Do mesmo modo, autores como Bush e colaboradores (2005), Larrabee (2005) e Walters e colaboradores (2008) defendem que os comportamentos de simulação variam entre diferentes níveis de fabricação ou exagero, numa perspetiva dimensional, o que se reflete quer na existência de subgrupos de simuladores (Larrabee, 2005), quer em diferentes combinações e padrões de desempenho em diversos TVDs (Larrabee, 2003).

1.2. Conceito de Simulação e suas implicações

A evolução do conceito e da abordagem à simulação culminou numa distinção pertinente entre diagnóstico e deteção (Heilbronner et al., 2009). O diagnóstico pressupõe a determinação explícita da *intencionalidade* do exagero ou da simulação de sintomas, a partir da exclusão de outras condições psicológicas, neurológicas e de desenvolvimento que possam justificar esse comportamento e da integração de múltiplos indicadores (Slick et al., 1999). De facto, a determinação da intenção não tem de assentar num evento singular e definitivo, podendo decorrer da combinação de diversos indicadores improváveis (Larrabee, Greiffenstein, Greve & Bianchini, 2007). Desta forma, é necessário considerar os desempenhos nos testes e/ou queixas sintomáticas improváveis, complementando-os, tanto quanto possível,

com informação de fontes adicionais e do historial apresentado durante a entrevista. Adicionalmente, de forma a garantir que o exagero intencional deriva da manifestação de simulação (por oposição ao que acontece noutras condições, por exemplo, na perturbação factícia), deve determinar-se o contexto da avaliação, especificamente se existe um ganho externo associado (Heilbronner et al., 2009). Torna-se, assim, pertinente esclarecer um conceito adicional, o de *ganho secundário*. Por oposição a um ganho primário, normalmente emocional, o ganho secundário inclui recompensas financeiras, baixas médicas profissionais, evitamento do dever militar e de consequências legais e obtenção de medicamentos, estando, por isso, presente em contextos que envolvem litígio, queixas de incapacidade, processos criminais e indemnizações (DSM-V; APA, 2014). Relativamente ao diagnóstico diferencial, diversos autores de referência reuniram-se (Heilbronner et al., 2009) e propuseram que existe a possibilidade de simulação de capacidades quando se verifica 1) disparidade entre as observações da vida real e o desempenho nos testes ou o autorrelato; 2) inconsistência entre o tipo ou a severidade da lesão e o desempenho nos testes; 3) inconsistência entre o comportamento do indivíduo quando sabe que está a ser avaliado *versus* quando não sabe; e 4) inconsistência ao longo de avaliações seriais, que não podem ser explicadas por processos neurológicos ou condições psiquiátricas conhecidas. Não obstante, o diagnóstico de simulação implica necessariamente a realização de uma inferência e, neste sentido, deve ser cientificamente baseada em conhecimento teórico e empírico atualizado (Heilbronner et al., 2009; Larrabee, 2005; Silck et al., 1999).

Por sua vez, a deteção da simulação pode considerar-se como um meio que possibilita o diagnóstico, podendo ocorrer em qualquer momento da avaliação neuropsicológica dos sintomas, das capacidades e/ou do funcionamento do indivíduo. Neste caso, a validade da avaliação depende do *esforço* investido na realização dos testes, que deriva inevitavelmente da motivação do indivíduo (Lezak, Howieson, Bigler & Tranel, 2012). O esforço e a simulação

são dois conceitos distintos, ainda que ambos possam ser vistos num *continuum*. O primeiro estende-se desde um *esforço reduzido* a um *esforço ótimo* (Larrabee, 2012) e a sua variabilidade pode expressar-se no desempenho quer de um mesmo teste, quer entre diferentes testes. Assim, muitas vezes, os baixos resultados nos testes podem representar *esforço reduzido* e não necessariamente um comportamento de simulação (Slick et al., 1999; Sweet et al., 1999). Da mesma forma, este cenário pode ser também explicado pelo *exagero de sintomas ou de défices* que, tal como o esforço, se traduz em estilos e enviesamentos de resposta específicos e que, nos testes, se manifestam em padrões (e.g., Greiffenstein, Baker, & Gola, 1996; Mittenberg, Azrin, Millsaps, & Heilbrunner, 1993) e níveis atípicos (Greiffenstein, 2007).

No entanto, estes conceitos são frequentemente confundidos e carecem de consenso e objetividade entre os autores da área (Bigler, 2012), pelo que Larrabee (2012) propôs a sua substituição. Uma vez que os indivíduos com perturbações neurológicas reais têm desempenhos ótimos nas medidas utilizadas para a deteção deste tipo de comportamentos, tal significa que estas requerem níveis mínimos de capacidade e, conseqüentemente, de esforço. Neste sentido, Larrabee (2012), apoiado por Bigler (2012), propôs a substituição dos conceitos de “esforço”, “validade de sintomas” e “enviesamento de resposta” por um termo mais descritivo, *validade de desempenho*, para referir em que medida o desempenho traduz uma representação precisa do nível de capacidade real. Além disso, o autor considera que o termo *validade de sintomas* está mais associado às queixas subjetivas do que ao desempenho em si, pelo que recomenda a sua utilização em referência à validade das queixas sintomáticas em medidas de autorrelato (Larrabee, 2012).

1.3. Deteção da Simulação: Metodologias

Tal como se verificou uma evolução na definição de simulação, também os métodos utilizados para a sua deteção foram sistematicamente atualizados, com base em avanços nos

conhecimentos teóricos e empíricos. De facto, nos seus primórdios, a deteção da simulação e de outros estilos de resposta era maioritariamente centrada na simulação de perturbações psiquiátricas (Rogers, 2008), segundo análises qualitativas e subjetivas das características “salientes” dos indivíduos. Estas eram detetadas com base em indivíduos suspeitos de simulação que, por sua vez, eram identificados a partir destas características, observando-se, portanto, um viés confirmatório (Borum, Otto & Golding, 1993; Rogers, 2008). Dadas as limitações inerentes a este método, numa tentativa de sustentar de forma mais válida as suas conclusões, os clínicos passaram a basear-se na análise qualitativa das respostas a testes tradicionais (e.g., WAIS-III-R; Wechsler, 1997, 2008; MMPI-2; Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989; AVLT; Binder, Villanueva, Howieson & Moore, 1993), especificamente nos padrões de resposta inconsistentes (Slick et al., 1999; Tombaugh, 1997), uma vez que se pressupõe que os padrões de resultados têm de fazer “sentido neuropsicológico” (Larrabee, 1990). Em geral, com base em evidência empírica, Larrabee (2005) conclui que os testes tradicionais mais sensíveis ao esforço reduzido são testes de funcionamento sensoriomotor, atenção, memória (reconhecimento) e resolução de problemas. Porém, considerando que os testes tradicionais não foram desenvolvidos para avaliar especificamente desempenhos não credíveis, só podem reportar, de forma geral, o desvio de um padrão esperado (Tracy, 2014). Neste sentido, foi introduzido um novo método, baseado num paradigma de escolha forçada, desenvolvido especificamente para avaliar a credibilidade dos desempenhos e sintomas reportados, os Testes de Validade de Sintomas (TVSs; Larrabee, 2005). Por sua vez, este tipo de teste pode ser utilizado para avaliar a validade do desempenho cognitivo e, neste sentido, existem Testes de Validade de Desempenho (TVDs; *Performance Validity Tests*), e/ou para avaliar a validade de sintomas, para a qual existem Testes de Validade de Autorrelato (TVARs; *Self-report Validity Tests*) (Larrabee, 2012; Merten et al., 2013).

Desta forma, para além de métodos complementares, como a observação comportamental e as informações colaterais (Merten et al., 2013), a deteção de comportamentos não-credíveis pode ser feita com base nos resultados em testes (neuro)psicológicos tradicionais e em Testes de Validade de Sintomas (TVSs), que se podem especificar em Testes de Validade de Desempenho (TVDs) e Testes de Validade de Autorrelato (TVARs). Independentemente do método utilizado, Rogers (2008) sugere que devem respeitar quatro pressupostos: 1) ser um método standardizado; 2) ter uma base conceptual; 3) ter validação empírica associada; 4) apresentar diferenciação sistemática; e 5) avaliar um estilo de resposta específico. De seguida, as metodologias mais utilizadas no contexto da avaliação neuropsicológica serão discutidas com maior detalhe.

1.3.1. Testes (Neuro)psicológicos tradicionais

A administração de testes (neuro)psicológicos tradicionais com o objetivo de detetar comportamentos não-credíveis justifica-se pelo facto de estes permitem derivar indicadores de validade, que se dizem *embedded*, úteis para essa identificação (e.g., Meyers & Volbrecht, 2003; Sherman, Boone, Lu & Razani, 2002). Exemplos das estratégias de deteção associadas aos testes tradicionais são a magnitude do erro, a curva de desempenho, a violação dos princípios de aprendizagem e o efeito chão (Rogers, 2008). De facto, diversas vantagens têm sido apontadas à utilização dos testes tradicionais (e.g., Rogers, 2008), entre as quais a) permitirem a expansão dos métodos de deteção da simulação desde os formatos de reconhecimento de memória iniciais; b) servirem diversos propósitos simultaneamente, avaliando quer os défices genuínos, quer os défices simulados; c) avaliarem retrospectivamente a validade das respostas em situações anteriores (Mittenberg, Aguila-Puentes, Patton, Canary & Heilbronner, 2002; d) economizar o tempo da avaliação; e) avaliar a validade do desempenho de forma mais contínua ao longo da sessão; e e) protegerem os desempenhos do *coaching* e da

pesquisa *online*, na medida em que são menos transparentes, o que faz com que os indivíduos não depreendam que a validade do seu desempenho está a ser avaliada (Rogers, 2008).

Desta forma, os testes mais utilizados, a partir dos quais se derivam indicadores de validade, são a Wechsler Adult Intelligence Scale (WAIS-III-R; Wechsler, 1997, 2008), especificamente os subtestes de Memória de Dígitos e de Vocabulário, a Wechsler Memory Scale (WMS; Wechsler, 1945; 2009), o Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989), o Auditory Verbal Learning Test (AVLT; Binder, Villanueva, Howieson & Moore, 1993; Boone, Lu & Wen, 2005), o California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan & Ober, 1987; Millis, Putnam, Adams & Ricker, 1995), o Wisconsin Card Sorting Test (WCST; Bernard et al., 1996; Suhr & Boyer, 1999), o Category Test (Tenhula & Sweet, 1996) e o Benton Visual Form Discrimination (BFD; Benton, Sivan, Hamsher, Varney & Spreen, 1994). Especificamente o subteste de Memória de Dígitos tem sido o mais estudado e, por isso, aquele que apresenta maior sustentação empírica para a deteção da simulação (Boone, 2007). A investigação até aqui realizada permitiu verificar que, de facto, esta é a medida com melhor precisão de diagnóstico. Por estas razões, esta foi a medida de validade tradicional selecionada para integrar o protocolo de avaliação do presente estudo e, neste sentido, será discutida com maior profundidade.

1.3.1.1. Subteste de Memória de Dígitos (WAIS-III-R; Wechsler, 1997, 2008; Pinho, 2012)

O subteste de Memória de Dígitos integra um instrumento de avaliação mais extenso e compreensivo do quociente de inteligência, a WAIS-III-R (1997, 2008), tendo sido integrado com o objetivo de avaliar a memória de trabalho, a atenção e a concentração. Não obstante, diversos estudos têm mostrado a eficácia deste subteste na deteção de comportamentos de simulação e/ou exagero de sintomas. Este consiste na repetição de sequências numéricas de

dificuldade crescente, sendo que em cada item é adicionado um número. O subteste é composto por duas condições: i) repetição das sequências em sentido direto, com um total de 8 itens, com 2 ensaios cada; e ii) repetição das sequências em sentido inverso, com 7 itens, com 2 ensaios cada. Os itens são cotados com 0 pontos se as sequências de ambos os ensaios forem repetidas incorretamente, com 1 ponto se um ensaio for repetido corretamente e com 2 pontos se o indivíduo reproduzir corretamente os dois ensaios. Em ambos os sentidos, a administração das sequências termina quando o participante reproduz incorretamente os dois ensaios de um mesmo item. No sentido direto, é possível obter uma pontuação máxima de 16 pontos e no sentido inverso é possível atingir 14 pontos, com uma pontuação máxima total de 30 pontos.

A partir do subteste de Memória de Dígitos é possível obter a pontuação do sentido direto, a pontuação do sentido inverso e resultados que têm sido usados como medidas de esforço, entre as quais o Digit Span Age-Corrected Scale Score (ACSS, Iverson, Franzen & McCracken, 1991; Iverson & Franzen, 1994, 1996; Suhr, Tranel, Wefel, & Barrash, 1997) e o Reliable Digit Span (RDS). Este último foi desenvolvido por Greiffenstein, Baker e Gola (1994), com base na WAIS-III (Wechsler, 1997), sendo calculado através da soma da sequência mais longa do sentido direto e do sentido inverso, nas quais ambos os ensaios foram corretamente reproduzidos. A investigação tem mostrado que o ACSS e o RDS não diferem significativamente entre si, pelo que o clínico pode optar pela utilização de apenas um deles (Jasinski, Berry, Shandera & Clark, 2011; Young, Sawyer, Roper & Baughman, 2012).

A eficácia do RDS na detecção de simulação e esforço insuficiente tem sido corroborada por diversos estudos, sendo que se encontra validado para amostras normativas de jovens adultos (e.g., Inman & Berry, 2002; Iverson & Franzem 1994), adultos (e.g., Babikian, Boone, Lu & Arnold, 2006; Shroeder et al., 2012) e adultos idosos (Choi et al., 2014); para amostras clínicas, em pacientes com TCE (e.g., Greiffenstein et al., 1994; Larrabee, 2003; Mathias, Greve, Bianchini, Houston & Crouch, 2002), síndrome pós-concussional (e.g., Greiffenstein et

al., 1994), lesão cerebral (e.g., Larrabee, 2003), perturbação intelectual (e.g., Greue et al., 2007; Marshall & Happe, 2007), perturbação do espectro autista (Marshall et al., 2010), perturbações psiquiátricas e perturbações de memória severas (Schroeder et al., 2012); para amostras forenses, com indivíduos suspeitos de simular síndrome pós-concussional (e.g., Greiffenstein et al., 1994; Ylioja, Baird & Podell, 2009), perturbação neurocognitiva (e.g., Etherton, Bianchini, Greve & Heinly, 2005; Larrabee, 2003; Mathias et al., 2002;) derivada da exposição a tóxicos (Greve et al., 2007), perturbação do espectro autista (Marshall et al., 2010), traumatismo cranioencefálico ligeiro e severo (e.g., Axelrod, Fichtenberg, Millis & Wertheimer, 2006; Heinly et al., 2005), perturbação intelectual (Greue et al., 2007) e dor crónica (Greve et al., 2010), com indivíduos envolvidos em pedidos de subsídios de incapacidade, devidos a perturbação médica ou psiquiátrica (Babikian et al., 2006) e com indivíduos em estabelecimentos prisionais (Duncan & Ausborn, 2002); e, finalmente, em paradigmas experimentais, nos quais estudantes são instruídos a simular défice mnésico (e.g., Etherton et al., 2005; Heaton, Smith, Lehman & Vogt, 1978; Inman & Berry, 2002; Strauss et al., 2000) e reclusos e pacientes psiquiátricos são instruídos a simular falhas de memória no contexto de litígio pessoal (Iverson & Franzen, 1994, 1996).

O ponto de corte mais adequado para o ACSS parece ser de ≤ 5 pontos, traduzindo valores de sensibilidade a partir de 42% e de especificidade a partir de 86,5%, bem como uma taxa de falsos positivos relativamente baixa (Axelroad et al., 2006; Babikian et al., 2006; Jasinski et al., 2011). Relativamente ao ponto de corte para o RDS, a proposta inicial era ≤ 7 pontos (Greiffenstein et al., 1994), porém, diversos estudos têm verificado que este ponto de corte leva a valores de especificidade inferiores a 90% para diversos grupos clínicos (Axelroad et al., 2006; Schroeder et al., 2012) e a uma taxa de falsos positivos com um valor inaceitável de 23% (Babikian et al., 2006). Desta forma, são vários os autores que sugerem um ponto de corte ≤ 6 que, embora reduza a sensibilidade para valores entre 45% e 63% (Babinkian et al.,

2006; Jasinski et al., 2011), aumenta a especificidade para valores acima de 90% para diversos grupos clínicos e reduz a taxa de falsos positivos para até 10% (Babinkian et al., 2006; Schroeder et al., 2012). Não obstante, Schroeder e colaboradores (2012) alertam para a inadequação deste ponto de corte para pacientes com AVCs, perturbações de memória severas e perturbação do desenvolvimento intelectual e para pessoas hispânicas e com o inglês como segunda língua (Schroeder et al., 2012).

Em Portugal, o subtteste de Memória de Dígitos, em contexto de deteção da simulação, de exagero de sintomas e/ou de esforço insuficiente foi utilizado em apenas dois estudos, que incluíram amostras normativas de adultos (Castro, 2015) e adultos idosos (Pinho, 2012); amostras clínicas, com grupos de DCL (Pinho, 2012) e de depressão (Castro, 2015); e paradigmas experimentais, nos quais foi pedido aos indivíduos que simulassem défice cognitivo (Pinho, 2012) e sintomas depressivos (Castro, 2012). No estudo de Castro (2015) foram apenas considerados o resultado padronizado e as pontuações de ambos os sentidos (direto e inverso). Desta forma, apenas o estudo de Pinho (2012) utilizou os índices usados no contexto de deteção de comportamentos suspeitos de simulação, o ACSS e o RDS. Usando um ponto de corte ≤ 5 para o ACSS, obteve uma sensibilidade de 10.34% para os adultos idosos simuladores de défice cognitivo e uma especificidade de 100% para o grupo controlo e de 94.12% para o grupo com DCL. Já para o RDS, o ponto de corte ≤ 6 resultou numa sensibilidade de 13.79% e numa especificidade de 96.77% para o grupo controlo e 70.59% para o grupo com DCL.

Em geral, alguns estudos têm mostrado que os resultados no subtteste de Memória de Dígitos são influenciados pelo nível educacional (e.g., Choi et al., 2013; Kaufman, Mclean & Reynolds, 1988), pela idade (e.g., Choi et al., 2013, Grégoire & Van der Linden, 1997) e pelo género, com desempenhos mais pobres do sexo feminino (Choi et al., 2013). Desta forma, Choi e colaboradores (2013) estratificaram os dados normativos para o sentido direto e inverso em

função das variáveis sociodemográficas referidas. Embora a influência destas variáveis possa ser, assim, minimamente contornável, o RDS tem mostrado sensibilidade adicional a défices cognitivos e intelectuais severos, bem como a perturbações psiquiátricas severas (e.g., Merten, Bossink & Schmand, 2006; Glassmire, Ross, Kenney & Nitch, 2016; Rickards, Cranston, Touradji & Bechtold, 2017).

Particularmente em amostras clínicas de adultos idosos observa-se uma redução da especificidade com o aumento da severidade do défice cognitivo, padrão observado no caso português (Pinho, 2012). De facto, Dean, Victor, Boone, Philpott e Hess (2009), utilizando um ponto de corte para o ACSS de ≤ 5 pontos numa população com demência, obtiveram uma especificidade de 73% e de 70% para o RDS, com um ponto de corte ≤ 6 . Para além de observarem variabilidade nos valores de especificidade em função do tipo de demência, verificaram também que estes valores diminuía com valores intervalares de MMSE mais baixos, ou seja, obtiveram uma especificidade de 84% para valores de MMSE entre 21 e 30, de 67% para valores entre 15 e 20 e de 33% para menos de 15 pontos no MMSE. Da mesma forma, Loring e colaboradores (2016) verificaram que mesmo utilizando um ponto de corte ≤ 6 , este traduzia uma taxa de falsos positivos elevada para os grupos de DCL e Doença de Alzheimer em estádios iniciais. Neste sentido, Zenisek, Millis, Banks e Miller (2016) compararam os resultados associados a três pontos de corte distintos (≤ 7 , ≤ 6 e ≤ 5), recomendando o valor mais baixo, uma vez que aproximadamente 96% da amostra clínica teve resultados superiores ao ponto de corte de ≤ 5 . Este ponto de corte foi igualmente recomendado por Loring e colaboradores (2016), já que foi o único que permitiu a redução da taxa de falsos positivos para menos de 10%. Desta forma, utilizar os pontos de corte tradicionais em grupos de demência leva a interpretações de desempenho incorretas, sendo recomendável o uso de pontos de corte mais baixos e de TVS complementares (Zenisek et al., 2016).

Tais considerações devem ser tidas em conta, uma vez que o subteste de Memória de Dígitos permite economizar tempo, evita os efeitos de *coaching* observados em alguns TVSS (Babikian et al., 2006), tem mostrado valores de especificidade elevados entre estudos, variando entre 90 e 100% (Rickards, Cranston, Touradji & Bechtold, 2017) e é uma medida válida e efetiva, quando comparada com outros TVSS (Rickards et al., 2017), sendo independente da versão da WAIS utilizada (Jasinski et al., 2011).

1.3.2. Testes de Validade de Sintomas/Testes de Validade de Desempenho

Este tipo de testes avalia a credibilidade dos comportamentos numa diversidade de domínios, desde a personalidade, a inteligência e a funcionalidade motora e afetivo-emocional à capacidade cognitiva, nas múltiplas funções (e.g., memória, atenção, funções executivas) (Simões, Sousa, Costa & Almiro, 2017). Qualquer que seja o domínio avaliado, para um diagnóstico de simulação é, como vimos, importante considerar diversos tipos de fontes de informação, nas quais se incluem as queixas subjetivas. De facto, a avaliação da sua credibilidade através do autorrelato cumpre um papel importante na avaliação dos diversos padrões de inconsistência que constituem alguns dos critérios (Slick et al., 1999). Em Portugal existem dois TVSS baseados no autorrelato validados, que recorreram a amostras normativas, forenses, clínicas e experimentais: o Inventário Estruturado de Simulação de Sintomas (Structured Inventory of Malingered Symptomatology – SIMS; Smith & Burger, 1997; Widows & Smith, 2005), validado por um total de oito estudos portugueses (Mota, 2008; Oliveira, 2008; Maior, 2008; Cunha, 2011; Domingues, 2012; Pinho, 2012; Soares, 2013; Almeida, 2015) e a Escala de Validade de Sintomas – Versão 2 (EVS-2; Simões, Nunes & Cunha, 2010), desenvolvida com base em estudos experimentais.

Quando o objetivo do clínico passa pela avaliação do desempenho em tarefas que avaliam especificamente a simulação de défice cognitivo, este recorre a Testes de Validade de

Desempenho (TVDs), que podem assumir duas categorias distintas (Larrabee, 2005). Por um lado, existem TVDs que se baseiam no efeito teto e que, assim, embora aparentem uma dificuldade acrescida, são tão simples que pessoas com défices neurológicos reais os desempenham a um nível ótimo (e.g., Rey-15IMT; Rey, 1964). Desta forma, infere-se que os indivíduos que têm desempenhos significativamente baixos e, sobretudo, piores do que os pacientes reais, assumem um comportamento não credível intencional. Por outro lado, existem testes baseados num paradigma de escolha forçada, com duas opções de resposta e que, segundo o teorema binominal, permitem calcular padrões de resposta abaixo do acaso (Frederick & Speed, 2007). Sabendo que a probabilidade de acertar aleatoriamente uma resposta entre duas opções é de 50%, os indivíduos que pontuam significativamente abaixo do nível do acaso ao longo dos ensaios fornecem evidência sobre a simulação intencional do défice (Vickery et al., 2001). De facto, os indivíduos que simulam ou exageram intencionalmente os seus défices tendem a apresentar resultados excessivamente baixos, que os pacientes com perturbações cognitivas raramente atingem (Schroeder et al., 2012). Utilizando o critério do desempenho abaixo do acaso, foi desenvolvida uma estratégia normativa de inclusão de grupos de comparação para colmatar os valores de sensibilidade (i.e., capacidade do teste para identificar corretamente os indivíduos simuladores; Reznick, 2005) inaceitavelmente baixos (e.g., Greiffenstein, Baker, & Gola, 1994; Iverson, Franzen, & McCracken, 1994; Slick, Hopp, Strauss, Hunter, & Pinch, 1994). Especificamente, o desempenho dos indivíduos suspeitos de simulação é comparado o padrão de resultados obtidos por indivíduos com défices neurológicos, psiquiátricos e/ou de desenvolvimento reais sem motivação para fabricar ou exagerar défices (Binder & Willis, 1991; Guilmette et al., 1994; Vickery et al., 2001). Desta forma, nos TVDs determina-se a existência de um comportamento suspeito de simulação quando os resultados nos TVDs estão significativamente abaixo do acaso e abaixo do padrão de desempenho do grupo clínico de referência (Vickery et al., 2001).

Inevitavelmente, esta estratégia levantou a preocupação adicional de evitar a existência de falsos positivos, i.e., identificar incorretamente indivíduos não-simuladores como simuladores. Por sua vez, este fator está dependente do valor de especificidade de um teste, i.e., a sua capacidade para identificar corretamente os indivíduos não simuladores (Rezneck, 2005). Para além das questões éticas associadas, a necessidade de uma taxa de falsos positivos mínima num TVD é também apoiada por uma razão estatística: o Poder Preditivo Positivo, que se refere à probabilidade do diagnóstico. O PPP é mais dependente da especificidade do que da sensibilidade (Strauss, Richardson, Glasziou & Haynes, 2005; Larrabee, 2012), pelo que os valores de especificidade assumem maior relevância. De facto, uma das características e critérios de um TVD adequado consiste em elevados valores de especificidade (acima de 90%; Larrabee, 2005; Vickery et al., 2001), de forma a garantir uma taxa de falsos positivos até 10% (Larrabee, 2005), com a consequência inevitável de valores de sensibilidade mais baixos (a partir de 50%; Larrabee, 2005; Vickety et al., 2001).

Segundo os critérios de Slick e colaboradores (1999), o diagnóstico definitivo de Simulação de Perturbação Neurocognitiva implica, em conjunto com os restantes critérios, a existência de um resultado abaixo do nível do acaso em um ou mais TVDs de escolha forçada. No entanto, diversos estudos (e.g., Larrabee, 2003, 2008; Victor, Boone, Serpa, Buehler & Ziegler, 2009) têm evidenciado o aumento da probabilidade de deteção de simulação em função dos resultados obtidos em múltiplos TVDs. De facto, usando apenas um TVD com 50% de sensibilidade e 90% de especificidade, a probabilidade de deteção da simulação é de apenas 36%, valor que sobe para 74% com a introdução de um segundo teste e para 93% com um terceiro teste (Larrabee, 2008). Desta forma, a recomendação atual é a de que sejam utilizados vários TVDs e/ou TVSSs, uma vez que aumentam a validade preditiva incremental e, assim, a confiança relativamente à credibilidade do desempenho (Heilbronner et al., 2009; Nelson, Boone, Duek, Wagener, Lu & Grills, 2003).

Em geral, relativamente às duas categorias de TVDs, enquanto alguns testes baseados no efeito teto têm sido criticados por não serem fáceis para alguns pacientes reais (e.g., perturbação mental severa e demência) e, assim, levarem a baixos desempenhos (van Gorp & Hassenstab, 2009; Schretlen, Brandt, Krafft, & van Gorp, 1991), os testes baseados no paradigma de escolha forçada não parecem ser influenciados pela capacidade neurocognitiva nem por fatores motivacionais (Larrabee, 2012). Não obstante, seguindo a recomendação de Heilbronner e colaboradores (2009), se forem utilizados múltiplos TVDs de ambas as categorias, as limitações subjacentes podem ser ultrapassadas.

No que diz respeito aos TVDs existentes encontra-se uma grande diversidade, desde os mais tradicionais, como o Rey's Dot-Counting Test (Boone et al., 2002; Lezak et al., 2004; Rey, 1964), o b Test (Boone et al., 2000) e o Rey-15 Item Memory Test (Rey-15IMT, Rey, 1964); aos testes de paradigma de escolha forçada, como o Digit Memory Test (DMT; Guilmette, Hart & Giuliano, 1993; Hiscock & Hiscock, 1989), o The Portland Digit Recognition Test (PDRT; Binder, 1993; Binder & Kelly, 1996), o Test of Memory Malingering (TOMM; Tombaugh, 1996; 1997), o Coin in Hand (CIH; Kapur, 1994) e o Coin in Hand-Extended Version (CIH-EV; Daugherty et al., no prelo), o Victoria Symptom Validity Test (VSVT; Grote et al., 2000; Slick et al., 1996) e a Computerized Assessment of Response Bias (CARB; Conder, Allen & Cox, 1992). Dentro do próprio paradigma de escolha forçada, podem ainda destacar-se testes que utilizam palavras como estímulo, tais como o 21-Item Test (Iverson, Franzen & Mc-Cracken, 1991, 1994), o Word Memory Test (WMT; Green, Iverson & Allen, 1999; Green, Lees-Haley & Allen, 2003), o Letter Memory Test (Inman et al., 1998) e o Validity Indicator Profile (VIP; Frederick, 1997; Frederick, Crosby & Wynkoop, 2000; Frederik, Sarfaty, Johnston & Powel, 1994).

Embora todos eles tenham vantagens associadas que justifiquem a sua administração em protocolos de avaliação neuropsicológica, os instrumentos utilizados no presente estudo serão, de seguida, discutidos com maior profundidade.

1.3.2.1. *Test of Memory Malingering* (TOMM; Cruz, 2008; Fernandes, 2009; Maior, 2008; Silva, 2011; Tombaugh, 1996;1997;)

O TOMM (Tombaugh, 1996;1997; Cruz, 2008; Fernandes, 2009; Maior, 2008; Silva, 2011) é um instrumento de reconhecimento visual que assenta num paradigma de escolha forçada. Foi originalmente desenvolvido para diferenciar o desempenho de indivíduos com défices cognitivos (mnésicos) reais de desempenhos alvo de simulação, exagero e/ou esforço insuficiente (Tombaugh, 1996). Com uma duração de 15 a 20 minutos, é constituído por dois Ensaios de Aprendizagem, nos quais são apresentados 50 estímulos visuais por três segundos cada, com um intervalo de um segundo entre eles. Imediatamente após cada um destes Ensaios são apresentados 50 pares de estímulos, com um estímulo original e um distrator, sendo que a tarefa consiste em identificar aquele que apareceu anteriormente. Para todas as respostas é fornecido um feedback, o que permite aumentar a disparidade entre os potenciais simuladores e aqueles com défices reais (Simões, Almeida & Gonçalves, 2017). Após 15 minutos, é possível administrar um Ensaio de Retenção opcional, no qual é pedido que se identifiquem os 50 estímulos apresentados nos Ensaios de Aprendizagem. Em todos os Ensaios as respostas corretas são cotadas com um ponto, perfazendo um total de 50 pontos possíveis em cada um. No seu trabalho original, Tombaugh (1996) propôs duas alternativas para a deteção do comportamento de simulação ou exagero de sintomas: 1) pontuações inferiores a 25 em qualquer um dos ensaios; e 2) pontuações inferiores a 45 no segundo ensaio de aprendizagem e/ou no ensaio de retenção. A última hipótese é a mais utilizada nos contextos de investigação

e clínica, tal como acontece também em Portugal (Simões et al., 2017), sendo que a maioria dos estudos corrobora a adequação deste valor para diversos grupos normativos e clínicos.

A validação deste instrumento inclui estudos empíricos com amostras normativas, com adultos e adultos idosos, com idades entre os 18 e os 84 anos (Teichner & Wagner, 2004; Tombaugh, 1997); amostras experimentais, nas quais um dos grupos é instruído a simular; e amostras clínicas. Estas últimas incluem grupos com traumatismo cranioencefálico (TCE), défice cognitivo, afasia, demência (e.g., Greve, Bianchini & Doane, 2006; Teichner & Wagner, 2004; Tombaugh, 1997), perturbação intelectual moderada (Hurley & Deal, 2006; Simon, 2007), depressão e ansiedade (e.g., Ashendorf, Constantinou & McCaffrey, 2004), perturbações psicóticas (e.g., Duncan, 2005), disfunção do lobo temporal (e.g., Hill, Ryan, Kennedy & Malamut, 2003) e dor crónica (e.g., Iverson, Page, Koehler, Shojanian & Badii, 2007). Da mesma forma, o TOMM está também validado para amostras forenses, em indivíduos em contexto psiquiátrico (Weinborn, Orr, Woods, Conover & Feix, 2003) e para indivíduos envolvidos em processos judiciais (e.g., Bounds, 2004; Delain, Stafford & Ben-Porath, 2003). Os estudos de validação mostram, ainda, que os défices de memória (e.g., Rees et al., 2001; Teichner & Wagner, 2004; Tombaugh, 1996), as variáveis sociodemográficas (e.g., O'Bryant et al., 2008; Rees, Tombaugh & Boulay, 2001; Tombaugh, 1996) e os estados afetivos (e.g., O'Bryant, Engel, Kleiner, Vasterling & Black, 2007; Rees et al., 2001; Tombaugh, 1996, 1997) não influenciam o desempenho no teste.

Em Portugal, o instrumento foi validado para amostras normativas de crianças e adolescentes, com idades compreendidas entre os 8 e os 16 anos (Silva, 2011), de adultos, com idades entre os 19 e os 30 anos (Cruz, 2008; Maior, 2008) e de adultos idosos, com idades entre 61 e 86 anos (Cruz, 2008; Fernandes, 2009; Maior, 2008); em amostras forenses, com adolescentes institucionalizados em Centros Educativos (Garcia, 2011), adultos em contexto prisional (Faustino, 2009; Mota, 2008; Soares, 2013), adultos em contexto médico-legal com

histórico de Traumatismo Crânioencefálico (TCE) e em adultos em situação de litígio (Fonseca, 2009); em amostras clínicas de crianças e adolescentes com dificuldades intelectuais ligeiras e moderadas (Silva, 2011), de adultos com depressão (Castro, 2015) e com declínio cognitivo ligeiro (Fernandes, 2009; Jesus, 2013; Pinho, 2012); e em amostras que envolvem um design experimental, no qual os participantes são instruídos a simular (Cruz, 2008; Pinho 2012). Relativamente à influência das variáveis sociodemográficas, Pinho (2012) encontrou diferenças de género, de idade e de escolaridade no segundo Ensaio de Aprendizagem do TOMM, ao contrário de Jesus (2013), que não observou uma correlação significativa destas variáveis e dos testes de rastreio cognitivo com os resultados do teste num grupo com DCL. Verifica-se, ainda, que os sintomas depressivos não influenciam os resultados do TOMM (Castro, 2015; Pinho, 2012).

No caso particular dos adultos idosos, têm surgido estudos que mostram que o ponto de corte inferior a 45 é adequado para grupos saudáveis e com DCL, com valores de especificidade e sensibilidade adequados, mas que o mesmo não acontece para grupos com demência moderada a severa (e.g., Rudman, Oyebode, Jones & Bentham, 2011), padrão também verificado no contexto português (Fernandes, 2009). Nesta população específica, os desempenhos são significativamente inferiores e o ponto de corte de 5 erros ou menos leva a uma taxa de falsos positivos inaceitavelmente elevada, que não melhora utilizando pontos de corte mais conservadores, como 8 ou 10 erros (Teichner & Wagner, 2004). De facto, se comparamos os estudos de Tombaugh (1996), Teichner e Wagner (2004) e Walter, Morris, Swier-Vosnos e Pliskin (2014), a variabilidade no desempenho dos grupos com demência pode ser explicada pelo grau de severidade de demência do grupo clínico utilizado, sendo que quanto maior a severidade do quadro demencial, pior o desempenho no TOMM (Dean, Victor, Boone, Philpott & Hess, 2009). Desta forma, é compreensível que pontos de corte alternativos levem

a diferentes valores de especificidade nos diversos estudos, já estes variam também em função da severidade da demência da amostra utilizada (McGuire, Crawford & Evans, 2019).

Apesar de ser um teste de reconhecimento visual, o TOMM implica uma codificação multimodal, sendo que os processos de produção verbal contribuem para a codificação dos estímulos visuais, o que pode levar a piores desempenhos nos participantes com dificuldades semânticas e de nomeação num quadro demencial (Rudman et al, 2011). Além disso, o TOMM é sensível ao treino e ao ensino (*coaching*) (Simões et al., 2017), a sua utilização tem um custo monetário associado e corresponde a apenas um dos critérios a ser cumpridos para o diagnóstico de simulação (i.e., falsidade ou exagero de sintomas manifestados), não sendo possível inferir a intencionalidade e a motivação desse comportamento, pelo que deve ser complementado com outras medidas ou procedimentos. Não obstante, é de fácil administração e cotação e tem elevados valores de especificidade e sensibilidade, com uma taxa de falsos positivos inferior a 1%. Ademais, trata-se de um dos instrumentos com mais investigação empírica associada, a qual corrobora consistentemente a sua precisão na identificação dos comportamentos de simulação de défices cognitivos (Barhon, Meares, Chekaluk & Shores, 2015). Além disso, o seu elevado número de itens aparenta um nível de dificuldade mais elevado do que o real, o que permite a existência de validade facial como teste de aprendizagem e memória (Simões et al., 2017). Todas estas vantagens justificam o facto de o TOMM ser um dos testes mais utilizados em contexto forense (Larrabee, 2003). Gavett, O'Bryant, Fisher e McCaffrey (2005) verificaram que os resultados no Ensaio 1 eram preditivos do desempenho no Ensaio 2 e no Ensaio de Retenção e, por isso, sugeriram que resultados iguais ou superiores a 45 acertos no Ensaio 1 seriam suficientes para identificar indivíduos sem comportamentos suspeitos de simulação, exagero ou esforço insuficiente, independentemente do nível e do tipo de perturbação. Neste sentido, e de forma a potenciar a utilização do TOMM nas avaliações neuropsicológicas, através da redução do seu tempo de administração, alguns autores

propuseram uma versão abreviada do instrumento, que inclui apenas o primeiro ensaio de aprendizagem (e.g., Bauer, O’Bryant, Lynch, McCaffrey & Fisher, 2007; O’Bryant et al., 2007; O’Bryant, Gavett, McCaffrey, Smitherman & Humphreys, 2008), mostrando elevada precisão de diagnóstico. Utilizando apenas o primeiro ensaio como rastreio de simulação, Hilsabeck, Gordon, Hietpas-Wilson e Zartman (2011) propõem valores iguais ou superiores a 41 acertos como indicativos de ausência de simulação, exagero ou esforço insuficiente e valores iguais ou inferiores a 25 como sugestivos destes mesmos comportamentos. Por sua vez, Webber e colaboradores (2018) sugerem um ponto de corte igual ou inferior a 40 acertos para ausência de comportamentos suspeitos de simulação, sendo que resultados entre 37 e 40 acertos deverão obrigar à aplicação dos restantes ensaios do TOMM. Também em Portugal, Fernandes (2009) corroborou a eficácia da administração exclusiva do Ensaio 1 deste instrumento, propondo um ponto de corte igual ou superior a 33 acertos para uma identificação de 100% dos não-simuladores, em amostras com adultos idosos saudáveis e com DCL, resultado apoiado por Pinho (2012).

Para além da administração exclusiva do Ensaio 1, Denning (2012) sugeriu a consideração adicional do número de erros nos dez primeiros itens deste ensaio, verificando valores de especificidade e sensibilidade adequados para um ponto de corte igual ou superior a 1 erro como identificativo do comportamento de simulação ou esforço insuficiente. Desta forma, o autor sugere a utilização dos dez primeiros itens do Ensaio 1 como uma forma de rastreio breve de simulação.

Não obstante às propostas de versões abreviadas do TOMM, a literatura mostra-se de acordo com a necessidade de administração de, pelo menos, mais um TVS complementar para inferir suspeitas de simulação, exagero de sintomas ou esforço insuficiente (Webber et al., 2018).

1.3.2.2. *Rey-15 Item Memory Test* (Rey-15IMT; Rey, 1964; Simões et al., 2010)

O Rey-15IMT (Rey-15IMT; Rey, 1964; Simões et al., 2010) foi desenvolvido como um teste de memória visual para avaliar a credibilidade das queixas cognitivas e para detetar comportamentos de simulação de défice cognitivo, exagero de sintomas e esforço insuficiente, associado a reduzida motivação (Rey, 1964).

A sua versão original inclui um Ensaio de Evocação Imediata, no qual é apresentado, durante 10 segundos, um cartão com 15 itens (3 itens dispostos por 5 linhas), composto por letras, números e formas geométricas, alvos de memorização. Apesar da sua aparente dificuldade, os 15 itens podem ser agrupados em 5 categorias conceptuais sobreaprendidas (letras maiúsculas, letras minúsculas, números árabes, números romanos e formas geométricas), pelo que o seu grau de dificuldade é reduzido (Reznek, 2005; Simões et al., 2017). Após a exposição dos estímulos, a folha é retirada, pedindo-se ao indivíduo que reproduza imediatamente o maior número de itens que conseguir recordar, independentemente da sua disposição espacial. Todos os itens corretamente evocados são cotados com 1 ponto, perfazendo um total possível de 15 pontos.

Uma vez que as limitações de especificidade e sensibilidade da versão original do instrumento punham em causa a sua utilização (Davidson, Suffield, Orenczuk, Nantau & Mandel, 1991; Rogers, Harrell & Liff, 1993), Boone, Salazar, Lu, Warner-Chacon e Razani (2002) desenvolveram um Ensaio de Reconhecimento complementar, administrado após o Ensaio de Evocação. Este consiste na apresentação de uma folha com 30 itens (15 itens originais e 15 itens distratores), sendo pedido ao participante que identifique os 15 itens originais. Neste ensaio contabiliza-se o número de itens corretamente reconhecidos (num total possível de 15 pontos), o número de falsos positivos (i.e., itens incorretamente identificados) e calcula-se um Resultado Combinado [número de itens corretamente evocados + (número de itens corretamente reconhecidos – falsos positivos)], com um resultado máximo possível de 30

pontos. A introdução deste ensaio permitiu, por um lado, a manutenção dos valores de especificidade (acima de 92%) e, por outro, o aumento dos valores de sensibilidade (de 47% para 71%) (Bonne et al., 2002).

O Rey 15-IMT revela-se um instrumento simples, de fácil e rápida aplicação (aproximadamente 5 minutos), de utilização de domínio público (Lezak, 1995), com facilidade de transporte, manuseio e cotação (Boone et al., 2002), com validade facial, uma vez que é apresentado como uma prova de memória imediata e de atenção (Simões et al., 2017) e que permite a obtenção de resultados quantitativos e qualitativos (Lezak, Howieson, Bigler & Tranel, 2012; Strauss, Sherman & Spreen, 2006), pelo que é um dos instrumentos de avaliação de validade de desempenho mais utilizado (Lopes, 2017). No entanto, diversas limitações têm sido consistentemente apontadas.

Em primeiro lugar, a forma como são tratados os erros “disléxicos” e os itens ambíguos não é clara (Griffin, Glassmire, Henderson & McCann, 1997). Adicionalmente, Goldberg e Miller (1986) alertam para o facto de o uso de perseverações e de reversões levar à identificação errónea de indivíduos como simuladores. Em terceiro lugar, ainda que o ponto de corte mais utilizado para o Ensaio de Evocação Imediata seja de 9 pontos (e.g., Boone et al., 2002; Green, Kirk, Connery, Baker & Kirkwood, 2014; Greiffenstein, Baker & Gola, 1996; Lezak et al., 2012; Taylor, Kreutzer & West, 2003), este não é consensual, uma vez que diversos autores propõem valores de ponto de corte distintos, como 6 (e.g., Arnett, Hammeke & Schwartz, 1995; Strutt, Scott, Shrestha & York, 2011); 7 (e.g., Lee, Loring & Martin, 1992); 8 (e.g., Bernard & Fowler, 1990; Schretlen, Brandt, Krafft & Van Gorp, 1991); e 10 (Bernard, Houston & Natoli, 1993). O mesmo acontece para o Resultado Combinado do Reconhecimento, sendo que, embora o ponto de corte mais utilizado seja uma pontuação inferior a 20 (Boone et al., 2002; Rabin, Barr & Burton, 2005), outros são propostos, por exemplo, 21 (Morse, Douglas-Newman, Mandel & Swirsky-Sacchetti, 2013) e 26 (Green et al., 2014). Esta problemática leva

inevitavelmente a outra, que se encontra documentada na meta-análise de Reznek (2005). Para além dos valores de sensibilidade reduzidos (cerca de 47%, variando entre 5% e 86%), ainda que os valores de especificidade sejam adequados (acima de 90%), estes variam em função das amostras analisadas, o que leva a elevadas taxas de falsos positivos (i.e., identificação de indivíduos não-simuladores como simuladores). Desta forma, os pontos de corte mais utilizados são impossíveis de generalizar a todos os grupos (Simões et al., 2017), requerendo-se urgentemente o seu ajustamento. Por si só, esta circunstância constitui uma quinta limitação deste instrumento. De facto, os testes de validade de sintomas devem ser insensíveis ao efeito de variáveis como a idade, escolaridade, género, quociente de inteligência, défices neurológicos e perturbações afetivas, de modo a garantir a sua validade (Simões et al., 2017). No entanto, diversos estudos mostram a influência da baixa escolaridade e do género (Robles, López, Salazar, Boone & Glaser, 2015; Strutt et al., 2011), da idade (Green et al., 2014; Griffin et al., 1997; Schretlen et al., 1991), dos défices mnésicos e de demência (Fazio, Fari & Yamout, 2017; Schretlen et al., 1991), eventualmente da cultura (Strutt et al., 2011), da epilepsia (Lee et al., 1992), das dificuldades intelectuais (Goldberg & Miller, 1996; Love, Glassmire, Zanolini & Wolf, 2014) e das perturbações psiquiátricas (Hays, Emmons & Lawson, 1993) nos desempenhos. Além disso, com base na sua experiência clínica, McGuire (2006) verificou que existe uma variável adicional que pode afetar o desempenho no teste, a ordem pela qual o indivíduo percebe os itens. De facto, embora se assuma que na sociedade ocidental exista uma codificação horizontal, no seu estudo, o autor verificou que 17% dos indivíduos do grupo com lesão cerebral e 10% da amostra clínica geral codificaram os 15 itens verticalmente. Tal tem implicações inevitáveis, uma vez que desta forma não existe uma sequência lógica entre os itens, o que aumenta a dificuldade de codificação e, assim, de recuperação dos estímulos, traduzindo-se em resultados significativamente mais baixos no teste e, consequentemente, levando a falsos positivos (McGuire, 2006).

Não obstante, o Rey-15IMT (Rey-15IMT; Rey, 1964; Simões et al., 2010) encontra-se validado para amostras normativas de adultos, com idades entre os 23 e os 54 anos (e.g., Bernard & Fowler, 1990; Beetar & Williams, 1995; Schretlen et al., 1991); para amostras forenses, especificamente com doenças mentais (Simon, 1994), em estabelecimentos prisionais e suspeitos de simulação (Frederik, Sarfaty, Sohnston & Powel, 1994) e pacientes de ambulatório em situação de litígio (Lee et al., 1992); e para amostras clínicas, que incluem perturbações psiquiátricas (Goldberg & Miller, 1986; Schretlen et al., 1991; Hays et al., 1993), perturbações neurológicas (Arnett, et al., 1995; Lee et al., 1992), perturbação intelectual (Golberg & Miller, 1986), lesão cerebral (Bernard & Fowler, 1990; Guilmette, Hart, Guiliano & Leininger, 1994; Schretlen et al., 1991; Taylor et al., 2003), demência e amnésia (Schretlen et al., 1991) e depressão (Lee et al., 1992; Guilmette et al., 1994).

Em Portugal, o instrumento está validado para uma amostra normativa de crianças, com idades entre os 8 e os 16 anos (Silva, 2011) e adultos e adultos idosos, com idades compreendidas entre os 19 e os 86 anos (Fernandes, 2009; Fonseca, 2009; Lages, 2016; Pinho, 2012; Simões et al., 2010); em amostras forenses, com adolescentes com medida tutelar educativa de internamento (Garcia, 2011), reclusos (Oliveira, 2008) e pacientes com TCE em situação de litígio (Fonseca, 2009); em amostras clínicas de adultos idosos com DCL (Fernandes, 2009; Pinho, 2012) e de adultos e adultos idosos com problemas cognitivos e depressão (Simões et al., 2010); e numa amostra experimental, em que adultos idosos foram instruídos a simular (Pinho 2012). Relativamente à influência de outras variáveis, o Rey-15IMT mostrou-se sensível às variáveis idade (Lages, 2016; Pinho 2012), género, (Lages, 2016), escolaridade (Lages, 2016), défice cognitivo (Jesus, 2013) e sintomatologia depressiva grave (Simões et al., 2010).

Particularmente em adultos idosos, Fazio e colaboradores (2017) verificaram que usando os pontos de corte mais consensuais, a taxa de falsos positivos era inaceitavelmente

elevada para indivíduos a partir de 59 anos. De modo específico, com amostras clínicas, no Ensaio de Evocação Imediata, Rudman e colaboradores (2011) obtiveram uma especificidade de 27% e Bortnik, Horner e Bachman (2013) de 28% nos grupos com demência moderada a severa. Em Portugal, mesmo reduzindo o ponto de corte do Ensaio de Evocação Imediata para <6, este traduziu uma especificidade de apenas 72,9% nos grupos de DCL e demência ligeira (Simões et al., 2010). No Resultado Combinado do Reconhecimento, o ponto de corte <18 mostrou-se adequado apenas para jovens adultos, obtendo valores de especificidade de 78,8% em adultos idosos saudáveis, 65,2% em adultos idosos com depressão e 37,8% em adultos idosos com problemas cognitivos. Desta forma, Simões e colaboradores (2010) sugeriram, para o Ensaio de Evocação Imediata, os pontos de corte de 1.90 para adultos idosos com problemas cognitivos e 5.70 para adultos idosos saudáveis. Relativamente ao Resultado Combinado, os valores propostos são 2 para adultos idosos com problemas cognitivos e 11.70 para idosos saudáveis.

De forma a dar resposta às limitações descritas e a aumentar a simplicidade e/ou a coerência lógica dos itens (Simões et al., 2017), facilitando, assim, a distinção entre simuladores e não-simuladores foram sugeridas versões alternativas do teste. Paul, Franzen, Cohen e Fremouw (1992) propuseram o Rey 16-Item, no qual eliminaram a sequência de figuras geométricas e acrescentaram 1 item às restantes, com uma disposição final de ABCD, 1234, abcd e I II III IIII. O procedimento de administração é o mesmo que o da versão original, porém, a instrução de memorização de 15 itens é substituída pela de 16 itens, enfatizando, assim, a maior dificuldade aparente da tarefa (Paul et al., 1992). Nesta versão alternativa é considerado o número de itens omissos e um resultado total, composto pelo número de itens corretamente evocados, na sua posição original. No entanto, Iverson e Franzen (1996) verificaram que o Rey-16 Item detetou apenas 22% dos participantes simuladores de lesão cerebral e no estudo de Franzen e Martin (1996) apenas dois indivíduos instruídos a simular

foram identificados. Mais tarde, Fisher e Rose (2005) compararam a versão original com a versão de 16 itens, verificando que a última detetou menos simuladores (31% comparativamente a 44% do Rey-15IMT) e apresentou menor poder preditivo.

Uma outra versão alternativa ao Rey-15IMT, o Rey II, foi desenvolvida por Griffin et al., (1997), com os objetivos de aumentar o poder discriminativo do teste, melhorar a sua validade facial, estandardizar as instruções e desenvolver um sistema de cotação qualitativo que não sofresse a influência da competência e/ou de perturbação clínica. Os autores elaboraram um novo cartão de 15 estímulos, no qual substituíram a sequência de figuras geométricas e a de letras minúsculas para aumentar a lógica interna e a redundância do padrão. Os resultados mostraram que o Rey II elicitou significativamente mais erros qualitativos no grupo instruído a simular, sem reduzir o número de itens corretamente reproduzidos nos grupos controlo e clínico. Além disso, verificaram que os resultados qualitativos não são contaminados pelo quociente de inteligência, idade, memória ou estado mental, permitindo, assim, maior precisão de classificação, com valores mais elevados de sensibilidade (mais 79%) e especificidade (mais 29%), comparativamente ao Rey-15IMT (Griffin et al., 1997). No entanto, num estudo de validade convergente com o TOMM, verificou-se que embora o Rey II fosse mais discriminativo, apresentou um poder preditivo positivo (i.e., probabilidade de um indivíduo ter a condição X, dado o resultado positivo no teste Y) e negativo (i.e., probabilidade de um indivíduo não ter a condição X, dado o resultado negativo no teste Y) inaceitavelmente baixo, de 0.62 e 0.64, respetivamente, produzindo um número de falsos positivos alarmante (Whitney, Hook, Steiner, Shepard & Callaway, 2008), pelo que o seu uso exclusivo não é recomendável (Witney et al., 2008).

1.3.2.3. *Coin in Hand-Extended Version* (CIH-EV; Kapur, 1994; Daugherty et al., no prelo)

O teste CIH foi originalmente documentado por Kapur (1994), com o fim de detetar o comportamento de simulação em indivíduos suspeitos de simularem baixos desempenhos mnésicos (Kapur, 1994). O teste assenta num paradigma de escolha forçada e no princípio do efeito de teto, sendo, no entanto, apresentado como um teste de memória difícil. De facto, administrando-o a uma amostra de pacientes com amnésia, Kapur (1994) verificou que todos obtiveram a pontuação máxima, enquanto os indivíduos suspeitos de simularem apresentaram um desempenho próximo ao nível do acaso. Particularmente, o teste consiste na apresentação de uma moeda na mão esquerda ou na mão direita por dois segundos, aparecendo 5 vezes em cada mão, de forma aleatória, num total de 10 ensaios. É pedido ao participante que memorize em que mão está a moeda e, após a exposição, deve fechar os olhos e contar a partir de 10, em ordem decrescente. No fim, deve abrir os olhos e escolher a mão onde ficou a moeda, sendo fornecido um feedback verbal para todas as respostas dadas.

Este teste está validado para amostras normativas de adultos, com idades compreendidas entre os 18 e os 37 anos (Cochrane, Baker & Meudell, 1998; Hanley, Baker & Ledson, 1999; Kelly, Baker, Broek, Jackson & Humphries, 2005; Kapur, 1994) e adultos idosos, com idades entre os 65 e os 76 anos (Yeh et al., 2019); amostras forenses, com indivíduos envolvidos em procedimentos médico-legais (Kapur, 1994; Colwell & Sjerven, 2005); em amostras clínicas, com indivíduos com défice mnésico (Kapur, 1994; Cochrane et al, 1998; Hanley et al., 1999; Kelly et al., 2005), perturbação intelectual (Colwell & Sjerven, 2005), lesão cerebral (Kelly et al, 2005; Hampson, Coughlan, Moulin & Bhakta, 2014), epilepsia (Hampson et al., 2014), doenças autoimunes (Ferreira, Gomes, Moreira, Silva & Cavaco, 2017) e demência (Yeh et al., 2019; Rudman et al., 2011; Schroeder, Peck, Buddin, Heinrichs & Baade, 2012); e em designs experimentais, nos quais os indivíduos são instruídos a simular défices mnésicos (Cochrane et al, 1998; Hanley et al., 1999) e demência (Yeah et al., 2019).

Relativamente aos pontos de corte, não existe um valor consensual entre os estudos realizados. De facto, na deteção de simuladores de défice mnésico, Cochrane et al. (1998), utilizaram um ponto de corte de <7 , que traduziu uma sensibilidade de 80%, e Hanley et al. (1999) sugeriram um ponto de corte ≤ 8 , com uma sensibilidade de 92.5% e uma especificidade de 87.5%. Já Kelly et al., (2005) utilizaram um valor <8.50 , que refletiu uma sensibilidade de 92,5% para simuladores de lesão cerebral e uma especificidade de 87,5%.

Até à data são conhecidos apenas três estudos com adultos idosos, que analisam particularmente o desempenho de amostras clínicas com demência no CIH (Rudman et al., 2011; Schroeder et al., 2012; Yeh et al., 2019). Nos estudos de Rudman e colaboradores (2011) e de Schroeder e colaboradores (2012), um ponto de corte ≤ 7 traduziu uma especificidade de 100% e de 98%, respetivamente, para os grupos de demência ligeira. Yeh e colaboradores (2019), por sua vez, verificaram que o ponto de corte mais adequado era ≤ 8 , com uma sensibilidade de 67% e uma especificidade de 88%. No entanto, Rudman e colaboradores (2011) verificaram que o valor de especificidade reduzia para 77% quando aplicado ao grupo com demência moderada/severa. De notar que a amostra de Yeh e colaboradores (2019) apresentava maioritariamente demência moderada/severa, com uma média de MMSE igual a 17.7. Tal pode explicar o facto de o valor de especificidade ser menor do que o encontrado para os grupos com demência ligeira nos estudos de Rudman et al., (2011) e de Schroeder et al., (2012), o que poderá sugerir uma influência da severidade do défice cognitivo nos resultados. Ainda que existam estudos que mostrem que não há influência do funcionamento cognitivo no desempenho do CIH (e.g., Hanley et al., 1999; Schroeder et al., 2012), outros apontam para a sua expressão, porém, menor do que o que acontece para o TOMM e o Rey-15IMT (e.g., Rudman et al., 2011; Yeh et al., 2019). Desta forma, a exploração desta influência torna-se pertinente.

O CIH tem sido apontado como um excelente instrumento de *screening* de simulação, exagero de sintomas e/ou esforço insuficiente, uma vez que não requer preparação dos estímulos e a sua administração é simples e rápida (Hanley et al., 1999; Kapur, 1994). Além disso, os seus resultados são de fácil interpretação, apresenta elevada validade facial (Kapur, 1994), apresenta melhor precisão de diagnóstico do que o TOMM e o REY-15IMT (e.g., Rudman et al., 2011; Yeh et al., 2018) e, ainda que variáveis, os valores de sensibilidade (entre 67 a 100%) e especificidade (entre 87.5 e 100%) são relativamente elevados (Daugherty et al., no prelo; Hanley et al., 1999). Ademais, o carácter visual dos estímulos permite reduzir a sua dependência da linguagem, relativamente a outros TVDs, bem como a sua administração a indivíduos não letrados ou com capacidades verbais diminuídas (Daugherty et al., no prelo). Adicionalmente, o desempenho no CIH parece não ser afetado por variáveis sociodemográficas, como a idade e o nível educacional (Ferreira et al., 2017; Schroeder et al., 2012), nem pelo estado emocional, especificamente ansiedade ou depressão (Rudman et al., 2011), aspetos fundamentais de um Teste de Validade de Desempenho.

Dadas as suas vantagens, o CIH foi recentemente adaptado para uma versão alargada/extensa (CIH-EV), informatizada e gratuita (Daugherty, Hidalgo-Ruzzante, Palma-Navarro, Bencrimo & García, 2017; Daugherty et al., no prelo). O procedimento de administração segue o mesmo racional que o proposto por Kapur (1994), mas nesta versão, o CIH-EV é apresentado como um teste que pretende medir a resistência da memória à distração e o seu nível de dificuldade percebida é aumentado ao serem contemplados 3 níveis de dificuldade (Binder, 1990; Hiscock & Hiscock, 1989; Iverson, Franzen, & McCracken, 1991) com 10 ensaios por nível, perfazendo um total de 30 ensaios. Durante o teste, o participante visualiza uma imagem com duas mãos e uma moeda numa delas. De seguida, as mãos fecham-se em punho, desaparecem e o ecrã fica escuro. Imediatamente a seguir, no primeiro nível, o participante deve contar a partir de 10, em ordem decrescente, durante 10 segundos; no segundo

nível conta a partir de 99, durante 15 segundos; e no terceiro nível conta a partir de 999, durante 20 segundos, sendo que para o segundo e terceiro níveis, é enfatizado o aumento da dificuldade. A contagem decrescente, em todos os ensaios, de todos os níveis, é visualmente guiada no ecrã do computador ou *tablet* utilizado. Posteriormente, as mãos reaparecem fechadas em punho, momento em que o participante deve dar a sua resposta. Ao longo dos 30 ensaios, todas as respostas são seguidas de um feedback visual.

Para além das vantagens associadas à versão original do CIH, a versão extensa vem permitir uma aplicação e cotação estandardizadas, bem como a sua facilitação de utilização, uma vez que se trata de uma ferramenta multiplataformas, que pode ser administrada em tablets ou computadores pessoais e em todos os sistemas operacionais (e.g., Windows, Mac) de forma gratuita (Daugherty et al., 2017). Desta forma, é possível administrar esta versão a um maior número de pessoas, estudar a sua equivalência cultural e traduzi-la em diversas línguas, permitindo, assim, a sua aplicação em diferentes contextos culturais, de forma a torná-la numa medida comum entre os estudos (Daugherty et al., no prelo). Para além disso, permite maior diversidade de resultados, já que para além do número total de acertos é ainda possível obter o número de acertos para cada nível de dificuldade e o tempo de resposta total e para cada um dos níveis, variáveis difíceis de controlar num contexto de avaliação mais tradicional (Daugherty et al., no prelo).

O CIH-EV foi recentemente validado para uma amostra de estudantes universitários portugueses, espanhóis, colombianos e americanos, segundo um design experimental com um grupo controlo e um grupo instruído a simular défice mnésico (Daugherty et al., no prelo). O estudo permitiu verificar que esta versão do instrumento apresenta elevada validade convergente com medidas de referência (e.g., Victoria Symptom Validity Test, TOMM e Subteste de Memória de Dígitos), o que indica que é uma alternativa a considerar na avaliação da credibilidade de desempenhos. Além disso, verificou-se que existe equivalência cultural,

uma vez que o ponto de corte mais adequado foi o mesmo para todos os países (≥ 27), traduzindo valores de sensibilidade a partir de 62% e de especificidade a partir de 95%, com uma taxa de falsos positivos de 5%. O CIH-EV revela-se, assim, eficaz na discriminação de indivíduos simuladores de não-simuladores, ainda que estes valores possam estar inflacionados. De facto, esta validação multicultural contou com amostras exclusivamente saudáveis, pelo que é necessário estendê-la a amostras com défice cognitivo real (Daugherty et al., no prelo), com o fim de verificar a necessidade de ajustamento do ponto de corte, bem como a manutenção da sua precisão diagnóstica.

1.3.2.4. Testes de Validade de Desempenho e Demência

Como discutido anteriormente, em Portugal existem apenas dois Testes de Validade de Desempenho (TVDs) validados para os adultos idosos saudáveis e com DCL, para os quais os pontos de corte se mostram adequados. Verifica-se, no entanto, a ausência de estudos com grupos com diagnóstico de demência. Este não é, porém, um cenário exclusivamente nacional, uma vez que os estudos internacionais com TVDs com estas amostras são também escassos (Dean et al., 2009). De facto, os grupos de adultos idosos e, sobretudo, os grupos de adultos idosos com demência são frequentemente excluídos dos estudos de validação deste tipo de testes (Simões et al., 2010), uma vez que a sua integração tem implicações associadas. Tal como discutido no Capítulo I, verifica-se que os pontos de corte dos TVDs mais utilizados não se adequam, muitas vezes, a estas amostras, produzindo elevadas taxas de falsos positivos (Dean et al., 2009) e, portanto, são excluídas, de modo a melhorar os valores de especificidade do teste (Boone et al., 2002). Além disso, a capacidade diagnóstica do teste pode ser posta em causa quando, por exemplo, em testes que se baseiam no efeito teto, este tipo de amostra apresenta dificuldades e não consegue atingir bons desempenhos, que se distingam significativamente dos indivíduos simuladores (van Gorp & Hassenstab, 2009).

Embora os estudos mostrem a viabilidade do uso dos TVDs em amostras com DCL, no caso de demência, a investigação mostra que quanto maior o grau de severidade, pior a precisão de diagnóstico (McGuire et al., 2019; Rudman et al., 2011). Especificamente, na sua meta-análise, Dean e colaboradores (2009) verificaram que os indivíduos com um MMSE >20 falharam cerca de 36% dos testes administrados, com um MMSE entre 15 a 20 falharam cerca de 47% e com pontuações MMSE abaixo de 15 falharam 83% dos testes. Ajustando os pontos de corte das medidas utilizadas, obtiveram-se valores de sensibilidade extremamente baixos, que inviabilizam a sua administração. Deste modo, este tipo de estudos mostra-se extremamente importante, já que se por um lado evidenciam a inadequação de alguns TVDs, por outro revelam aqueles que são verdadeiramente úteis nestes tipos de amostras (Dean et al., 2009).

Considerando o quadro demográfico atual e o padrão de inadequação dos pontos de corte atuais dos TVDs existentes, torna-se indubitavelmente necessário estudar os desempenhos específicos dos adultos idosos saudáveis e com quadros clínicos, particularmente demenciais (Boone et al., 2002), de forma a que, por um lado, possam ser delineadas soluções alternativas que passam, muitas vezes, pelo ajustamento dos pontos de corte (Simões et al., 2010; van Gorp & Hassenstab, 2009) e, por outro, que se tenha conhecimento sobre os testes que podem ou não ser utilizados nestes grupos (Dean et al., 2009; Rudman et al., 2011). A importância dos estudos com população idosa saudável e clínica toma ainda mais revelado se considerarmos que é com base no conhecimento empírico que se podem fundamentar as decisões clínicas que, por sua vez, têm implicações diretas na vida dos indivíduos.

1.3.2. A dicotomia entre medidas *embedded* (integradas) versus *stand-alone* (isoladas)

Atualmente, as medidas de avaliação da credibilidade do desempenho mais utilizadas são testes independentes, especificamente desenvolvidos para avaliar determinadas capacidades cognitivas – *stand-alone* ou *free-standing* TVSSs ou TVSSs isolados – e os indicadores de validade integrados em testes tradicionais – *embedded* TVSSs ou TVSSs integrados (Larrabee, 2012). De facto, os indicadores integrados têm ganho popularidade na área da deteção da simulação, sendo que, atualmente, todos os domínios neuropsicológicos têm pelo menos um teste com um índice de validade integrado (Miele, Gunner, Lynch & McCaffrey, 2012). A crescente afirmação dos indicadores dos testes tradicionais é justificada pelas já mencionadas vantagens em relação aos TVSSs isolados, das quais se destacam a economia de tempo de avaliação, já que este não é necessário ser estendido para administração de um teste adicional; o facto de a monitorização do esforço investido não ocorrer num momento pontual, mas ao longo da sessão, considerando que o esforço é variável (Bonne, 2008); e o facto de resultados significativamente baixos num único TVSS não ser diagnóstico de simulação, o que implica recorrer a outros meios de informação (Rogers, 2008). Não obstante, é necessário pesar até que ponto as suas vantagens compensam o seu uso em detrimento dos TVSSs isolados, ou seja, até que ponto não comprometem a identificação precisa dos comportamentos não credíveis.

A investigação sobre a validade de diagnóstico das medidas isoladas e integradas não é consensual (e.g., An, Kaploun, Erdodi & Abeare, 2017; Miele et al., 2012), principalmente porque os valores de especificidade e de sensibilidade de ambos os tipos de medidas variam em função de determinadas condições, especificamente dos pontos de corte (mais liberais ou mais conservadores) e das amostras utilizadas (An et al., 2017). Dentro da própria análise dos índices integrados, os resultados são contraditórios, com uns estudos a mostrarem que a combinação de, por exemplo, nove índices consegue discriminar indivíduos simuladores de não-simuladores (Meyers, Volbrecht, Axelrod & Reinsch-Boothby, 2011; Meyers &

Volbrecht, 2003) e outros estudos a recomendar a utilização de apenas um índice com boas referências discriminativas (e.g., subteste de Memória de Dígitos da WAIS-III-R), uma vez que a integração mais índices não contribui para a validade preditiva incremental (Miele et al., 2012). Além disso, são reportados casos em que a combinação de índices integrados traduz uma taxa de falsos-negativos elevada (entre 67% e 100%), pelo que o seu uso exclusivo não é recomendável (Miele et al., 2012).

Deste modo, a tendência empírica parece apontar quer para um maior Poder Preditivo Positivo (PPP), quer para uma menor taxa de falsos positivos dos TVDS isolados (Miele et al., 2012), tornando mais pertinente a sua utilização. Não obstante, se o clínico pretender beneficiar das vantagens associadas às medidas integradas, deve utilizá-las em conjunto com TVSS isolados (Miele et al., 2012).

1.4. O presente estudo

1.4.1. Pertinência do Teste de Validade de Desempenho CIH-EV

O presente estudo teve como objetivo a validação do teste CIH-EV (Daugherty et al., no prelo) com adultos idosos portugueses. De facto, apesar da área da avaliação da simulação estar em expansão, em Portugal a investigação com testes especificamente desenvolvidos para examinar a credibilidade das queixas e dos desempenhos cognitivos é ainda escassa (Simões et al., 2010), razão pela qual existem poucos testes validados para a população portuguesa. Este cenário tende a facilitar o *coaching*, a pesquisa *online* e, assim, a familiarização com os poucos instrumentos que existem (Suhr & Gunstad, 2007), permitindo contornar a deteção dos comportamentos de simulação e/ou exagero de sintomas em contexto forense e/ou médico-legal. Além disso, por um lado, a avaliação destes comportamentos deixou de ser exclusiva a este tipo de contextos e passou a ser recomendada em todas as avaliações (neuro)psicológicas (e.g., Bush et al., 2005; Heilbroner et al., 2009; Larrabee, 2007) e, por outro, a investigação

mostra que a precisão de diagnóstico aumenta com o número de TVDs administrados, pelo que é necessário existir alguma diversidade de instrumentos validados.

Por estas razões, foi recentemente validado um novo instrumento, o CIH-EV (Daugherty et al., no prelo), para estudantes universitários portugueses com idades entre os 17 e os 50 anos, dadas as vantagens que apresenta. Entre elas destacam-se o facto de ser um teste de domínio público e informatizado, o que permite a sua utilização mais vasta e estandardizada, com maior variedade de resultados (Daugherty et al., no prelo); a independência de variáveis demográficas e do funcionamento neurocognitivo (e.g., Ferreira, Gomes, Moreira, Silva, & Cavaco, 2017; Schroeder, et al., 2012); a sua equivalência multicultural (Daygherty et al., no prelo); a elevada especificidade que apresenta em diversas populações clínicas (entre 88.1 e 100%; e.g., Hampson et al., 2014; Rudman et al., 2011) e a sensibilidade modesta a elevada em grupos de simuladores (entre 67 e 100%; e.g., Cochrane et al., 1998; Ferreira et al., 2017; Kelly et al., 2005; Yeh et al., 2019). Além disso, o CIH-EV apresenta ainda vantagens sobre outros TVDs, nomeadamente, o facto de apresentar uma menor taxa de falsos positivos do que o TOMM (Rudman et al., 2011; Yeh et al. 2019) e o Rey-15IMT (Rudman et al., 2011) em amostras com demência. Estas vantagens fundamentaram a seleção deste instrumento para complementar os protocolos de avaliação de validade de desempenho existentes em Portugal, em particular na população adulta idosa.

1.4.2. Pertinência das amostras selecionadas: adultos idosos saudáveis e com demência

No presente estudo, a validação do CIH-EV foi então estendida a amostras de adultos idosos portugueses (com idades iguais ou superiores a 65 anos) saudáveis e com diagnóstico de demência. À semelhança do que acontece a nível mundial, em Portugal observa-se um envelhecimento demográfico acentuado, sendo que se estima a existência de aproximadamente

2 213 274 adultos idosos, representando, assim, 21.5% da população portuguesa total (Instituto Nacional de Estatística [INE], 2018). Este número reflete-se num Índice de Envelhecimento de 153.4 pessoas idosas por cada 100 jovens, que se prevê aumentar até 2060 (INE, 2017). Subjacente a estes números está a melhoria da qualidade de vida e das condições e serviços de saúde prestados que contribui simultaneamente para uma esperança média de vida atual de 81.6 anos (INE, 2017), com perspectivas de aumento. Este quadro demográfico obriga, inevitavelmente, a alterações nas políticas sociais, entre as quais, o aumento da idade da reforma. De facto, a idade da reforma foi, entretanto, aumentada para 66 anos e 5 meses de idade, de acordo com a Portaria n.º 25/2018, de 18 de janeiro (Diário da República, 2018). Como consequência, os pedidos de reforma antecipada tendem também a aumentar.

Sabe-se também que a idade é o maior fator de risco para a demência (Alzheimer's Society, 2016; van der Flier & Scheltens, 2005) e, de facto, de acordo com as taxas de envelhecimento nacionais, Portugal ocupa o quarto lugar da Organização para a Cooperação e Desenvolvimento Económico (OCDE) com o maior número de pessoas com demência (OCDE, 2018), prevendo-se a existência de mais de 205 mil casos nacionais (Alzheimer Portugal, 2017). Desta forma, podem existir casos em que os adultos idosos pretendem obter ganhos secundários como a reforma antecipada, subsídios de incapacidade e/ou de dependência, indemnizações, beneficiar de seguros de saúde ou evitar consequências decorrentes de determinados processos judiciais, pelo que recorrem à simulação de incapacidade (Simões et al., 2010; Dean et al., 2009). Em muitos casos, tirando partido da prevalência significativa de demência em Portugal, podem, então, recorrer à simulação de défices cognitivos típicos de um quadro demencial, de forma a obter este tipo de ganhos externos. Por sua vez, este quadro pode levar a descredibilizar queixas apresentadas por indivíduos com défices reais. Deste modo, é premente existirem meios em Portugal que permitam detetar este tipo de comportamentos não-

credíveis e validar diagnósticos mais precisos, de forma a preservar os recursos necessários para indivíduos que apresentam diagnósticos reais (Lippa, 2018).

A um nível pragmático, é necessário que existiam instrumentos de avaliação (neuro)psicológica validados que permitam fazer a distinção entre défice cónico real e simulado ou exagerado. No entanto, em Portugal existem apenas dois Testes de Validade de Desempenho validados para a população adulta idosa saudável e com Declínio Cognitivo Ligeiro (DCL), o TOMM (Fernandes, 2009; Jesus, 2013; Pinho, 2012) e o Rey-15IMT (Fernandes, 2009; Pinho, 2012; Simões et al., 2010). Para além de serem os únicos, o primeiro tem custos de utilização associados, o que faz com que, muitas vezes, os profissionais tenham de prescindir da sua administração (Daugherty et al., no prelo), e o segundo tem limitações significativas que põem em causa a sua utilização (ver capítulo I).

1.4.3. Objetivos

O objetivo geral do presente estudo foi o de validar o CIH-EV (Daugherty et al., no prelo) para adultos idosos portugueses, segundo um paradigma de simulação e utilizando um grupo clínico de idosos com demência para comparação dos desempenhos. Os objetivos específicos foram: a) analisar a relação entre as variáveis demográficas e o desempenho no CIH-EV dos três grupos (idosos saudáveis não-simuladores, idosos saudáveis simuladores e idosos com diagnóstico de demência); b) realizar um estudo comparativo dos diferentes níveis de dificuldade do CIH-EV e do desempenho total no CIH-EV, quer relativamente à precisão, quer em relação ao tempo de resposta entre os três grupos; c) analisar e determinar os pontos de corte do CIH-EV, quer para idosos saudáveis, quer para idosos com demência, determinando os seus valores de sensibilidade e especificidade; e d) realizar um estudo de validade convergente do CIH-EV com a versão portuguesa do Test of Memory Malingering (TOMM; Fernandes et al., 2009) e do Rey 15-Item Memory Test (15-IMT; Simões et al., 2010;

Simões et al., 2011) e a medida RDS do subteste de Memória de Dígitos da WAIS-III-R (Wechsler, 1997, 2008; Pinho, 2012).

Uma vez que a distribuição dos idosos saudáveis participantes pelos grupos simulador e não-simulador foi aleatória, espera-se que estes não difiram entre si relativamente às variáveis sociodemográficas e ao desempenho nos testes de rastreio cognitivo. É também esperado que os grupos clínico e simulador não difiram relativamente às variáveis sociodemográficas, diferindo no que se refere ao nível do desempenho cognitivo (tratando-se, no caso do grupo clínico, de idosos com um diagnóstico de demência) e à variável habitação (uma vez que todos os idosos do grupo clínico se encontram institucionalizados).

Foram colocadas as seguintes hipóteses de estudo:

H1. Em relação ao CIH-EV, com base no pressuposto do efeito teto dos TVDs, é expectável que o grupo de adultos idosos não-simulador e o grupo clínico tenham um melhor desempenho do que o grupo de idosos simulador.

H2. O ponto de corte no CIH-EV mais adequado ao grupo de idosos saudáveis é semelhante ao observado para jovens adultos para discriminar simuladores de não-simuladores, uma vez que estudos anteriores (e.g., Ferreira et al., 2017; Schroeder et al., 2012) apontam para a insensibilidade do CIH a variáveis como a idade.

H3. Relativamente ao grupo clínico, espera-se que o ponto de corte mais adequado no CIH-EV seja inferior ao do grupo de idosos saudáveis não-simuladores, na medida em que estudos anteriores (e.g., McGuire et al., 2019) apontam para a diminuição da especificidade dos pontos de corte em função da severidade da demência.

H4. Finalmente, é expectável que o CIH-EV apresente boa validade convergente com as restantes medidas de validade de desempenho (TOMM, Rey-15IMT e RDS - subteste de Memória de Dígitos da WAIS-III-R), em concordância com estudos anteriores (e.g., Daugherty et al., no prelo) e uma vez que estas se mostram úteis na deteção do comportamento de

simulação. Porém, de acordo com evidência anterior (e.g., Hanley et al.,1999; Rudman et al., 2011; Schroeder et al., 2012; Yeh et al. 2019) em amostras de adultos idosos com demências, espera-se que o CIH-EV revele menor percentagem de falsos positivos do que os restantes testes de validade de desempenho (TOMM, Rey-15IMT e RDS).

II. Método

2.1. Participantes

Para a presente investigação foi recolhida uma amostra de conveniência, recrutada no distrito de Lisboa, com um total de 81 adultos idosos (66 mulheres e 15 homens). Foi, no entanto, necessário excluir 5 participantes, por se verificarem critérios de exclusão, pelo que a amostra final contou com 76 participantes (64 mulheres e 12 homens) com idades compreendidas entre os 65 e os 94 anos ($M=75,78$; $DP=0,92$).

Com base num paradigma de simulação análogo, recomendado e utilizado na investigação para a validação de instrumentos (Merten, Green, Blaskewitz & Brockhaus, 2005; Tydecks, Merten & Gubbay, 2006; Rogers 1997, 2008), os participantes foram distribuídos aleatoriamente por dois grupos experimentais: um grupo controlo (não-simulador), no qual os participantes foram instruídos a ter o seu melhor desempenho e um grupo experimental (simulador), no qual os indivíduos foram instruídos a simular um défice mnésico típico de uma demência, com o objetivo de obter algum tipo de pensão, subsídio ou reforma antecipada. O grupo não-simulador ($n=30$) foi constituído por 25 participantes do sexo feminino e 5 do sexo masculino, apresentando idades compreendidas entre 65 e 94 anos ($M=75,20$; $DP=8,90$), com uma média de 8,43 anos de escolaridade ($DP=3,94$). O grupo simulador ($n=29$) foi constituído por 25 participantes do sexo feminino e 4 do sexo masculino, com idades entre os 65 e 87 anos ($M=73,31$; $DP=6,40$), com uma média de 9,07 anos de escolaridade ($DP=3,36$). De forma a potenciar a generalização e a comparação dos resultados (Rogers 1997, 2008), foi incluído adicionalmente um grupo clínico ($n=17$), composto por indivíduos com diagnóstico de demência. Deste grupo fizeram parte 14 participantes do sexo feminino e 3 do sexo masculino, com idades compreendidas entre os 68 e os 92 anos ($M=81,00$; $DP=6,82$) e uma média de 8,35 anos de escolaridade ($DP=3,66$).

Todos os participantes deviam ter idade igual ou superior a 65 anos e dominar a língua portuguesa. Relativamente aos critérios de exclusão aplicados, designaram-se problemas médicos crónicos que interferissem com o funcionamento cognitivo (e.g., doença cerebrovascular, convulsões), história de traumatismo crânio-encefálico ou de abuso de substâncias, dificuldades visuais e/ou auditivas não corrigidas e dificuldades de compreensão oral. Particularmente, a presença de défice cognitivo foi critério de exclusão para os participantes dos grupos simulador e não-simulador e critério de inclusão para os participantes do grupo clínico, devendo estar associado a diagnóstico médico de demência. De notar que nenhum dos participantes se encontrava envolvido em processos de carácter forense e/ou médico-legal.

2.2. Instrumentos

No total, foram administrados sete instrumentos: um questionário de dados sociodemográficos, dois instrumentos de rastreio cognitivo (MMSE e MoCA), uma medida para avaliação da simulação (do subteste de Memória de Dígitos da WAIS-III-R) e três testes de validade de desempenho (Rey 15-IMT, TOMM e CIH-EV).

2.2.1. Questionário de Dados Sociodemográficos

Foram realizadas questões para recolha de dados sociodemográficos, tais como idade, sexo, ocupação, estado civil, número de filhos, tipo de habitação, orçamento familiar, habilitações literárias e atividade profissional.

2.2.2. Mini-Mental State Examination (MMSE; Folstein, Folstein & McHugh, 1975; Guerreiro et al., 1994; Guerreiro, 1998)

Instrumento de rastreio breve do estado cognitivo global, aplicável a indivíduos com idade igual ou superior a 25 anos. É constituído por diversas tarefas que visam avaliar as seguintes funções cognitivas: orientação espaço-temporal (10 pontos), repetição (3 pontos),

atenção e cálculo (5 pontos), memória (3 pontos), linguagem (8 pontos) e capacidade construtiva (1 ponto), possibilitando um resultado máximo de 30 pontos. Relativamente aos pontos de corte para a população portuguesa (Guerreiro et al., 1994; Guerreiro, 1998), considera-se a existência de défice cognitivo com uma pontuação igual ou inferior a 22 no caso de indivíduos com 1 a 11 anos de escolaridade e com uma pontuação igual ou inferior a 27 para indivíduos com escolaridade superior a 11 anos. Para o grupo clínico utilizou-se um ponto de corte inferior a 26 pontos (Freitas et al., 2012a; Freitas et al., 2012b; Santana et al., 2016). A inclusão deste instrumento no protocolo de avaliação deve-se ao facto de este ser o teste de rastreio cognitivo mais utilizado, amplamente validado e mais referenciado na literatura (Simões, Santana & Grupo de Estudos de Envelhecimento Cerebral e Demência, 2015).

2.2.3. *Montreal Cognitive Assessment* (MoCA; Nasreddine et al., 2005; Freitas, Simões, Alves & Santana, 2011)

Instrumento de avaliação cognitiva breve, desenvolvido para avaliar formas mais ligeiras de declínio cognitivo, permitindo distinguir entre alterações cognitivas normativas e patológicas em indivíduos com idade igual ou superior a 25 anos. As suas tarefas permitem avaliar as seguintes funções cognitivas: funcionamento executivo (4 pontos), capacidade visuoespacial (4 pontos), memória a curto-prazo (5 pontos), memória de trabalho, atenção e concentração (5 pontos), linguagem (6 pontos) e orientação spatiotemporal (6 pontos). De acordo com as normas da população portuguesa (Freitas et al., 2011), no presente estudo consideraram-se os pontos de corte abaixo de 1.5 desvios-padrão para os indivíduos com mais de 65 nos, em função do seu nível educacional. Assim, determina-se a existência de défice cognitivo com uma pontuação igual ou inferior a 16 para 1 a 4 anos de escolaridade, a 20 pontos para 5 a 9 anos de escolaridade, a 22 pontos para 10 a 12 anos de escolaridade e a 24 pontos para mais de 12 anos de escolaridade. Para o grupo clínico foi utilizado um ponto de corte inferior a 17 (Freitas et al., 2012a; Freitas et al., 2012b; Freitas et al., 2013). Este instrumento

foi considerado no presente protocolo, uma vez que, comparativamente ao MMSE, avalia adicionalmente as funções executivas e se mostra mais sensível a estádios iniciais de demência (Nasreddine et al., 2005; Freitas et al., 2010).

2.2.4. *Subteste de Memória de Dígitos da WAIS-III-R* (Wechsler, 1997; 2008)

Trata-se de um dos subtestes da WAIS-III-R, que pretende avaliar as capacidades mnésicas, de atenção e de concentração. É pedido ao participante que repita exatamente as mesmas sequências numéricas, que aumentam progressivamente a sua extensão e, assim, a sua dificuldade. Este é composto por duas condições: repetição no sentido direto, com uma pontuação máxima de 16 pontos e repetição em sentido inverso, com uma pontuação máxima de 14 pontos. A partir deste subteste é possível obter a pontuação do sentido direto, com um ponto de corte ≤ 5 , a pontuação do sentido inverso, com um ponto de corte ≤ 2 e o Índice de Fiabilidade da Memória de Dígitos (*Reliable Digit Span – RDS*), com um ponto de corte ≤ 6 (Castro, 2015; Pinho 2012). Este subteste foi incluído no protocolo de avaliação, uma vez que a literatura (e.g., Babikian, Boone, Lu & Arnold, 2006; Jasinsky, Berry Shandera & Clark, 2011) evidencia a sua pertinência na deteção de simulação de défices neurocognitivos.

2.2.5. *Rey 15-Item Memory Test* (Rey 15-IMT; Rey, 1964; Boone, Salazar, Lu, Warner-Chacon & Razani, 2002; Simões et al., 2010)

Este instrumento foi desenvolvido para avaliar o esforço em tarefas de memória e o exagero ou a simulação de queixas de memória. É composto por um cartão com 15 itens (letras, formas geométricas e números), com três itens distribuídos por cinco linhas, apresentado durante 10 segundos, após os quais é retirado. De seguida, passa-se ao Ensaio de Evocação Livre, em que o participante deve reproduzir imediatamente o maior número de itens que conseguir recordar, cotando-se o número de itens corretamente reproduzidos, independentemente da sua localização espacial, com uma pontuação máxima de 15 pontos.

Posteriormente é administrado um Ensaio de Reconhecimento, onde se apresentam os 15 itens originais e 15 distratores, pedindo-se ao participante que identifique o maior número de itens originais que conseguir. Ao nível da cotação, considera-se o número de itens corretamente identificados, o número de falsos positivos (i.e., itens incorretamente identificados) e um resultado combinado [número de itens corretamente evocados + (número de itens corretamente reconhecidos – número de falsos positivos)], com uma pontuação máxima possível de 30 pontos. Relativamente aos pontos de corte, utilizaram-se pontuações inferiores a 9 para o Ensaio de Evocação Livre Imediata e inferiores a 20 para o Resultado Combinado (Simões et al., 2010). Este instrumento foi considerado na presente investigação, por ser um dos testes de validade de desempenho mais utilizados e por estar validado para a população adulta idosa portuguesa (Simões et al., 2010), sendo, assim, útil para a análise da validade convergente com o CIH-EV (Daugherty et al., no prelo).

2.2.6. *Test of Memory Malingering* (TOMM; Tombaugh, 1996; Mota, 2008; Fernandes, 2009)

Trata-se de um teste de reconhecimento visual, construído para avaliar a validade de desempenho. É composto por dois Ensaios de Aprendizagem, nos quais são apresentadas 50 figuras durante 3 segundos cada, com um intervalo de 1 segundo. Imediatamente a seguir, estas são novamente apresentadas em pares, cada uma emparelhada com uma figura distratora. Uma vez que esta é uma tarefa de escolha forçada, o participante deve escolher, para cada par, qual das figuras viu anteriormente. Desta forma, os ensaios diferem em termos da ordem de apresentação das 50 figuras e das figuras distratoras utilizadas. A todas as respostas dadas é fornecido um *feedback* e por cada resposta correta é atribuído 1 ponto, perfazendo um total máximo de 50 pontos para cada ensaio. Como ponto de corte, utilizou-se uma pontuação inferior a 45 para o segundo ensaio (e.g., Cruz, 2008; Fernandes, 2009; Maior, 2008). O TOMM é um dos TVS de referência na investigação e na prática clínica, estando validado para

a população adulta idosa portuguesa (Cruz, 2008; Fernandes, 2009; Maior, 2008), sendo, portanto, pertinente para a análise da validade convergente no presente estudo.

2.2.7. *Coin in Hand – Extended Version* (CIH-EV; Kapur, 1994; Daugherty et al., no prelo)

Instrumento inicialmente desenvolvido para distinguir entre pacientes com perturbações neurocognitivas e pacientes que exageram ou simulam queixas de memória (Kapur, 1994). O participante é inicialmente informado que o teste pretende medir de que forma a memória consegue resistir à distração. A tarefa é de escolha forçada, sendo que o participante observa um par de mãos com uma moeda na mão esquerda ou na mão direita. De seguida, as mãos fecham-se em punho e decorre uma contagem decrescente, após a qual surgem novamente os punhos cerrados, momento em que o participante deve seleccionar a mão onde se encontra a moeda. Após cada resposta é fornecido feedback. No presente estudo foi utilizada a versão extensa e informatizada (CIH-EV; REF.), que inclui três níveis de dificuldade, com 10 ensaios cada um (i.e., a moeda aparece 5 vezes na mão esquerda e 5 vezes na mão direita, aleatoriamente), num total de 30 ensaios. Os níveis diferem em termos de contagem decrescente, ou seja, no primeiro nível o participante conta a partir de 10 durante 10 segundos, fazendo depois a sua escolha; no segundo nível conta a partir de 99 durante 15 segundos; e no terceiro nível conta a partir de 999 durante 20 segundos, sendo que no segundo e terceiro níveis o participante é informado de que a dificuldade do teste irá aumentar. O ponto de corte determinado para jovens adultos, incluindo portugueses (Daugherty et al., no prelo) foi ≤ 27 , que apresenta equivalência cultural.

2.3. Procedimento

A presente investigação foi aprovada pela Comissão de Deontologia do Conselho Científico da Faculdade de Psicologia da Universidade de Lisboa, obedecendo aos critérios éticos e deontológicos dispostos no seu regulamento.

Os participantes dos grupos Simulador e Não-Simulador foram testados em três Universidades Sêniores, tendo sido distribuídos aleatoriamente pelos dois grupos. As testagens dos participantes do grupo Clínico foram realizadas em duas Residências Sêniores, sendo que o diagnóstico de demência foi previamente atribuído pela equipa médica e (neuro)psicológica de ambas as Residências. Todas as Instituições pertenciam ao distrito de Lisboa, tendo-lhes sido inicialmente dirigida uma carta informativa acerca dos objetivos e procedimentos do presente do estudo. Além disso, todos os participantes assinaram um consentimento informado que apresentava o objetivo do estudo e o tipo e a duração das tarefas aplicadas e que garantia a participação voluntária, o anonimato e a confidencialidade, a possibilidade de desistência sem prejuízo e a possibilidade de divulgação posterior dos resultados da presente investigação. Após o seu consentimento, foi atribuído um código alfanumérico a cada participante, de forma a garantir o anonimato dos dados.

A administração do presente protocolo de investigação foi realizada individualmente, com uma duração aproximada de 50 minutos. No caso específico do grupo clínico, a aplicação dos instrumentos foi dividida em duas sessões de aproximadamente 30 minutos cada. As sessões foram iniciadas pelo questionamento acerca dos critérios de exclusão. O protocolo de investigação iniciou-se pelo preenchimento do Questionário Sociodemográfico, com exceção dos participantes do grupo clínico, cujas informações foram disponibilizadas pelas Instituições correspondentes, e posterior administração do MMSE e do MoCA. Após o rastreio cognitivo, os participantes do grupo Simulador foram instruídos a desempenhar os testes de validade de desempenho (subteste de Memória de Dígitos, do TOMM, do Rey-15IMT e do CIH-EV)

simulando um déficit mnésico típico de uma demência, com o objetivo de obter algum tipo de pensão, subsídio ou reforma antecipada. A ordem de aplicação destes testes foi aleatorizada. A administração do TOMM e do CIH-EV foi realizada com recurso a um computador portátil, com um ecrã de 13,3" e, relativamente ao CIH-EV. De acordo com Daugherty et al. (no prelo), todos participantes foram inicialmente informados sobre a existência de três níveis de dificuldade no teste.

A análise estatística dos resultados foi realizada com recurso ao *Statistical Package for Social Sciences* (SPSS, v.25.0).

III. Resultados

Análise das diferenças entre grupos nas variáveis sociodemográficas e no desempenho cognitivo

Compararam-se as diferenças entre os grupos não-simulador e simulador relativamente às variáveis sociodemográficas e aos resultados do MMSE e do MoCA através de testes *t*. Não se verificaram diferenças significativas entre os dois grupos (ver Tabela 1). Para comparar os grupos simulador e clínico recorreu-se ao teste não-paramétrico *U* de Mann-Whitney, que revelou diferenças significativas quanto à idade ($U = 103.50, p < .001$), à habitação ($U = .00, p < .001$) e aos resultados dos instrumentos de rastreio cognitivo (MMSE: $U = 37.00, p < .001$; e MoCA: $U = 3.00, p < .001$) (ver Tabela 1).

Tabela 1

Médias, desvios-padrão e percentagens para as variáveis sociodemográficas e para o rastreio cognitivo dos três grupos.

	Grupo Não-Simulador (n=30)	Grupo Simulador (n=29)	<i>t</i> ^a	Grupo Clínico (n=17)	<i>U</i> ^b
Sexo			.30		237.50
Feminino	25 (83.33%)	25 (86.21%)		14 (82.35%)	
Masculino	5 (16.67%)	4 (13.79%)		3 (17.65%)	
Idade	75.20 ± 8.90	73.31 ± 6.40	.93	81.00 ± 6.82	103.50*
Anos de Escolaridade					
0-4	9 (30.00%)	4 (13.79%)		3 (17.65%)	
5-9	10 (33.33%)	13 (44.83%)		7 (41.18%)	
10-12	8 (26.67%)	10 (34.48%)		5 (29.41%)	
>12	3 (10.00%)	2 (6.90%)		2 (11.76%)	
Média	8.43 ± 3.94	9.07 ± 3.36	- .67	8.35 ± 3.66	223.50
Estado Civil			.07		221.50
Solteiro	3 (10.00%)	3 (9.69%)		4 (20.00%)	
Casado	13 (43.33%)	11 (35.48%)		5 (25.00%)	
Viúvo	5 (16.67%)	11 (35.48%)		9 (45.00%)	
Divorciado	9 (30.00%)	6 (19.35%)		2 (10.00%)	
Filhos	1.73 ± 1.14	1.42 ± 0.85	1.21	1.60 ± 1.35	222.50

(Continua)

	Grupo Não-Simulador (n=30)	Grupo Simulador (n=29)	t^a	Grupo Clínico (n=17)	U^b
Habitação			.39		
Sozinho	14 (46,67%)	18 (58,06%)		1 (5,00%)	.00*
Cônjuge	12 (40,00%)	8 (25,81%)		2 (10,00%)	
Famíliares	3 (10,00%)	5 (16,13%)		0 (0,00%)	
Instituição	1 (3,33%)	0 (0,00%)		17 (85,00%)	
Orçamento familiar			1.21		.87
≤ 1200€	18 (60,00%)	21 (67,74%)		-	
Entre 1200€ a < 1800€	6 (20,00%)	9 (29,03%)		-	
≥ 1800€	6 (20,00%)	1 (3,23%)		-	
Profissão durante mais tempo			.96		181.50
Baixa	12 (40,00%)	14 (48,28%)		11 (64,71%)	
Intermédia	12 (40,00%)	12 (41,38%)		4 (23,53%)	
Alta	6 (20,00%)	3 (10,34%)		2 (11,76%)	
Média	Intermédia (M=1,80 ± .76)	Intermédia (M=1,62 ± .68)		Baixa (M=1,47±1,71)	
Última profissão					
Baixa	12 (40,00%)	13 (44,83%)		11 (64,71%)	170.50
Intermédia	13 (43,33%)	13 (44,83%)		4 (23,53%)	
Alta	5 (16,67%)	3 (10,34%)		2 (11,76%)	
Média	Intermédia (M=1,77 ± .73)	Intermédia (M=1,66 ± .67)	.61	Baixa (M=1,47±0,72)	
MMSE	29.07 ± .17	28.86 ± .18	.85	19.59 ± 1.36	37.00*
MoCA	24.67 ± .49	23.69 ± .41	1.53	11.41 ± 1.15	3.00*

Análise da relação entre as variáveis sociodemográficas, o desempenho cognitivo e o desempenho nos testes

Utilizou-se o coeficiente de correlação de Pearson para analisar a relação entre as variáveis sociodemográficas, o desempenho cognitivo e o resultado no CIH-EV para cada grupo. Não se observaram correlações significativas em nenhum dos grupos (ver Tabela 2).

Tabela 2

Correlação entre as variáveis sociodemográficas, o desempenho cognitivo e o desempenho no CIH-EV para cada grupo.

	Desempenho no CIH-EV		
	Grupo Não-Simulador (n=30)	Grupo Simulador (n=29)	Grupo Clínico (n=17)
Sexo	-.09	.24	.14
Idade	-.01	.32	-.09
Anos de Escolaridade	.07	-.07	.01
Estado Civil	.02	.07	.04
Filhos	.29	-.02	.06
Habitação	-.12	-.02	-
Orçamento familiar	.35	.14	-
Profissão durante mais tempo	-.29	.03	-.10
Última profissão	-.29	-.05	-.10
MMSE	.26	.14	.19
MoCA	.11	.21	.27

Nota. Valores de correlação com base no *r* de Pearson.

A mesma análise foi realizada para as restantes medidas de validade de desempenho utilizadas (ver Tabela 3). Para o grupo de adultos idosos saudáveis não simuladores, todos os testes apresentaram correlações significativas com, pelo menos, uma variável sociodemográfica. No grupo de idosos saudáveis simuladores, um dos instrumentos de rastreio cognitivo (MoCA) e dois testes de validade de desempenho (Rey-15IMT e TOMM) mostraram correlações significativas com variáveis sociodemográficas. Finalmente, no grupo com diagnóstico de demência verificou-se que apenas uma variável sociodemográfica se correlacionou significativamente com um dos testes de rastreio cognitivo (MoCA).

Tabela 3

Correlação entre as variáveis sociodemográficas e o desempenho nos diferentes testes para cada grupo.

	MMSE	MoCA	RDS	Rey Evoc	Rey RC	TOMMI	TOMM 2
Não-simuladores							
Sexo	-.23	-.01	.17	-.07	-.03	-.28	-.38*
Idade	-.35	-.35	-.36	-.52**	-.43*	-.33	-.16

(Continua)

	MMSE	MoCA	RDS	Rey Evoc	Rey RC	TOMM1	TOMM2
Escolaridade	.32	.50**	-.10	.39*	.09	-.03	-.17
Profissão+	-.08	-.34	-.09	-.50**	-.04	-.27	-.25
Última profissão	-.13	-.39*	.01	-.45*	-.05	-.23	-.21
Clínico							
Sexo	.12	.16	.47	.02	.15	-.06	-.07
Idade	-.13	-.20	.37	-.26	-.25	-.03	-.15
Escolaridade	.38	.49*	.19	.22	.40	-.09	-.15
Profissão+	-.26	-.35	.14	-.22	-.29	.10	.20
Última profissão	-2.8	-.35	.14	-.22	-.29	.10	.20

Nota. Rey Evoc = Ensaio de Evocação do Rey-15IMT. Rey RC = Resultado Combinado do Reconhecimento do Rey-15IMT; TOMM1 = Primeiro Ensaio do TOMM; TOMM2 = Segundo Ensaio do TOMM. Profissão+ = Profissão durante mais tempo.

* $p < .05$.

** $p < .01$.

Análise da precisão e do tempo de resposta dos três grupos para os níveis de dificuldade e para o resultado total do CIH-EV

Na Tabela 4 são apresentados os desempenhos dos três grupos no CIH-EV por nível de dificuldade, relativamente à precisão e ao tempo de resposta.

Para averiguar a existência de diferenças significativas entre os três níveis de dificuldade do CIH-EV realizaram-se duas ANOVAS 3x3, com o nível de dificuldade (nível 1, 2 e 3) e o grupo (não-simulador, simulador e clínico) como variáveis independentes e com a precisão e o tempo de resposta como variáveis dependentes. Foi ainda realizada uma ANOVA adicional com os mesmos grupos para a precisão (i.e., a soma de acertos nos três níveis) e o tempo de resposta médios totais entre os três níveis de dificuldade.

Tabela 4

Média e desvio-padrão da precisão e do tempo para cada nível do CIH-EV para os três grupos.

	Grupo não-simulador			Grupo simulador			Grupo clínico		
	N1	N2	N3	N1	N2	N3	N1	N2	N3
Precisão	9.77	9.83	9.80	4.14	4.72	4.03	8.359	8.35	8.29
<i>M(SE)</i>	(.57)	(.46)	(.48)	(1.85)	(1.41)	(1.43)	(1.00)	(1.46)	(1.53)
Tempo	2501.40	2386.15	2410.40	3108.47	3118.69	3199.12	4331.24	4440.94	4498.18
<i>M(SE)</i>	(1371.64)	(822.01)	(732.52)	(1332.62)	(1259.09)	(1558.57)	(2123.76)	(2123.76)	(2299.87)

Nota. N1 = nível 1, N2 = nível 2, N3 = nível 3. Precisão = número de acertos. Tempo = tempo em segundos. *M* = média. *SE* = Standard Error.

Relativamente à precisão, a ANOVA realizada não revelou um efeito significativo de nível de dificuldade [$F(2), 146=1.31, p>.10$], mas verificou-se um efeito significativo de grupo [$F(2), 73=312.78, p<.001., \eta_p^2 =.89$]. Além disso, a interação entre nível de dificuldade e grupo [$F(4), 146=1.04, p>.10$] não foi significativa. Segundo as análises *post-hoc*, o grupo de adultos idosos não-simulador apresentou melhor desempenho ($M=9.08$) do que o grupo clínico ($M=8.33$) e ambos tiveram melhor desempenho do que o grupo de simuladores ($M=4.30$) (ver Figura 1).

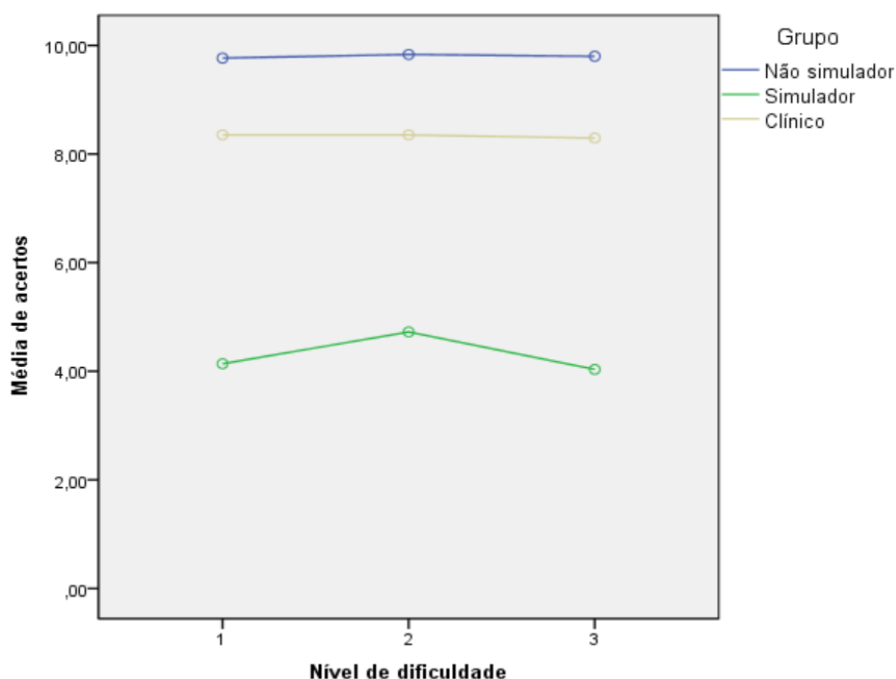


Figura 1. Precisão de cada grupo nos três níveis de dificuldade do CIH-EV.

Em relação ao tempo de resposta, a ANOVA revelou a ausência de um efeito significativo de nível de dificuldade [$F(2), 146 = .058, p > .10$], tendo-se observado um efeito significativo de grupo [$F(2), 73 = 567.43, p > .001, \eta_p^2 = .94$]. Não se verificou adicionalmente uma interação significativa entre o nível de dificuldade e o grupo [$F(4), 146 = .101, p > .10$]. As análises *post-hoc* revelaram que o grupo de adultos idosos não-simuladores teve um melhor tempo de resposta ($M=2432.65$) do que o grupo simulador ($M=3142,09$) e ambos deram respostas mais rápidas do que o grupo clínico ($M=4423.45$) (ver Figura 2).

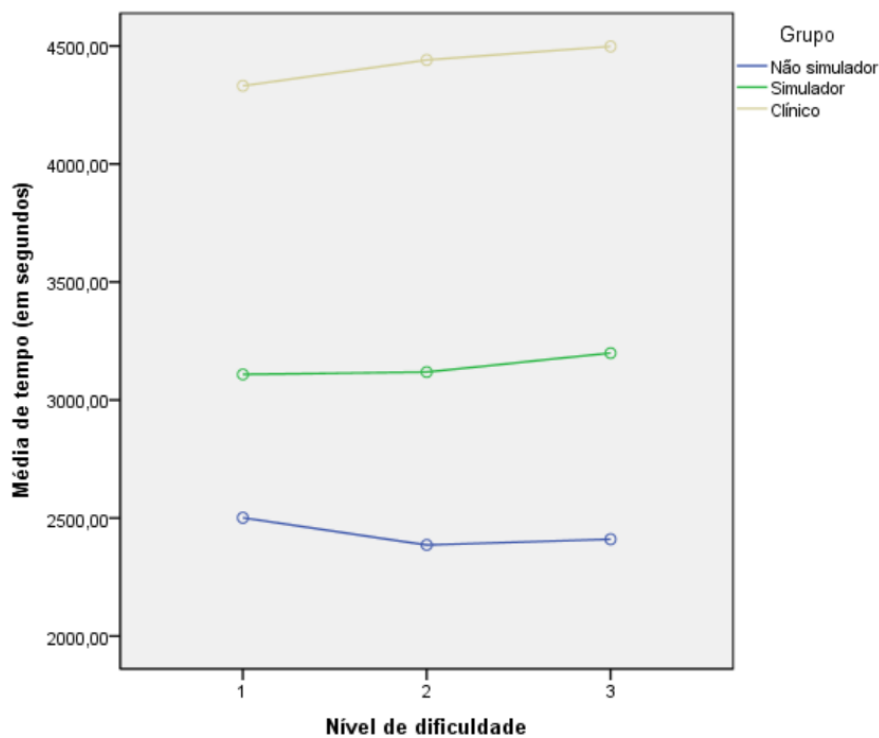


Figura 2. Média do tempo de resposta por grupo e nível de dificuldade no CIH-EV.

Adicionalmente, agruparam-se os níveis de dificuldade e analisaram-se os resultados de precisão e de tempo de resposta totais em função do grupo. No caso da precisão, verificou-se um efeito de grupo [$F(2), 75 = 312.78, p > .001, \eta_p^2 = .89$], sendo que o grupo simulador teve

um pior desempenho ($M=12.90$) do que o grupo clínico ($M=25.00$) que, por sua vez, teve um pior desempenho do que o grupo não-simulador ($M=29.40$) (ver Tabela 5 e Figura 3).

Tabela 5

Média e desvio-padrão da precisão e do tempo de resposta totais no CIH-EV para os três grupos.

	Grupo não-simulador	Grupo simulador	Grupo clínico	<i>F</i>
Precisão	29.40	12.90	25.00	312.78*
<i>M(SE)</i>	(.19)	(.58)	(.81)	
Tempo	2373.38	2999.55	4205.62	14.48*
<i>M(SE)</i>	(135.03)	(182.41)	(424.48)	

Nota. *M* = média. *SE* = *Standard Error*.

* $p < .01$.

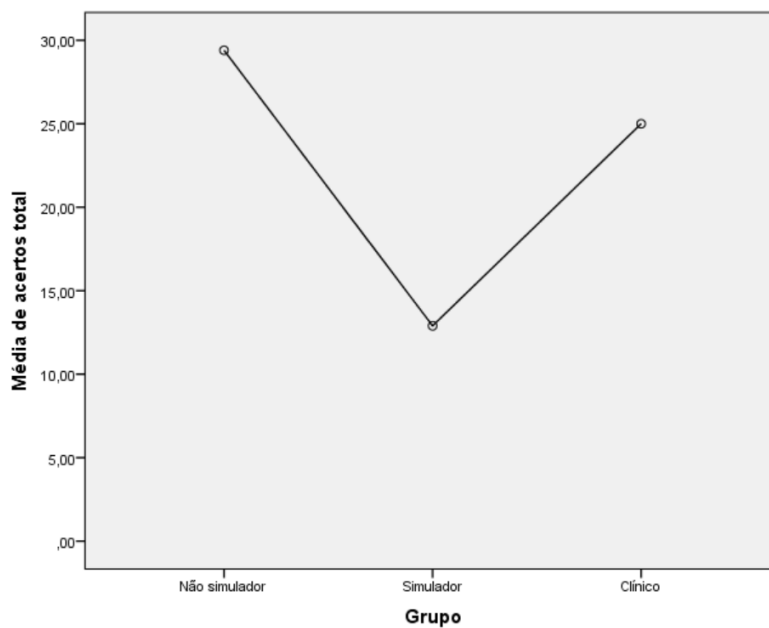


Figura 3. Média total de acertos no CIH-EV de cada grupo.

Relativamente ao tempo de resposta, observou-se igualmente um efeito de grupo [$F(2)$, $75 = 14.48$, $p < .001$, $\eta_p^2 = .28$]. O grupo não-simulador apresentou um melhor tempo de resposta ($M=2373.38$) do que o grupo simulador ($M=2999.55$) e ambos foram mais rápidos do que o grupo clínico ($M=4205.62$) (ver Figura 4).

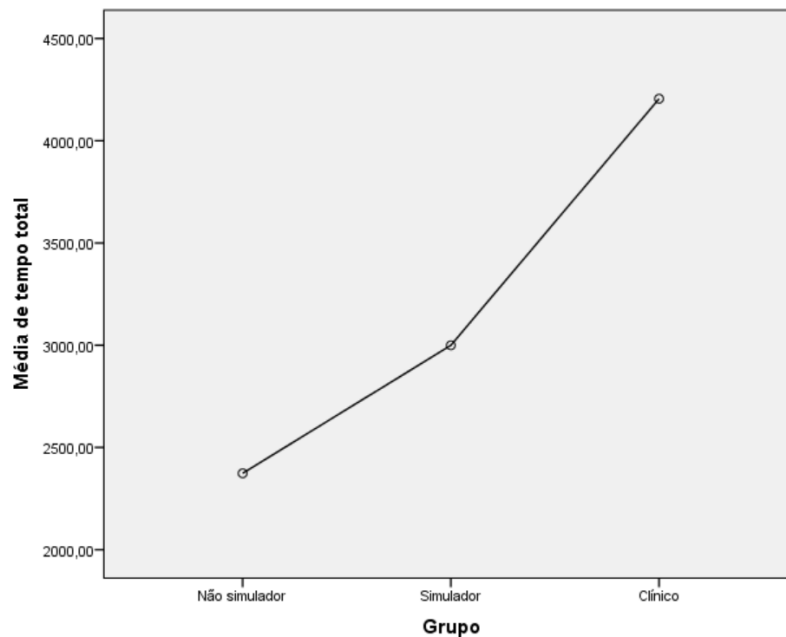


Figura 4. Média total dos tempos de resposta no CIH-EV de cada grupo.

Análise e determinação dos pontos de corte do CIH-EV para idosos saudáveis

De acordo com um dos objetivos do presente estudo, utilizou-se uma análise de curva ROC para estabelecer os pontos de corte adequados para adultos idosos saudáveis. A análise da curva ROC relativamente à comparação entre os grupos de idosos saudáveis simulador e não-simulador revelou que o tempo de resposta não se mostrou discriminativo entre os dois grupos, apresentando uma área abaixo da curva ROC (AUC) reduzida (AUC =.30).

A comparação da precisão do desempenho no CIH-EV dos grupos simulador e não-simulador revelou uma boa capacidade discriminativa (níveis de dificuldade 1: AUC=.98; níveis de dificuldade 2 e 3: AUC=1.00) entre os grupos, pelo que se determinou o ponto de corte mais adequado para cada nível. Este foi o mesmo para todos os níveis (≤ 8), traduzindo sensibilidades entre 97% e 100% e especificidades entre 93% e 97% (ver Tabela 6).

A análise da curva ROC que comparou os grupos simulador e não-simulador quanto ao número total de respostas corretas revelou uma área abaixo da curva excelente (AUC=1.00) e,

por isso, os pontos de corte foram determinados em função desta variável. Desta forma, para adultos idosos saudáveis selecionou-se o ponto de corte ≤ 26 para o número total de acertos no CIH-EV, que demonstrou uma sensibilidade de 100% e uma especificidade de 97% (ver Tabela 6).

Tabela 6.

Sensibilidade e especificidade dos pontos de corte do CIH-EV para os adultos idosos saudáveis.

Nível de dificuldade	Acertos	Sensibilidade	Especificidade
1	≤ 8	.97	.93
	≤ 9	.97	.83
	≤ 10	1	.00
2	≤ 8	1	.97
	≤ 9	1	.87
	≤ 10	1	.00
3	≤ 8	1	.97
	≤ 9	1	.83
	≤ 10	1	.00
Total	≤ 26	1	.97
	≤ 27	1	.90
	≤ 28	1	.87
	≤ 29	1	.67
	≤ 30	1	.00

Análise e determinação dos pontos de corte do CIH-EV para idosos com diagnóstico de demência

Análises de curva ROC foram levadas a cabo para estabelecer os pontos de corte adequados a idosos com diagnóstico de demência. Assim, foram comparados os grupos de idosos saudáveis simuladores com o grupo clínico (idosos com demência).

Relativamente aos níveis de dificuldade, estes mostraram-se discriminativos (AUC=.95 para o nível 1; AUC=.96 para os níveis 2 e 3), pelo que foram determinados os pontos de corte mais adequados. Para os níveis de dificuldade 1 e 3, o ponto de corte ≤ 5 representou uma sensibilidade de 94% e uma sensibilidade de 79% e 83%, respetivamente. Para o nível 2, um

ponto de corte ≤ 6 refletiu uma especificidade de 88% e uma sensibilidade de 97% (ver Tabela 7).

Para comparação dos grupos de idosos simulador e clínico quanto ao número total de acertos (precisão do desempenho), a análise da curva ROC apresentou uma área abaixo da curva excelente (AUC=.99), determinando-se o ponto de corte ≤ 17 acertos como o mais adequado para o grupo de idosos com demência, traduzindo uma sensibilidade de 94% e uma especificidade de 97% (ver Tabela 7).

Tabela 7

Sensibilidade e especificidade dos pontos de corte do CIH-EV para o número total de acertos para adultos idosos com diagnóstico de demência.

Nível de dificuldade	Acertos	Sensibilidade	Especificidade
1	≤ 4	1	.69
	≤ 5	.94	.79
	≤ 6	.94	.31
	≤ 7	.94	.66
	≤ 8	.53	.66
	≤ 9	.00	.66
2	≤ 5	.94	.66
	≤ 6	.88	.97
	≤ 7	.77	.00
3	≤ 4	.94	.59
	≤ 5	.94	.83
	≤ 6	.94	.00
Total	≤ 16	1	.83
	≤ 17	.94	.97
	≤ 18	.88	.00
	≤ 19	.77	.00
	≤ 20	.70	.00

Análise da validade convergente

Utilizou-se o coeficiente de correlação de Pearson para analisar a validade convergente do CIH-EV com as restantes medidas de validade de desempenho utilizadas. Relativamente ao Rey-15IMT, verificou-se uma correlação significativa, baixa, entre o CIH-EV e o Ensaio de Evocação ($r=.24$, $p<.05$) e uma correlação moderada com o Resultado Combinado ($r=.62$, $p<.001$). Com o RDS observou-se uma correlação elevada ($r=.73$, $p<.01$) e, finalmente, com

o TOMM, as correlações revelaram-se muito elevadas, quer para o Ensaio 1 ($r=.92, p<.01$), quer para o Ensaio 2 ($r=.93, p<.01$). Desta forma, o CIH-EV apresenta boa validade convergente com as medidas utilizadas para avaliar a validade de desempenho.

Além disso, analisaram-se as percentagens de indivíduos corretamente identificados como não-simuladores (i.e., verdadeiros negativos) e incorretamente identificados como simuladores (i.e., falsos positivos). Para tal, estas percentagens foram comparadas utilizando os pontos de corte propostos na literatura para o TOMM (Cruz, 2008; Maior, 2008; Tommbaugh, 1996), o Rey-15IMT (Simões et al., 2010) e o RDS (Castro, 2015; Pinho 2012) e o ponto de corte recomendado no presente estudo para o CIH-EV. Verificou-se que o CIH-EV e o Resultado Combinado do Rey-IMT são os instrumentos com menor percentagem de falsos positivos (5.9% e 0.0%, respetivamente). Porém, entre estas duas medidas, o CIH-EV foi o instrumento que detetou corretamente mais simuladores (96.6%) e que identificou corretamente mais indivíduos não-simuladores (100.0%) (ver Tabela 8).

Tabela 8

Comparação das percentagens de verdadeiros negativos e falsos positivos entre os testes.

Grupo	Ponto de Corte	Condição Real	
		Não-simulador	Simulador
Não-simulador			
CIH-EV_total	≤26	100.0%	0.0%
TOMM1	≤33	100.0%	0.0%
TOMM2	≤45	100.0%	0.0%
Rey_Evoc	5.70	48.4%	0.0%
Rey_RC	11.70	76.7%	23.3%
RDS	≤6	96.7%	3.3%
Simulador			
CIH-EV	≤26	3,4%	96.6%
TOMM1	≤33	6.9%	93.1%
TOMM2	≤45	0.0%	100%
Rey_Evoc	5.70	96.6%	3,4%
Rey_RC	11.70	6.9%	93.1%

Grupo	Ponto de Corte	Condição Real	
		Não-simulador	Simulador
RDS	≤6	20.7%	79.3%
Clínico			
CIH-EV	≤17	94.1%	5.9%
TOMM1	≤33	70.6%	29.4%
TOMM2	≤45	35.3%	64.7%
Rey_Evoc	1.90	70.6%	29.4%
Rey_RC	2	100.0%	0.0%
RDS	≤6	64.7%	35.3%

Nota. CIH-EV_total = número total de acertos no CIH-EV. Rey_Evoc = Ensaio de Evocação do Rey-15IMT. Rey_RC = Resultado Combinado do Reconhecimento do Rey-15IMT; TOMM1 = Primeiro Ensaio do TOMM; TOMM2 = Segundo Ensaio do TOMM.

Analisou-se ainda a adequabilidade dos pontos de corte do Rey-15IMT, do TOMM e do RDS na discriminação entre o grupo simulador e o grupo clínico, através de uma análise de curva ROC. Relativamente ao Rey-15IMT, este apresentou uma área abaixo da curva reduzida no Ensaio de Evocação Imediata (AUC=.10), no Ensaio de Reconhecimento (AUC=.64) e no Resultado Combinado (AUC=.35) (ver Figura 5), portanto não discriminativa. Assim, não se selecionaram os pontos de corte para este teste.

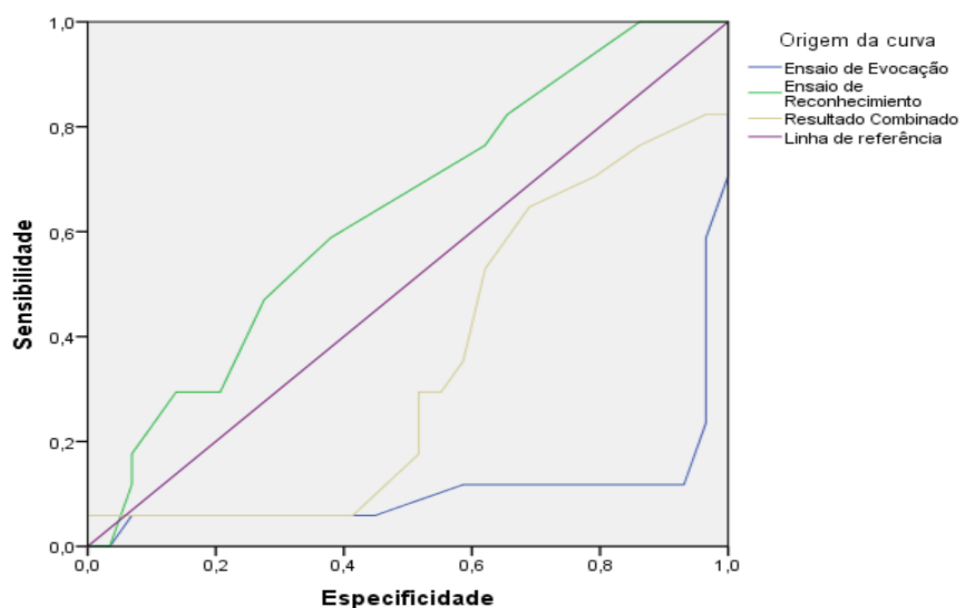


Figura 5. Curva ROC relativa à capacidade discriminativa do Rey-15IMT.

Da mesma forma, analisou-se a adequabilidade dos pontos de corte do TOMM na discriminação entre o grupo simulador e o grupo clínico. A análise da curva ROC apresentou uma elevada área abaixo da curva, quer no Ensaio 1 (AUC=.96), quer no Ensaio 2 (AUC = .99) (ver Figura 6) e, por isso, procedeu-se à análise dos pontos de corte. Para o Ensaio 1, a pontuação mais discriminativa foi ≤ 32 , com uma sensibilidade de 71% e uma especificidade de 93%. Para o Ensaio 2, o ponto de corte mais adequado (≤ 34) traduziu uma sensibilidade de 94% e uma especificidade de 97% (ver Tabela 9).

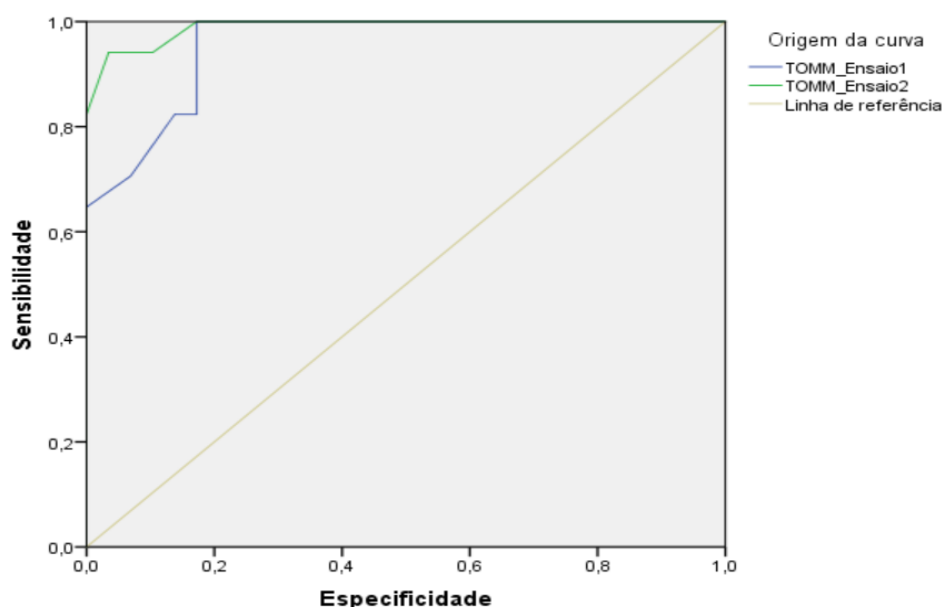


Figura 6. Curva ROC relativa à capacidade discriminativa do TOMM.

Tabela 9

Sensibilidade e especificidade dos pontos de corte do Rey-15-IMT para idosos com diagnóstico de demência.

Ensaio	Pontuação	Sensibilidade	Especificidade
Ensaio 1	≤ 28	.82	.83
	≤ 29	.82	.86
	≤ 32	.71	.93
	≤ 34	.65	1
Ensaio 2	≤ 32	.94	.93
	≤ 34	.94	.97
	≤ 35	.82	1

Finalmente, para a discriminação entre os grupos simulador e clínico com a RDS, a análise da curva ROC revelou uma área abaixo da curva relativamente elevada ($AUC=.80$) (ver Figura 7), procedendo-se à análise dos seus pontos de corte. A pontuação selecionada (≤ 5) traduziu uma sensibilidade de 65% e uma especificidade de 79% (ver Tabela 10).

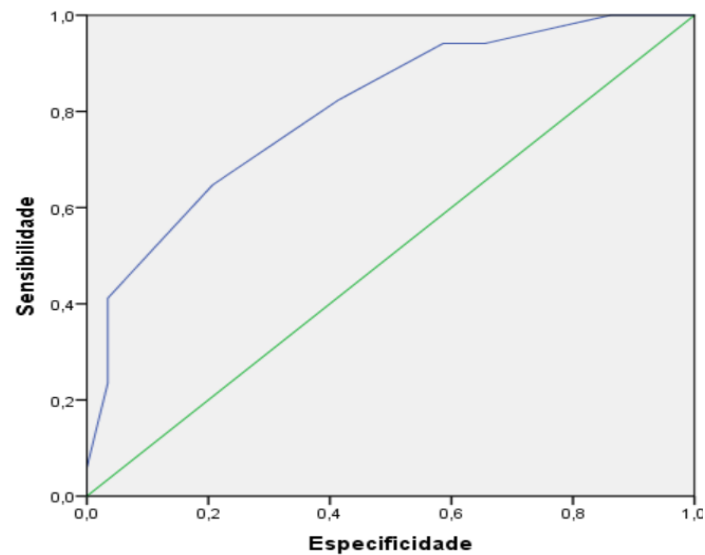


Figura 7. Curva ROC relativa à capacidade discriminativa do RDS.

Tabela 10

Sensibilidade e especificidade dos pontos de corte da RDS para idosos com diagnóstico de demência.

Índice	Pontuação	Sensibilidade	Especificidade
RDS	≤ 4	.82	.59
	≤ 5	.65	.79
	≤ 6	.41	.97

IV. Discussão e conclusões

Com o presente estudo pretendeu-se validar o Teste de Validade de Desempenho CIH-EV (Daugherty et al., no prelo) com adultos idosos saudáveis e com diagnóstico de demência, segundo um paradigma de simulação (Merten et al., 2005; Tydecks et al., 2006; Rogers, 2008.). Particularmente, visou-se i) analisar a relação entre variáveis sociodemográficas e o desempenho no CIH-EV dos três grupos testados (idosos saudáveis não-simuladores, idosos saudáveis simuladores e idosos com diagnóstico de demência); ii) realizar um estudo comparativo dos diferentes níveis de dificuldade do CIH-EV e do desempenho total no CIH-EV, quer relativamente à precisão, quer em relação ao tempo de resposta entre os três grupos; iii) analisar e determinar os pontos de corte do CIH-EV, quer para idosos saudáveis, quer para idosos com demência, identificando-se os seus valores de sensibilidade e especificidade; e iv) realizar um estudo de validade convergente do CIH-EV com as versões portuguesas do *Test of Memory Malingering* (TOMM; Fernandes et al., 2009) e do *Rey 15-Item Memory Test* (15-IMT; Simões et al., 2010; Simões et al., 2011) e a medida RDS do subteste de Memória de Dígitos da WAIS-III-R (Wechsler, 1997, 2008; Pinho, 2012).

Para levar a cabo os objetivos descritos, como referido anteriormente, recorreu-se a três grupos de adultos idosos: idosos saudáveis não-simuladores, instruídos a ter o melhor desempenho possível, idosos saudáveis simuladores, instruídos a simular um défice de memória típico de uma demência, com o objetivo de obter algum tipo de pensão, subsídio ou reforma antecipada e um grupo clínico, composto por adultos idosos com diagnóstico de demência. A todos os participantes foi administrado um protocolo de avaliação composto por um questionário sociodemográfico, o MMSE (Folstein et al., 1975; Guerreiro et al., 1994; Guerreiro, 1998), o MoCA (Freitas et al., 2011; Nasreddine et al., 2005), o Rey-15IMT (Boone et al., 2002; Rey, 1964; Simões et al., 2010), o TOMM (Fernandes, 2009; Mota, 2008;

Tombaugh, 1996), o subteste de Memória de Dígitos da WAIS-III-R (Wechsler, 1997; 2008) e o CIH-EV (Daugherty et al., no prelo).

Relativamente à análise da relação entre as variáveis demográficas e o desempenho no CIH-EV dos três grupos, que constituiu o primeiro objetivo deste estudo, os grupos simulador e não-simulador não diferiram entre si relativamente ao desempenho cognitivo (MMSE e MoCA), tal como esperado, uma vez que os participantes foram distribuídos aleatoriamente pelas condições. Contrariamente, observaram-se diferenças esperadas entre os grupos simulador e clínico nestas medidas, já que os últimos apresentam um diagnóstico de demência. Além disso, os dois grupos diferiram ao nível da habitação, também como esperado, dado que o grupo clínico era composto por indivíduos institucionalizados. Embora idealmente os grupos simulador e clínico não devessem diferir nas restantes variáveis demográficas, o grupo clínico revelou-se significativamente mais velho, diferença também reportada por Techneir e Wagner (2004) com uma amostra clínica de pessoas com diagnóstico de demência. De facto, em adultos idosos mais velhos (van der Flier & Scheltens, 2005) e institucionalizados (Matthews & Denning, 2002), verifica-se uma maior prevalência de demência, com maior incidência nas mulheres (Alzheimer's Society, 2016). Desta forma, o recurso a um grupo com demência, institucionalizado e maioritariamente do sexo feminino traduz-se, com grande probabilidade, numa idade média mais elevada. Não obstante, esta diferença vem fornecer maior robustez aos resultados encontrados, na medida em que, mesmo com um quadro demencial e uma idade significativamente mais avançada, o grupo clínico apresentou um melhor desempenho do que o grupo simulador de idosos saudáveis, resultado que será discutido com mais detalhe posteriormente.

No presente estudo, o CIH-EV (Daugherty et al., no prelo) mostrou-se insensível a todas as variáveis sociodemográficas e ao funcionamento neurocognitivo, quer em adultos idosos com demência, à semelhança do que Schroeder e colaboradores (2012) verificaram com o CIH

(Kapur, 1994.), quer em adultos idosos saudáveis. Pelo contrário, as restantes medidas de validade de desempenho utilizadas evidenciaram correlações com, pelo menos, uma variável sociodemográfica e um teste de desempenho cognitivo. Particularmente em relação ao Rey-15 IMT (Rey, 1964; Simões et al., 2010), em adultos idosos saudáveis, o Ensaio de Evocação Imediata e o Resultado Combinado mostraram-se correlacionados com a idade e a escolaridade, resultados que tinham sido já reportados por Strutt e colaboradores (2011) e Pinho (2012) com adultos idosos saudáveis. No presente estudo observou-se ainda uma correlação significativa entre o Ensaio de Evocação e a profissão. Do mesmo modo, o Rey-15IMT mostrou-se sensível ao funcionamento neurocognitivo quer em adultos idosos saudáveis, quer no grupo clínico. De facto, Fazio e colaboradores (2017) tinham já reportado que o Rey-15IMT era sensível a défices mnésicos e quadros demenciais, resultado corroborado pelo presente estudo. Em relação ao segundo ensaio do TOMM (Tombaugh, 1996, 1997), no grupo de idosos saudáveis não-simuladores observou-se uma correlação com o sexo. Este resultado tinha sido já verificado por Pinho (2012), porém, apenas para idosos saudáveis simuladores. Embora o desempenho do grupo de idosos com diagnóstico de demência no TOMM não tenha apresentado qualquer relação com as variáveis sociodemográficas neste estudo, Teichner e Wagner (2004) reportaram a influência da idade neste teste num grupo de adultos idosos com demência. Além disso, no presente estudo, o desempenho no primeiro ensaio do TOMM mostrou ser também sensível à idade em adultos idosos saudáveis simuladores. A conjugação dos resultados deste estudo com os dos referidos estudos anteriores mostram-se, portanto, inconclusivos, face ao tipo de variáveis que exercem influência nos resultados deste teste, porém, sugerem que mesmo um teste considerado *gold standard* (Fazio, Sanders & Denney, 2015; Slick et al., 1999) pode ser sensível a, pelo menos, uma variável demográfica em adultos idosos saudáveis e com demência. Finalmente, relativamente ao desempenho de adultos idosos saudáveis na RDS, este mostrou-se sensível à escolaridade, de forma concordante com o reportado por Choi et al.,

(2014) e à profissão em adultos idosos saudáveis não-simuladores. Pinho (2012) não encontrou estas relações, mas observou uma correlação adicional entre a RDS e a idade num grupo de idosos saudáveis. Além disso, observou também uma influência da escolaridade nesta medida, mas apenas para um grupo de idosos com DCL, o que não se observou no grupo de idosos com demência no presente estudo. Por um lado, os resultados do presente estudo apontam para a importância de considerar a profissão na avaliação da influência de variáveis sociodemográficas no desempenho dos testes, sendo que esta raramente é integrada e, por outro lado, fornecem evidência sólida para considerar o CIH-EV (Daugherty et al., no prelo) como um instrumento válido para a detecção de simulação, uma vez que foi o único que não se mostrou sensível a qualquer tipo de variável sociodemográfica nem ao funcionamento neurocognitivo, quer em adultos idosos saudáveis, quer em adultos idosos com demência.

Para cumprir o segundo objetivo, realizou-se o estudo comparativo dos desempenhos nos três níveis de dificuldade e no desempenho total no CIH-EV entre os três grupos. Como esperado (H1), o grupo de adultos idosos não-simulador e o grupo clínico apresentaram um melhor desempenho do que o grupo simulador, quer entre os níveis de dificuldade do CIH-EV, quer no total de acertos, corroborando, assim, a primeira hipótese do presente estudo. O facto de não se terem verificado diferenças de precisão de desempenho entre os níveis de dificuldade sugere que, em adultos idosos, se possa administrar apenas um deles. Por um lado, o CIH-EV (Daugherty et al., no prelo) tinha já demonstrado boa capacidade discriminativa entre grupos de estudantes universitários simuladores e não-simuladores, sendo que os últimos revelaram também um desempenho significativamente menos preciso no número total de acertos e entre os níveis de dificuldade. Por outro lado, usando a versão original do CIH (Kapur, 1994) com grupos de idosos saudáveis (não-simuladores e simuladores) e de idosos com demência, Yeh e colaboradores (2019) haviam já reportado o desempenho superior dos grupos não-simulador e clínico com demência relativamente ao não-simulador. Desta forma, o presente estudo vem

replicar o padrão de resultados obtidos com a versão original do CIH (Kapur, 1994) em adultos idosos saudáveis e com demência, utilizando a versão extensa e informatizada do teste (CIH-EV; Daugherty et al., no prelo). Além disso, os presentes resultados vêm corroborar também quer o cumprimento do pressuposto do efeito teto no CIH-EV (Daugherty et al., no prelo), na medida em que adultos idosos com demência obtiveram resultados mais precisos do que os adultos idosos simuladores, quer a inferência dos comportamentos de simulação intencionais associados a desempenhos significativamente mais baixos (Vickery et al., 2001). Estes aspetos reforçam a adequação do CIH-EV como um bom Teste de Validade de Desempenho. Não obstante, verificou-se que os adultos idosos com demência apresentaram tempos de resposta significativamente superiores aos grupos de idosos saudáveis, resultado que não deixa de ser expectável, considerando que nestes quadros clínicos, funções como a velocidade de processamento e a velocidade de execução se encontram significativamente mais comprometidas e, assim, poderão influenciar o tempo de resposta (e.g., Gainotti, Marra & Villa, 2001).

Relativamente ao terceiro objetivo do presente estudo, a determinação do ponto de corte mais adequado ao grupo de idosos saudáveis, verificou-se, como esperado (H2), um mesmo valor ou próximo ao observado para jovens adultos na discriminação entre simuladores e não-simuladores (Daugherty et al., no prelo). Particularmente para o número de acertos nos níveis de dificuldade, o ponto de corte mais adequado foi o mesmo para os três níveis (≤ 8), traduzindo sensibilidades entre 97% e 100% e especificidades entre 93% e 97%. De facto, este foi o mesmo ponto de corte sugerido para estudantes universitários (Daugherty et al., no prelo), demonstrando sensibilidades entre 84% e 95% e especificidades de 95% entre os níveis no CIH-EV. Utilizando a versão original do CIH (Kapur, 1994), Yeah e colaboradores (2019) determinaram igualmente um ponto de corte ≤ 8 para adultos idosos saudáveis. Em relação ao número total de acertos, o ponto de corte sugerido por Daugherty e colaboradores (no prelo)

para os jovens adultos (≤ 27) revelou 95% de sensibilidade e de especificidade. A opção por este valor para idosos saudáveis do presente estudo mostra-se válida, com 100% de sensibilidade e 90% de especificidade. Porém, optando por um ponto de corte inferior, mas próximo (≤ 26), é possível aumentar a especificidade para 97% e manter a sensibilidade de 100%. Desta forma, embora se possa optar pelo mesmo ponto de corte que o dos jovens adultos, sugere-se a adoção de um ponto de corte de ≤ 26 , mais conservador, indo ao encontro da maior relevância dada à especificidade comparativamente à sensibilidade defendida na literatura (e.g., Rezneck, 2005). De salientar que a semelhança encontrada entre os pontos de corte para ambas as amostras saudáveis (jovens adultos e adultos idosos) corroboram a evidência de que o CIH (Kapur, 1994) é insensível a variáveis como a idade (e.g., Ferreira et al., 2017; Schroeder et al., 2012), resultado verificado no presente estudo para a versão extensa do instrumento (CIH-EV; Daugherty et al., no prelo).

Os pontos de corte foram também determinados para adultos idosos com diagnóstico de demência e, tal como esperado (H3), revelaram-se inferiores aos valores sugeridos para os adultos idosos saudáveis não-simuladores. Para o primeiro e terceiro níveis de dificuldade, o ponto de corte mais adequado foi ≤ 5 , traduzindo uma sensibilidade de 94% e uma especificidade de 79% e 83%, respetivamente. Para o segundo nível, o valor selecionado (≤ 6) representou 88% de sensibilidade e 97% de especificidade. Adicionalmente, para o número total de acertos nos idosos com diagnóstico de demência observou-se um ponto de corte inferior (≤ 17) ao dos idosos saudáveis (≤ 26) demonstrando uma sensibilidade de 94% e uma especificidade de 97%. De facto, a necessidade de ajustamento dos pontos de corte para o grupo com demência traduz a relação inversa entre a precisão de diagnóstico e o grau de severidade do défice cognitivo reportada na literatura (McGuire et al., 2019; Rudman et al., 2011). Ainda assim, o ajustamento dos pontos de corte torna-se, para diversos testes, inviável, na medida em que representa uma sensibilidade extremamente reduzida (Dean et al., 2009). Este não é,

porém, o caso do CIH-EV (Daugherty et al., no prelo) que, no presente estudo, mostrou valores de sensibilidade e de especificidade que cumprem largamente o critério de, pelo menos, 50% de sensibilidade e de, pelo menos, 90% de especificidade proposto por Sugarman e Axelrod (2015).

Finalmente, de acordo com o quarto objetivo do presente estudo, foi analisada a validade convergente do CIH-EV (Daugherty et al., no prelo) com as restantes medidas de validade de desempenho (Rey-15IMT; Boone et al., 2002; Rey, 1964; Simões et al., 2010; TOMM; Fernandes, 2009; Mota, 2008; Tombaugh, 1996; e RDS; Wechsler, 1997; 2008). Como esperado (H4), o CIH-EV apresentou boa validade convergente com todos os testes administrados. Observou-se, no entanto, uma correlação baixa com o Rey-15IMT. Este resultado não se revela surpreendente se considerarmos que este último teste apresenta valores de sensibilidade reduzidos, que limitam a sua utilidade na deteção de simulação (e.g., Reznick, 2005), razão pela qual se correlaciona de forma reduzida também com outros TVDs (e.g., TOMM; Lopes, 2017). As correlações mais elevadas foram observadas com o TOMM, à semelhança do que Yeh e colaboradores (2019) reportaram também com adultos idosos saudáveis e com demência, o que se justifica pela comprovada precisão de diagnóstico deste instrumento (e.g., Barhon et al., 2015). Além disso, o facto de se terem observado correlações moderadas entre o CIH-EV e o RDS sugere que, tal como Miele e colaboradores (2012) reportaram, esta última é uma medida de validade de desempenho *embedded* com boa capacidade de deteção de simulação. Daugherty e colaboradores (no prelo) verificaram igualmente a boa validade convergente do CIH-EV com estes instrumentos.

No entanto, estudos anteriores sugerem que tanto o TOMM (Dean et al., 2009; Rudman et al., 2009; Teichner & Wagner, 2004; Swier-Vosnos & Pliskin, 2014) como a RDS (Dean et al., 2009; Loring et al., 2016; Pinho, 2012) e o Rey-15IMT (Bortnik et al., 2013; Fazio et al., 2017; Rudman et al., 2011) mostram uma redução da precisão de diagnóstico quando aplicadas

a grupos com demência, originando uma percentagem de falsos positivos inaceitavelmente elevada (Rudman et al., 2011). Esperou-se (H4) que o CIH-EV (Daugherty et al., no prelo) apresentasse menor percentagem de falsos positivos do que os restantes testes. De facto, nos idosos saudáveis, o CIH-EV detetou corretamente 100% dos não-simuladores e 96.6% dos simuladores, à semelhança do TOMM, pelo que ambos têm especificidades e sensibilidades muito elevadas. No entanto, em idosos com demência, verificou-se que o TOMM apresentou uma percentagem de falsos positivos inaceitavelmente alta para ambos os ensaios (29.4% e 64.7%, respetivamente, sendo que o recomendável é até 10%; Sugerman & Axelrod, 2015). Se, por um lado, o Resultado Combinado do Rey-15IMT apresentou uma percentagem de falsos positivos nula para o grupo de idosos com demência, por outro, apresentou uma percentagem muito elevada (23.3%) para o grupo de idosos saudáveis (23.3%). Além disso, as análises permitiram verificar a inadequabilidade do Ensaio de Evocação nos adultos idosos saudáveis, para identificar idosos não-simuladores (48.4%), idosos simuladores (3,4%) e adultos idosos com demência, para os quais revela uma percentagem de 29.4% de falsos positivos. Relativamente à RDS, embora nos adultos idosos saudáveis tenha mostrado relativamente bons valores de especificidade (96.7%) e de sensibilidade (79.3%), para adultos idosos com demência apresentou uma percentagem de falsos positivos inaceitavelmente elevada (35.3%). Desta forma, o TOMM e a RDS mostraram-se eficazes na deteção de simulação de adultos idosos saudáveis, mas não de adultos idosos com diagnóstico de demência. Para este grupo clínico, o Resultado Combinado do Rey-15IMT mostrou-se adequado, mas não para adultos idosos saudáveis. Estes resultados apoiam a necessidade que o clínico tem de conhecer o tipo de grupos em que os TVD são adequados e aqueles em que não devem ser administrados (Dean et al., 2009), para que a deteção de simulação possa ser válida e precisa. Além disso, os resultados demonstram que o único teste que apresentou valores elevados de especificidade e de sensibilidade, bem como reduzidas percentagens de

falsos positivos quer em adultos idosos saudáveis, quer em adultos idosos com demência foi o CIH-EV (Daugherty et al., no prelo), o que indica que este teste é válido para a detecção de simulação de jovens adultos saudáveis (Daugherty et al., no prelo), adultos idosos saudáveis e com demência (Yeh et al., 2019) com a garantia de boa precisão de diagnóstico.

O presente estudo teve ainda como objetivo analisar a adequabilidade dos pontos de corte sugeridos na literatura para adultos idosos com demência nos restantes testes TVD (Rey-15IMT; Boone et al., 2002; Rey, 1964; Simões et al., 2010; TOMM; Fernandes, 2009; Mota, 2008; Tombaugh, 1996; e RDS; Wechsler, 1997; 2008). No presente estudo, o Rey-15IMT (Simões et al., 2010) apresentou uma capacidade discriminativa muito reduzida, quer para o Ensaio de Evocação Imediata, quer para o Resultado Combinado do Reconhecimento, pelo que não foi possível determinar pontos de corte. Assim, este resultado sugere a inadequação da utilização do Rey-15IMT em adultos idosos com demência, contrariamente ao observado por Simões e colaboradores (2010) com adultos idosos com problemas cognitivos.

Relativamente ao TOMM (Tombaugh, 1996; Mota, 2008; Fernandes, 2009), apresentou uma capacidade discriminativa elevada, verificando-se a necessidade de ajustar os pontos de corte para adultos idosos com demência. No primeiro ensaio, utilizando um ponto de corte ≤ 32 , obteve-se uma sensibilidade de 71% e uma especificidade de 97%. Fernandes (2009) e Pinho (2012), haviam sugerido um valor de ≤ 33 para idosos com DCL para uma especificidade de 100%, no entanto, não referem o valor de sensibilidade do mesmo. No presente estudo, o ponto de corte ≤ 32 obteve, simultaneamente, valores adequados para a sensibilidade e especificidade, que cumprem o critério de Sugerman e Axelrod (2015) e, por isso, foi o valor proposto. Para o segundo ensaio, o ponto de corte tradicional < 45 não se mostrou adequado, ao contrário do que acontece para adultos idosos saudáveis e com DCL (Jesus, 2013; Pinho, 2012), em conformidade com o sugerido por Teichner e Wagner (2004) e Walter e colaboradores (2014). Assim, para adultos idosos com demência há a necessidade de ajustar o ponto de corte, para

≤34, no segundo ensaio do TOMM, de forma a obter uma sensibilidade de 94% e uma especificidade de 97%.

Relativamente à RDS (Wechsler, 1997, 2008; Pinho, 2012), embora Pinho (2012) tenha proposto um ponto de corte ≤6 para adultos idosos com DCL (com 70.59% de especificidade), este valor não se revelou adequado para adultos idosos com demência. Houve, assim, a necessidade de ajustar o valor para ≤5, que traduziu uma sensibilidade de 65%% e uma especificidade de 79%. De facto, Loring e colaboradores (2016) verificaram que entre os pontos de corte ≤7, ≤6 e ≤5, o último era o mais adequado para o grupo de idosos com demência, permitindo uma percentagem de falsos positivos mais baixa, resultados corroborados pelo presente estudo.

O padrão verificado para os três testes (Rey-15IMT; Boone et al., 2002; Rey, 1964; Simões et al., 2010; TOMM; Fernandes, 2009; Mota, 2008; Tombaugh, 1996; e RDS; Wechsler, 1997; 2008) revela que para idosos com demência há uma a necessidade de reduzir os pontos de corte relativamente aos adotados para adultos idosos saudáveis e com DCL . Esta necessidade está de acordo com o que Dean e colaboradores (2009) e McGuire e colaboradores (2019) defendem, relativamente à redução da adequabilidade dos pontos de corte em função da severidade do declínio cognitivo e da existência de demência. Assim, a determinação de pontos de corte ajustados revela-se indubitavelmente pertinente, tal como o presente estudo sugere.

O presente estudo contribuiu para que em Portugal sejam adotados pontos de corte adequados para adultos idosos com demência no TOMM (Fernandes, 2009; Mota, 2008; Tombaugh, 1996), na RDS (Wechsler, 1997, 2008; Pinho, 2012) e no CIH-EV (Daugherty et al., no prelo), para além de demonstrar a inadequabilidade do uso do Rey-15IMT (Boone et al., 2002; Rey, 1964; Simões et al., 2010) neste grupo.

Não obstante, o presente estudo apresenta algumas limitações. Em primeiro lugar, a amostra clínica apresenta um tamanho reduzido e maioritariamente feminino, o que pode ter implicações na generalização dos resultados para a população. Em segundo lugar, a amostra não contou com um grupo clínico que cumprisse critério(s) de Slick et al. (1999.), i.e., indivíduos com suspeita de simulação, incentivo externo ou ganho secundário. De facto, o grupo simulador foi instruído a simular, sendo que alguns autores (e.g., Rogers, 2008) afirmam que, nestas condições, por não terem um ganho secundário a motivá-los, tal como acontece com os simuladores reais, os participantes não conseguem simular exatamente, o que dificulta a generalização dos resultados para a população. No entanto, por exemplo, Haines e Norris (1995) incorporaram um incentivo financeiro no grupo simulador e verificaram que não existiam diferenças entre os simuladores com incentivo e os simuladores apenas instruídos. Além disso, relativamente ao TOMM, embora estudos anteriores (e.g., Larrabee, Greiffenstein, Greve, & Bianchini, 2007) sugiram que a administração do Ensaio de Retenção permite reduzir o número de falsos positivos e que a sua não administração pode implicar não detetar cerca de 3% de casos suspeitos de simulação, Tombaugh (1996) e O'Bryan e colaboradores (2008) reportaram que a aplicação dos dois ensaios de aprendizagem se mostra suficiente para detetar a presença de comportamentos de simulação. De facto, o protocolo de avaliação utilizado em jovens adultos (Daugherty et al., no prelo) recorreu apenas a ambos os ensaios de aprendizagem, uma vez que i) existe a necessidade de economizar os protocolos de avaliação (neuro)psicológica (Boone, 2007) e, ii) foram administradas outras medidas de deteção de simulação (Rey-15IMT; Boone et al., 2002; Rey, 1964; Simões et al., 2010; CIH-EV; Daugherty et al., no prelo; e RDS; Wechsler, 1997; 2008), razões que justificaram a opção de administrar apenas os ensaios de aprendizagem no presente estudo.

No estudo atual, poderia ter sido interessante verificar se existiam diferenças de desempenho nos testes em função do tipo de demência sendo esta uma sugestão de investigação

futura. Da mesma forma, em estudos futuros poderá ser pertinente incluir grupos clínicos com diferentes níveis de deterioração cognitiva, de modo a perceber com que nível de deterioração o ponto de corte deixa de ser semelhante ao de idosos saudáveis. Além disso, incluir um quarto grupo, de simuladores clínicos, como proposto por Rogers (2008) em estudos futuros permitirá aumentar a generalização dos resultados para a população. De forma mais particular, poderá ser pertinente realizar uma análise qualitativa das estratégias de simulação aplicadas no Rey-15IMT (Strutt, Scott, Shrestha e York, 2011), de modo a ter uma compreensão mais abrangente sobre a sua insensibilidade.

Em síntese, a validação do CIH-EV (Daugherty et al., no prelo) em adultos idosos saudáveis e com diagnóstico de demência revelou-se extremamente pertinente. De facto, veio possibilitar a existência de um Teste de Validade de Desempenho adicional em Portugal. Verificou-se que os grupos não-simulador e clínico tiveram um desempenho mais preciso no CIH-EV do que o grupo simulador, com o grupo clínico a apresentar tempos de resposta superiores; determinou-se o ponto o ponto de corte (≤ 26) para adultos idosos saudáveis e para adultos idosos com diagnóstico de demência (≤ 17), com elevados valores de sensibilidade e especificidade; verificou-se a existência de boa validade convergente com as diferentes medidas de validade de desempenho utilizadas (TOMM, Rey-15IMT e RDS); observou-se que o CIH-EV não é sensível às variáveis sociodemográficas e ao funcionamento neurocognitivo em idosos saudáveis nem em idosos com demência; e, por fim, demonstrou ser o instrumento com maior precisão de diagnóstico nos três grupos utilizados. Além disso, o presente estudo corroborou ainda a necessidade de adequação dos pontos de corte dos Testes de Validade de Desempenho (TVDs) nestas amostras e, particularmente, propôs os pontos de corte adequados para os TVDs validados em Portugal para adultos idosos.

Referências

- Almeida, D. (2015). *Comportamento de Simulação e exagero de sintomas em contexto médico-legal: Novos Estudos de validação do Structured Inventory of Malingered Symptomatology (SIMS)*. Dissertação de Mestrado em Psicologia, especialidade em Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Alzheimer Portugal. (2017). *Portugal é o 4º país da OCDE com mais pessoas com demência por cada 1000 habitantes*. Retirado de http://alzheimerportugal.org/pt/news_text-77-8-827-portugal-e-o-4-pais-da-ocde-com-mais-pessoas-com-demencia-por-cada-1000-habitante.
- Alzheimer's Society. (2016). *Risk factors for dementia (Factsheet 450LP)*. England: Alzheimer's Society.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Arlington, VA, US: American Psychiatric Publishing, Inc.
- American Psychiatric Association. (2014). *Manual de Diagnóstico e Estatístico das Perturbações Mentais* (5ª Edição). Lisboa: Climepsi Editores.
- An, K. Y., Kaploun, K., Erdodi, L. A., & Abeare, C. A. (2017). Performance validity in undergraduate research participants: A comparison of failure rates across tests and cutoffs. *The Clinical Neuropsychologist*, 31(1), 193-206.
- Arnett, P. A., Hammeke, T. A., & Schwartz, L. (1995). Quantitative and qualitative performance on Rey's 15-Item Test in neurological patients and dissimulators. *The Clinical Neuropsychologist*, 9(1), 17-26.

- Ashendorf, L., Constantinou, M., & McCaffrey, R. J. (2004). The effect of depression and anxiety on the TOMM in community-dwelling older adults. *Archives of Clinical Neuropsychology, 19*(1), 125-130.
- Axelrod, B. N., Fichtenberg, N. L., Millis, S. R., & Wertheimer, J. C. (2006). Detecting incomplete effort with digit span from the Wechsler Adult Intelligence Scale—Third Edition. *The Clinical Neuropsychologist, 20*(3), 513-523.
- Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*(1), 145-159.
- Barhon, L. I., Batchelor, J., Meares, S., Chekaluk, E., & Shores, E. A. (2015). A comparison of the degree of effort involved in the TOMM and the ACS Word Choice Test using a dual-task paradigm. *Applied Neuropsychology: Adult, 22*(2), 114-123.
- Bauer, L., O'Bryant, S. E., Lynch, J. K., McCaffrey, R. J., & Fisher, J. M. (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.
- Beetar, J. T., & Williams, J. M. (1995). Malingering response styles on the Memory Assessment Scales and symptom validity tests. *Archives of Clinical Neuropsychology, 10*(1), 57-72.
- Benton, A. L., Sivan, A. B., Hamsher, K. D., Varney, N. R., & Spreen, O. (1994). *Contributions to neuropsychological assessment: A clinical manual*. New York: Oxford University Press.

- Bernard, L. C., & Fowler, W. (1990). Assessing the validity of memory complaints: Performance of brain-damaged and normal individuals on Rey's task to detect malingering. *Journal of Clinical Psychology, 46*(4), 432-436.
- Bernard, L. C., Houston, W., & Natoli, L. (1993). Malingering on neuropsychological memory tests: Potential objective indicators. *Journal of Clinical Psychology, 49*(1), 45-53.
- Bernard, L. C., McGrath, M. J., & Houston, W. (1996). The differential effects of simulating malingering, closed head injury, and other CNS pathology on the Wisconsin Card Sorting Test: Support for the "pattern of performance" hypothesis. *Archives of Clinical Neuropsychology, 11*(3), 231-245.
- Bigler, E. D. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(4), 632-640.
- Binder, L.M. (1990). Malingering following minor head trauma. *The Clinical Neuropsychologist, 4*, 25-36.
- Binder, L. M. (1993). An abbreviated form of the Portland digit recognition test. *The Clinical Neuropsychologist, 7*(1), 104-107.
- Binder, L. M., & Kelly, M. P. (1996). Portland Digit Recognition Test performance by brain dysfunction patients without financial incentives. *Assessment, 3*(4), 403-409.
- Binder, L. M., Villanueva, M. R., Howieson, D., & Moore, R. T. (1993). The Rey AVLT recognition memory task measures motivational impairment after mild head trauma. *Archives of Clinical Neuropsychology, 8*(2), 137-147.
- Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*(2), 175-181.

- Boone, K. B. (Ed.). (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford.
- Boone, K. B. (2008). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, 23, 729–741.
- Boone, K. B., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., ... & Berman, N. G. (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. B., Lu, P., & Wen, J. (2005). Comparison of various RAVLT scores in the detection of noncredible memory performance. *Archives of Clinical Neuropsychology*, 20(3), 301-319.
- Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). 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.
- Bortnik, K. E., Horner, M. D., & Bachman, D. L. (2013). Performance on standard indexes of effort among patients with dementia. *Applied Neuropsychology: Adult*, 20(4), 233-242.
- Borum, R., Otto, R., & Golding, S. (1993). Improving clinical judgment and decision making in forensic evaluation. *The Journal of Psychiatry & Law*, 21(1), 35-76.
- Bounds, T. A. (2004). *The Test of Memory Malingering: Detection of malingered cognitive deficits in a criminal forensic population*. University of Southern Mississippi.
- Brandt, J. (1988). Malingered amnesia. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 65-83). New York: Guilford Press.

- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., ... & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, 20(4), 419-426.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis, MN: Univer.
- Castro, V. S. S. (2015). *Deteção da simulação em indivíduos com e sem diagnóstico de depressão*. Dissertação de Mestrado em Psicologia da Saúde e Reabilitação Neuropsicológica. Faculdade de Psicologia da Universidade de Aveiro.
- Choi, H. J., Lee, D. Y., Seo, E. H., Jo, M. K., Sohn, B. K., Choe, Y. M., ... & Jhoo, J. H. (2014). A normative study of the digit span in an educationally diverse elderly population. *Psychiatry investigation*, 11(1), 39-43.
- Cima, M., Merckelback, H., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R., & Muris, P. (2003). The other side of malingering: Supernormality. *The Clinical Neuropsychologist*, 17(2), 235-243.
- Cochrane, H. J., Baker, G. A., & Meudell, P. R. (1998). Simulating a memory impairment: Can amnesics implicitly outperform simulators?. *British Journal of Clinical Psychology*, 37(1), 31-48.
- Colwell, K., & Sjerven, E. R. (2005). The "Coin-in hand" stratagem for the forensic assessment of malingering. *American Journal of Forensic Psychology*, 23(1), 83-90.
- Conder, R., Allen, L., & Cox, D. (1992). *Computerized assessment of response bias test manual*. Durham, NC: CogniSyst.

- Cruz, A. P. S. (2008). *Quatro estudos exploratórios com o Test of Memory Malingering (TOMM)*. Dissertação de Mestrado em Avaliação Psicológica. Coimbra: Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra.
- Cunha, C. (2011). *Escala de Validade de Sintomas – Versão Experimental (EVS-VE): Estudos de “simulação” e validação em amostra da comunidade*. Dissertação de Mestrado em Psicologia, especialidade em Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of clinical and experimental neuropsychology*, 39(3), 257-264.
- Dandachi-FitzGerald, B., Ponds, R. W., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology*, 28(8), 771-783.
- Daugherty, J. C., Hidalgo-Ruzzante, N., Palma-Navarro, A. B., Bencrimo, D., & García, M. P. (2017). Versión computarizada del test de esfuerzo Coin in Hand: Datos preliminares. Unpublished. <https://doi.org/10.13140/RG.2.2.26697.03689>
- Daugherty, J., Querido, L., Quiroz, N., Wang, D., Hidalgo-Ruzzante, N., Fernandes, S., ... & Valera, E. (no prelo). The Coin in Hand-Extended Version: Development and validation of a multicultural performance. *Assessment*.
- Davidson, H., Suffield, B., Orenczuk, S., Nantau, K., & Mandel, A. (1991, February). Screening for malingering using the Memory for Fifteen Items Test (MFIT). Poster

presented at the International Neuropsychological Society 19th Annual Meeting, San Antonio.

Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M., & Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist*, 23(1), 133-152.

Delain, S. L., Stafford, K. P., & Ben-Porath, Y. S. (2003). Use of the TOMM in a criminal court forensic assessment setting. *Assessment*, 10(4), 370-381.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult Version*. San Antonio, TX: The Psychological Corporation.

Denning, J. H. (2012). The efficiency and accuracy of the Test of Memory Malingering Trial 1, errors on the first 10 items of the Test of Memory Malingering, and five embedded measures in predicting invalid test performance. *Archives of Clinical Neuropsychology*, 27, 417-432.

Diário da República. (2018). Trabalho, solidariedade e segurança social (1.^a série, N.º 13, 18 de janeiro de 2018). Governo Português.

Domingues, A. F. (2012). Escala de Validade de Sintomas – Versão Experimental (EVS-VE): Estudo de validação em amostras médico-legal e da comunidade. Dissertação de Mestrado em Psicologia, especialidade em Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.

Duncan, A. (2005). The impact of cognitive and psychiatric impairment of psychotic disorders on the Test of Memory Malingering (TOMM). *Assessment*, 12(2), 123-129.

Duncan, S. A., & Ausborn, D. L. (2002). The use of reliable digits to detect malingering in a criminal forensic pretrial population. *Assessment*, 9(1), 56-61.

- Etherton, J. L., Bianchini, K. J., Greve, K. W., & Heinly, M. T. (2005). Sensitivity and specificity of reliable digit span in malingered pain-related disability. *Assessment, 12*(2), 130-136.
- Faust, D., & Ackley, M.A. (1998). Did you think it was going to be easy? Some methodological suggestions for the investigation and development of malingering detection techniques. In C. R. Reynolds (Ed.), *Detection of malingering during head injury litigation* (pp. 261-286). New York: Plenum Press.
- Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56*, 578-582.
- Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19*(5), 508-515.
- Faustino, J. D. P. R. (2008). Violência, funcionamento cognitivo e “esforço reduzido”: Estudos com um grupo de reclusos. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Fazio, R. L., Faris, A. N., & Yamout, K. Z. (2017). Use of the Rey 15-Item Test as a performance validity test in an elderly population. *Applied Neuropsychology: Adult, 26*(1), 28-35.
- Fazio, R. L., Sanders, J. F., & Denney, R. L. (2015). Comparison of performance of the Test of Memory Malingering and Word Memory Test in a criminal forensic sample. *Archives of Clinical Neuropsychology, 30*(4), 293-301.

- Ferguson, K. E. (2004). Detecting malingering in forensic neuropsychological evaluations in litigants with mild traumatic brain injury. In W. T. O'Donohue & E. R. Levensky (Eds.), *Handbook of forensic psychology* (pp. 301-314). London, UK: Elsevier Academic Press.
- Fernandes, S. P. (2009). Test of Memory Malingering (TOMM): Estudos de validação em adultos idosos com Declínio Cognitivo Ligeiro. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicogerontologia Clínica. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Ferreira, I., Gomes, F., Moreira, I., Silva, A., & Cavaco, S. (2017). Medir a validade da avaliação neuropsicológica em doentes com esclerose múltipla ou lúpus eritematoso sistémico. *Revista Evidências*, 27.
- Fisher, H. L., & Rose, D. (2005). Comparison of the effectiveness of two versions of the Rey memory test in discriminating between actual and simulated memory impairment, with and without the addition of a standard memory test. *Journal of clinical and experimental neuropsychology*, 27(7), 840-858.
- Folstein, M., Folstein, M., & McHugh, P. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189-198.
- Fonseca, M. S. J. (2009). Comportamentos de “esforço reduzido” numa amostra de traumatismos crânio-encefálicos avaliados em contexto médico-legal. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.

- Franzen, M. D., & Martin, N. (1996). Do people with knowledge fake better?. *Applied Neuropsychology*, 3, 82–85.
- Frederick, R. I. (1997). Validity indicator profile manual. Minnetonka, MN: NCS Assessments.
- Frederick, R. I., Crosby, R. D., & Wynkoop, T. F. (2000). Performance curve classification of invalid responding on the Validity Indicator Profile. *Archives of Clinical Neuropsychology*, 15(4), 281-300.
- Frederick, R. I., Sarfaty, S. D., Johnston, J. D., & Powel, J. (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology*, 8(1), 118-125.
- Frederick, R. I., & Speed, F. M. (2007). On the interpretation of below-chance responding in forced-choice tests. *Assessment*, 14(1), 3-11.
- Freitas, S., Simões, M. R., Alves, L., Duro, D., & Santana, I. (2012a). Montreal Cognitive Assessment (MoCA): Validation study for Frontotemporal Dementia. *Journal of Geriatric Psychiatry and Neurology*, 25(3), 146-154.
- Freitas, S., Simões, M. R., Alves, L., & Santana, I. (2011). Montreal Cognitive Assessment (MoCA): Normative study for the portuguese population. *Journal of Clinical and Experimental Neuropsychology*, 33(9), 989-996.
- Freitas, S., Simões, M. R., Alves, L., & Santana, I. (2013). Montreal Cognitive Assessment (MocA): Validation study for mild cognitive impairment and Alzheimer's disease. *Alzheimer Disease and Associated Disorders*, 27(1), 37-43.
- Freitas, S., Simões, M. R., Alves, L., Vicente, M., & Santana, I. (2012b). Montreal Cognitive Assessment (MoCA): Validation study for vascular dementia. *Journal of International Neuropsychology Society*, 18(6), 1031-1040.

- Gainotti, G., Marra, C., & Villa, G. (2001). A double dissociation between accuracy and time of execution on attentional tasks in Alzheimer's disease and multi-infarct dementia. *Brain, 124*(4), 731-738.
- Garcia, S. M. G. (2011). Test of Memory Malinger (TOMM) e Rey-15 Item Memory Test (FIT): Estudos de validação numa amostra de adolescentes institucionalizados em Centros Educativos. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Gavett, B. E., O'Bryant, S. E., Fisher, J. M., & McCaffrey, R. J. (2005). Hit rates of adequate performance based on the test of memory malinger (TOMM) Trial 1. *Applied Neuropsychology, 12*(1), 1-4.
- Goldberg, T. O., & Miller, H. R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of Clinical Psychology, 42*, 792-795.
- Graue, L. O., Berry, D. T. R., Clark, J. A., Sollman, M. J., Cardi, M., Hopkins, J., & Werline, D. (2007). Identification of feigned mental retardation using the new generation of malinger detection instruments: Preliminary findings. *The Clinical Neuropsychologist, 21*, 929-942.
- Green, C. M., Kirk, J. W., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The use of the Rey 15-Item Test and recognition trial to evaluate noncredible effort after pediatric mild traumatic brain injury. *Journal of clinical and experimental neuropsychology, 36*(3), 261-267.
- Green, P., Iverson, G. L., & Allen, L. (1999). Detecting malinger in head injury litigation with the Word Memory Test. *Brain Injury, 13*(10), 813-819.

- Green, P., Lees-Haley, P. R., & Allen, L. M. (2003). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology*, 2(3-4), 97-124.
- Grégoire, J., & Van Der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, neuropsychology, and cognition*, 4(2), 140-149.
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6(3), 218-224.
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1996). Comparison of multiple scoring methods for Rey's malingered amnesia measures. *Archives of Clinical Neuropsychology*, 11(4), 283-293.
- Greiffenstein, M. F., Fox, D., & Lees-Haley, P. R. (2007). The MMPI-2 Fake Bad Scale in detection of noncredible brain injury claims. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 210–235). NY: Guilford Publications.
- Greve, K. W., & Bianchini, K. J. (2006). Should the Retention trial of the Test of Memory Malingering be optional?. *Archives of Clinical Neuropsychology*, 21(1), 117-119.
- Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-groups analysis. *Journal of Clinical and Experimental Neuropsychology*, 28(7), 1176-1190.
- Greve, K. W., Bianchini, K. J., Etherton, J. L., Meyers, J. E., Curtis, K. L., & Ord, J. S. (2010). The Reliable Digit Span test in chronic pain: Classification accuracy in detecting malingered pain-related disability. *The Clinical Neuropsychologist*, 24(1), 137-152.

- Griffin, G. E., Glassmire, D. M., Henderson, E. A., & McCann, C. (1997). Rey II: Redesigning the Rey screening test of malingering. *Journal of Clinical Psychology, 53*(7), 757-766.
- Grote, C. L., Kooker, E. K., Garron, D. C., Nyenhuis, D. L., Smith, C. A., & Mattingly, M. L. (2000). Performance of compensation seeking and non-compensation seeking samples on the Victoria Symptom Validity Test: Cross-validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology, 22*(6), 709-719.
- Guilmette, T. J., Hart, K. J., & Giuliano, A. J. (1993). Malingering detection: The use of a forced-choice method in identifying organic versus simulated memory impairment. *The Clinical Neuropsychologist, 7*(1), 59-69.
- Guilmette, T. J., Hart, K. J., Giuliano, A. J., & Leininger, B. E. (1994). Detecting simulated memory impairment: Comparison of the Rey Fifteen-Item Test and the Hiscock forced-choice procedure. *The Clinical Neuropsychologist, 8*(3), 283-294.
- Hampson, N. E., Kemp, S., Coughlan, A. K., Moulin, C. J., & Bhakta, B. B. (2014). Effort test performance in clinical acute brain injury, community brain injury, and epilepsy populations. *Applied Neuropsychology: Adult, 21*(3), 183-194.
- Hanley, J. R., Baker, G. A., & Ledson, S. (1999). Detecting the faking of amnesia: A comparison of the effectiveness of three different techniques for distinguishing simulators from patients with amnesia. *Journal of Clinical and Experimental Neuropsychology, 21*(1), 59-69.
- Hays, J. R., Emmons, J., & Lawson, K. A. (1993). Psychiatric norms for the Rey 15-item visual memory test. *Perceptual and Motor Skills, 76*(3), 1331-1334.

- Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*, 892-900.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*(7), 1093-1129.
- Hill, S. K., Ryan, L. M., Kennedy, C. H., & Malamut, B. L. (2003). The relationship between measures of declarative memory and the Test of Memory Malingering in patients with and without temporal lobe dysfunction. *Journal of Forensic Neuropsychology, 3*(3), 1-18.
- Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11*(6), 967-974.
- Hurley, K. E., & Deal, W. P. (2006). Assessment instruments measuring malingering used with individuals who have mental retardation: Potential problems and issues. *Mental Retardation, 44*(2), 112-119.
- Inman, T. H., & Berry, D. T. (2002). Cross-validation of indicators of malingering: A comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. *Archives of Clinical Neuropsychology, 17*(1), 1-23.
- Inman, T. H., Vickery, C. D., Berry, D. T., Lamb, D. G., Edwards, C. L., & Smith, G. T. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The letter memory test. *Psychological Assessment, 10*(2), 128-139.

Instituto Nacional de Estatística [INE]. (2017). *Projeções de População Residente: 2015 a 2080*. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística [INE]. (2018). *Anuário Estatístico de Portugal* (ISSN 0871-8741). Lisboa: Instituto Nacional de Estatística.

Iverson, G. L., & Franzen, M. D. (1994). The recognition memory test, digit span, and kox cube test as markers of malingered memory impairment. *Assessment, 1*(4), 323-334.

Iverson, G. L., & Franzen, M. D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology, 18*(1), 38-51.

Iverson, G. L., Franzen, M. D., & McCracken, L. M. (1991). Evaluation of an objective assessment technique for the detection of malingered memory deficits. *Law and Human Behavior, 15*(6), 667-676.

Iverson, G. L., Franzen, M. D., & McCracken, L. M. (1994). Application of a forced-choice memory procedure designed to detect experimental malingering. *Archives of Clinical Neuropsychology, 9*(5), 437-450.

Iverson, G. L., Page, J. L., Koehler, B. E., Shojania, K., & Badii, M. (2007). Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. *The Clinical Neuropsychologist, 21*(3), 532-546.

Jasinski, L. J., Berry, D. T., Shandera, A. L., & Clark, J. A. (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.

Jesus, S. A. S. (2013). Desempenho cognitivo não credível no Defeito Cognitivo Ligeiro: Estudos com o Test of Memory Malingering (TOMM) e o Rey-15 Item Memory Test

- (Rey 15-IMT). Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicogerontologia Clínica. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Kapur, N. (1994). The coin-in-the-hand test: a new "bed-side" test for the detection of malingering in patients with suspected memory disorder. *Journal of neurology, neurosurgery, and psychiatry*, 57(3), 385-386.
- Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region, and education differences on the 11 WAIS-R subtests. *Journal of clinical psychology*, 44(2), 231-248.
- Kelly, P. J., Baker, G. A., Van den Broek, M. D., Jackson, H., & Humphries, G. (2005). The detection of malingering in memory performance: The sensitivity and specificity of four measures in a UK population. *British Journal of Clinical Psychology*, 44(3), 333-341.
- Lages, H. J. (2016). Rey 15-Item Memory Test (Rey 15-IMT): Estudo de normalização para a população portuguesa de adultos e adultos idosos. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Larrabee, G. J. (1990). Cautions in the use of neuropsychological evaluations in legal settings. *Neuropsychology*, 4(4), 239-247.
- Larrabee, G. J. (1992). Interpretive strategies for evaluation of neuropsychological data in legal settings. *Forensic Reports*, 5(3), 257-264.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17(3), 410-425.

- Larrabee, G. J. (Ed.). (2005). *Forensic neuropsychology: A scientific approach*. New York: Oxford University Press.
- Larrabee, G. J. (Ed.). (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.
- Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18(4), 625–630.
- Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334–371). New York: Oxford University Press.
- Lee, G. P., Loring, D. W., & Martin, R. C. (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.
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York: Oxford University Press.
- Lezak, M. D., Howieson, D. B., & Loring, D. W. (1995). *Neuropsychological assessment*. New York: Oxford University Press.
- Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*. New York: Oxford University Press.
- Lippa, S. M. (2018). Performance validity testing in neuropsychology: a clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, 32(3), 391-421.
- Lopes, A. B. C. C. (2017). Rey 15–Item Memory Test (Rey 15-IMT) e Test of Memory Malingering (TOMM): Contributos para os estudos de validação e normalização para a

população portuguesa de jovens adultos e adultos. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.

Loring, D. W., Goldstein, F. C., Chen, C., Drane, D. L., Lah, J. J., Zhao, L., & Larrabee, G. J. (2016). False-positive error rates for reliable digit span and auditory verbal learning test performance validity measures in amnesic mild cognitive impairment and early Alzheimer disease. *Archives of Clinical Neuropsychology*, *31*(4), 313-331.

Love, C. M., Glassmire, D. M., Zanolini, S. J., & Wolf, A. (2014). Specificity and false positive rates of the Test of Memory Malingering, Rey 15-Item Test, and Rey Word Recognition Test among forensic inpatients with intellectual disabilities. *Assessment*, *21*(5), 618-627.

Maior, F. S. (2008). Structured Inventory of malingered symptomatology (SIMS) e Test of memory malingering (TOMM): Estudos de validação numa amostra normativa. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.

Marshall, P., & Happe, M. (2007). The performance of individuals with mental retardation on cognitive tests assessing effort and motivation. *The Clinical Neuropsychologist*, *21*, 826-840.

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist*, *24*, 1204-1237.

- Mathias, C. W., Greve, K. W., Bianchini, K. J., Houston, R. J., & Crouch, J. A. (2002). Detecting malingered neurocognitive dysfunction using the reliable digit span in traumatic brain injury. *Assessment, 9*(3), 301-308.
- Matthews, F. E., & Denning, T. (2002). Prevalence of dementia in institutional care. *The Lancet, 360*(9328), 225-226.
- McGuire, B. E. (2006). Response mode and performance on the Rey 15-item test: A preliminary study. *Brain injury, 20*(6), 647-651.
- McGuire, C., Crawford, S., & Evans, J. J. (2018). Effort testing in dementia assessment: A systematic review. *Archives of Clinical Neuropsychology, 34*(1), 114-131.
- Merten, T., Dandachi-FitzGerald, B., Hall, V., Schmandd, B. A., Santamaría, P., & González-Ordi, H. (2013). Symptom validity assessment in European countries: Development and state of the art. *Clínica y Salud, 24*(3), 129-138.
- Meyers, J. E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology, 18*(3), 261-276.
- Meyers, J. E., Volbrecht, M., Axelrod, B. N., & Reinsch-Boothby, L. (2011). Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology, 26*(1), 8-15.
- Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2011). Are embedded validity indices equivalent to free-standing symptom validity tests?. *Archives of Clinical Neuropsychology, 27*(1), 10-22.

- Millis, S. R., Putnam, S. H., Adams, K. M., & Ricker, J. H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment, 7*(4), 463-471.
- Mittenberg, W., Aguila-Puentes, G., Patton, C., Canyock, E. M., & Heilbronner, R. L. (2002). Neuropsychological profiling of symptom exaggeration and malingering. *Journal of Forensic Neuropsychology, 3*, 227-240.
- Mittenberg, W., Azrin, R., Millsaps, C., & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale—Revised. *Psychological Assessment, 5*(1), 34-40.
- Morgan, J. E., & Sweet, J. J. (Eds.). (2009). *Neuropsychology of malingering casebook*. New York: American Academy of Clinical Neuropsychology.
- Morse, C. L., Douglas-Newman, K., Mandel, S., & Swirsky-Sacchetti, T. (2013). Utility of the Rey-15 recognition trial to detect invalid performance in a forensic neuropsychological sample. *The Clinical Neuropsychologist, 27*(8), 1395-1407.
- Mota, M. (2008). Test of Memory Malingering (TOMM): Estudos de validação numa amostra de reclusos. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699.

- Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *The Clinical Neuropsychologist*, *17*(2), 263-272.
- O'Bryant, S. E., Engel, L. R., Kleiner, J. S., Vasterling, J. J., & Black, F. W. (2007). Test of memory malingering (TOMM) trial 1 as a screening measure for insufficient effort. *The Clinical Neuropsychologist*, *21*(3), 511-521.
- O'Bryant, S. E., Gavett, B. E., McCaffrey, R. J., O'Jile, J. R., Huerkamp, J. K., Smitherman, T. A., & Humphreys, J. D. (2008). Clinical utility of Trial 1 of the Test of Memory Malingering (TOMM). *Applied Neuropsychology*, *15*(2), 113-116.
- Organisation for Economic Co-operation and Development. (2018). *Care needed: Improving the lives of people with dementia*. OECD Health Policy Studies. OECD Publishing, Paris.
- Oliveira, C. S. S. (2008). Estudos de validação com o Test of Memory Malingering (TOMM), o Rey 15-Item Visual Memory Teste (Rey-15IMT) e o Structured Inventory of Malingered Symptomatology (SIMS) numa amostra de reclusos em “cadeias especiais”. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Pankratz, L. (1988). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 169-194). New York: The Guilford Press.
- Pankratz, L., & Binder, L. M. (1997). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 223-236). New York: The Guilford Press.

- Pankratz, L., & Erickson, R.C. (1990). Two views of malingering. *The Clinical Neuropsychologist*, 4, 379-389.
- Paul, D. S., Franzen, M. D., Cohen, S. H., & Fremouw, W. (1992). An investigation into the reliability and validity of two tests used in the detection of dissimulation. *International Journal of Clinical Neuropsychology*, 14, 1-9.
- Pinho, J. I. S. S. (2012). Testes de validade de sintomas: Validação de um protocolo em amostras de adultos idosos. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicogerontologia Clínica. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, 20(1), 33-65.
- Rees, L. M., Tombaugh, T. N., & Boulay, L. (2001). Depression and the test of memory malingering. *Archives of Clinical Neuropsychology*, 16(5), 501-506.
- Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Universitaires de France.
- Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury*, 19(7), 539-543.
- Rickards, T. A., Cranston, C. C., Touradji, P., & Bechtold, K. T. (2017). Embedded performance validity testing in neuropsychological assessment: Potential clinical tools. *Applied Neuropsychology: Adult*, 25(3), 219-230.
- Robles, L., López, E., Salazar, X., Boone, K. B., & Glaser, D. F. (2015). Specificity data for the b Test, Dot Counting Test, Rey-15 Item Plus Recognition, and Rey Word

- Recognition Test in monolingual Spanish-speakers. *Journal of Clinical and Experimental Neuropsychology*, 37(6), 614-621.
- Rogers, R. (1990a). Development of a new classificatory model of malingering. *Bulletin of the American Academy of Psychiatry & the Law*, 18(3), 323-333.
- Rogers, R. (1990b). Models of feigned mental illness. *Professional Psychology: Research and Practice*, 21(3), 182-188.
- Rogers, R. (2008). Detection strategies for malingering and defensiveness. *Clinical assessment of malingering and deception*, 3, 14-35.
- Rogers, R., Harrell, E. H., & Liff, C.D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review*, 13, 255-274.
- Rudman, N., Oyebode, J. R., Jones, C. A., & Bentham, P. (2011). An investigation into the validity of effort tests in a working age dementia population. *Aging & Mental Health*, 15(1), 47-57.
- Schretlen, D., Brandt, J., Krafft, L., & Van Gorp, W. (1991). Some caveats in using the Rey 15-Item Memory Test to detect malingered amnesia. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(4), 667-672.
- Schroeder, R. W., Peck, C. P., Buddin Jr, W. H., Heinrichs, R. J., & Baade, L. E. (2012). The Coin-in-the-Hand Test and dementia: More evidence for a screening test for neurocognitive symptom exaggeration. *Cognitive and Behavioral Neurology*, 25(3), 139-143.

- Sherman, D. S., Boone, K. B., Lu, P., & Razani, J. (2002). Re-examination of a Rey Auditory Verbal Learning Test/Rey Complex Figure discriminant function to detect suspect effort. *The Clinical Neuropsychologist*, *16*(3), 242-250.
- Silva, D. C. (2011). Test of Memory Malinger (TOMM): Estudo de validação numa amostra de crianças e adolescentes com diagnóstico de deficiência mental. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Simões, M. R., Almeida, L. S., & Gonçalves, M. M. (2017). *Psicologia forense: Instrumentos de avaliação*. Lisboa: Pactor.
- Simões, M. R., Nunes, A. P., & Cunha, C. (2010). *Escala de Validade de Sintomas (EVS-Versão Experimental)*. Coimbra: Serviço de Avaliação Psicológica. Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra.
- Simões, M. R., Sousa, L., Duarte, P., Firmino, H., Pinho, M. S., Gaspar, N., ... & Ferreira, A. R. (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*, *28*(1), 209-226.
- Simões, M. R., Sousa, L. B., Marques-Costa, C., & Almiro, P.A. (2017). Avaliação neuropsicológica em contextos forenses: fundamentos, protocolo e instrumentos de avaliação. In Fernando Vieira, Ana S. Cabral & Carlos B. Saraiva (Eds.), *Manual de Psiquiatria Forense* (pp. 373-410). Lisboa: Pactor/Lidel.
- Simon, M. J. (1994). The use of the Rey Memory Test to assess malingering in criminal defendants. *Journal of Clinical Psychology*, *50*(6), 913-917.

- Simon, M. J. (2007). Performance of mentally retarded forensic patients on the Test of Memory Malingering. *Journal of Clinical Psychology, 63*(4), 339-344.
- Slick, D., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's forced-choice memory test. *Journal of Clinical and Experimental Neuropsychology, 16*(3), 472-481.
- Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*(4), 545-561.
- Smith, G. P., & Burger, G. K. (1997). Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law Online, 25*(2), 183-189.
- Soares, D. M. (2013). Detecção de comportamentos de esforço reduzido e exagero de sintomas numa amostra de reclusos do Estabelecimento Prisional de Coimbra. Dissertação de Mestrado em Psicologia, especialidade de Psicologia Clínica e da Saúde, subespecialidade de Psicologia Forense. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Strauss, S. E., Richardson, W.S., Glasziou, P., Haynes, R. B. (2005). *Evidence-based medicine: How to practice and teach EBM* (3rd ed.). New York: Elsevier Churchill Livingstone.
- Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.
- Strutt, A. M., Scott, B. M., Shrestha, S., & York, M. K. (2011). The Rey 15-item memory test and Spanish-speaking older adults. *The Clinical Neuropsychologist, 25*(7), 1253-1265.

- Sugarman, M. A., & Axelrod, B. N. (2015). Embedded measures of performance validity using verbal fluency tests in a clinical sample. *Applied Neuropsychology: Adult*, 22(2), 141-146.
- Suhr, J. A., & Boyer, D. (1999). Use of the Wisconsin Card Sorting Test in the detection of malingering in student simulator and patient samples. *Journal of Clinical and Experimental Neuropsychology*, 21(5), 701-708.
- Suhr, J. A., & Gunstad, J. (2007). Coaching and malingering: A review. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 287-311). New York: Oxford University Press.
- Suhr, J., Tranel, D., Wefel, J., & Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology*, 19(4), 500-514.
- Sweet, J. J. (1999). Malingering: Differential diagnosis. In J. J. Sweet (Ed.), *Studies on neuropsychology, development, and cognition. Forensic neuropsychology: Fundamentals and practice* (pp. 255-285). Lisse, Netherlands: Swets & Zeitlinger Publishers.
- Taylor, L. A., Kreutzer, J. S., & West, D. D. (2003). Evaluation of malingering cut-off scores for the Rey 15-Item Test: a brain injury case study series. *Brain Injury*, 17(4), 295-308.
- Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, 19(3), 455-464.

- Tenhula, W. N., & Sweet, J. J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. *The Clinical Neuropsychologist*, *10*(1), 104-116.
- Tombaugh, T. N. (1996). *Test of memory malingering: TOMM*. New York: Multi Health Systems.
- Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, *9*(3), 260-268.
- Tracy, D. K. (2014). Evaluating malingering in cognitive and memory examinations: a guide for clinicians. *Advances in psychiatric treatment*, *20*(6), 405-412.
- van der Flier, W. M., & Scheltens, P. (2005). Epidemiology and risk factors of dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, *76*.
- van Gorp, W., & Hassenstab, J. (2009). Why questions regarding effort and malingering are always raised in forensic neuropsychological evaluations. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of Malingering Casebook* (pp. 9-20). Psychology Press.
- Vickery, C. D., Berry, D. T., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*(1), 45-73.
- Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, *23*(2), 297-313.
- Walter, J., Morris, J., Swier-Vosnos, A., & Pliskin, N. (2014). Effects of severity of dementia on a symptom validity measure. *The Clinical Neuropsychologist*, *28*(7), 1197-1208.

- Walters, G. D., Rogers, R., Berry, D. T. R., Miller, H. A., Duncan, S. A., McCusker, P. J., ... & Granacher Jr., R. P. (2008). Malingering as a categorical or dimensional construct: The latent structure of feigned psychopathology as measured by the SIRS and MMPI-2. *Psychological Assessment, 20*(3), 238-247.
- Webber, T. A., Bailey, K. C., Alverson, W. A., Critchfield, E. A., Bain, K. M., Messerly, J. M., ... & Soble, J. R. (2019). Correction to: Further validation of the Test of Memory Malingering (TOMM) Trial 1 Performance Validity Index: Examination of false positives and convergent validity. *Psychological Injury and Law, 12*(1), 88-89.
- Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19*, 87-95.
- Wechsler, D. (1997). *WAIS-3: Wechsler Adult Intelligence Scale: Administration and Scoring Manual*. Psychological Corporation.
- Wechsler, D. (2008). *Escala de Inteligência de Wechsler para Adultos – 3ª Edição (WAIS-III)*. Lisboa: CEGOC.
- Wechsler, D. (2009). *WMS-IV: Wechsler Memory Scale*. San Antonio, TX: Pearson.
- Weinborn, M., Orr, T., Woods, S. P., Conover, E., & Feix, J. (2003). A validation of the Test of Memory Malingering in a forensic psychiatric setting. *Journal of Clinical and Experimental Neuropsychology, 25*(7), 979-990.
- Wessely, S. (2003). Malingering: Historical perspectives. In P. W. Halligan, C. Bass & D. A. Oakley (Eds.), *Malingering and illness deception* (pp. 31-41). New York: Oxford University Press.

- Whitney, K. A., Hook, J. N., Steiner, A. R., Shepard, P. H., & Callaway, S. (2008). Is the Rey 15-Item Memory Test II (Rey II) a valid symptom validity test?: Comparison with the TOMM. *Applied neuropsychology, 15*(4), 287-292.
- Widows, M. R., & Smith, G. P. (2005). *SIMS: Structured Inventory of Malingered Symptomatology*. Odessa: Psychological Assessment Resources, Inc.
- Yeh, T. C., Chou, Y. C., Weng, J. P., Yeh, H. W., Kao, Y. C., Chiang, W. S., ... & Tzeng, N. S. (2019). Detection of malingering in the memory of patients with dementia: A pilot study on coin-in-the-hand test in a Northern Taiwan Memory Clinic. *Journal of Medical Sciences, 39*(2), 81-89.
- Ylioja, S. G., Baird, A. D., & Podell, K. (2009). Developing a spatial analog of the Reliable Digit Span. *Archives of Clinical Neuropsychology, 24*, 729-739.
- Young, J. C., Sawyer, R. J., Roper, B. L., & Baughman, B. C. (2012). Expansion and re-examination of Digit Span effort indices on the WAIS-IV. *The Clinical Neuropsychologist, 26*(1), 147-159.
- Zenisek, R., Millis, S. R., Banks, S. J., & Miller, J. B. (2016). Prevalence of below-criterion Reliable Digit Span scores in a clinical sample of older adults. *Archives of Clinical Neuropsychology, 31*(5), 426-433.