



UNIVERSIDADE D
COIMBRA

Carolina Duarte de Almeida

ESTUDOS COM *DESIGN* DE SIMULAÇÃO ANÁLOGO EM
TESTES DE VALIDADE DE SINTOMAS:
UMA REVISÃO DA LITERATURA

Dissertação no âmbito do Mestrado em Psicologia, área de especialização em Psicologia Clínica e da Saúde, área de subespecialização em Psicologia Forense, orientada pelo/a Professor Doutor Mário R. Simões e apresentada à Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra.

Outubro de 2020

Estudos com *design* de simulação análogo em Testes de Validade de Sintomas (TVS): Uma revisão da literatura

Resumo

A área da avaliação da validade de sintomas tem assumido uma importância acrescida no domínio da avaliação psicológica. Em contextos forenses, a evidência dos ganhos secundários alcançáveis com a simulação ou exagero de sintomas psicopatológicos e a prevalência deste tipo de comportamentos por parte dos avaliados, constitui um problema relevante. Os Testes de Validade de Sintomas (TVS) têm como objetivo detetar a presença de comportamentos de simulação ou exagero de sintomas permitindo, aos profissionais de saúde mental, realizar interpretações mais fidedignas dos resultados obtidos através de uma avaliação psicológica, de forma a refletir o funcionamento psicológico real do sujeito.

Neste âmbito, releva-se a importância das investigações na área, para que os clínicos consigam recorrer à utilização dos TVS com confiança nos seus resultados. Uma das mais comuns metodologias de investigação na simulação de sintomas é o *design* de simulação análogo, no qual são formados grupos de controlo e de simulação instruída, para estudar a identificação e controlo dos instrumentos ao comportamento referido. No entanto, esta metodologia encontra-se muito associada a críticas relativas à sua fraca validade externa, devido às diferenças existentes entre as condições experimentais e as condições da vida real, nas quais esses comportamentos podem estar presentes. Neste sentido, acresce a importância de serem respeitados certos padrões de construção metodológica nas investigações realizadas, de forma a que as conclusões retiradas sejam generalizadas com confiança.

Assim, na presente dissertação, é realizada uma revisão dos estudos existentes com recurso ao *design* de simulação análogo, com destaque para os contextos forenses (criminal e médico-legal). Considerando o número reduzido de investigações encontradas nestes contextos, são, adicionalmente, apresentados estudos em contextos comunitários. As conclusões retiradas indicam que são, ainda, cometidos diversos erros associados ao desenvolvimento das instruções de simulação, o que coloca em risco a fiabilidade com que são interpretados os comportamentos de simulação reproduzidos pelos “simuladores experimentais”. São feitas recomendações de natureza metodológica orientadas para assegurar um maior grau de representatividade e generalização dos resultados obtidos neste *design* de simulação análogo para os contextos reais.

Palavras chave: Simulação, exagero de sintomas, validade de sintomas, testes de validade de sintomas; *design* de simulação análogo.

Studies with analogue simulation design in Symptom Validity Tests (SVT): A literature review

Abstract

Symptom validity assessment is an area becoming increasingly important in the field of psychological assessment. In forensic contexts, the presence of evident secondary gains achievable with malingering or symptom overreporting on the part of the evaluated, and the prevalence of this kind of performance, constitutes a relevant problem. Symptom Validity Tests (SVT) aim to detect the presence of malingered symptomatology, allowing mental health professionals to produce more reliable interpretations of the results obtained through a psychological assessment, in order to reflect the subject's real psychological functioning.

Therefore, the attainment of investigations in the area is highlighted, to help clinicians to resort to the use SVT with confidence in their results. One of the most common investigation methodologies to evaluate the identification and control of malingering by the SVT, is the analogue simulation *design*, according to which participants are allocated to control or simulation conditions. Nevertheless, this methodology is closely associated with criticisms regarding its weak external validity, due to the differences between the experimental and real-life conditions. Thus, investigators should respect certain methodological standards, so that the conclusions drawn from the investigations are generalized with confidence.

In the present dissertation, a review of the existing studies using analogue simulation design is carried out. Emphasis is placed on forensic (criminal and medico-legal) contexts, with addition of community contexts, due to the reduced number of investigations in the first ones. The conclusions drawn indicate several errors associated with the development of simulation instructions, which lays risk in the reliability with which the simulation behaviours reproduced by experimental simulators are interpreted. Methodological recommendations oriented to ensure a higher degree of representativeness and generalization of the results obtained in this analog simulation design are made.

Key Words: Malingering, symptom overreporting, symptom validity, symptom validity testing; analogue simulation design.

Agradecimentos

No final desta tão importante etapa, não posso deixar de agradecer a todos os que contribuíram para me ajudar a chegar onde estou.

À minha família, pelo apoio e confiança que sempre depositaram em mim, e por caminharem sempre ao meu lado. Em especial: aos meus pais, os maiores pilares da minha vida, pela oportunidade que me deram, e pelo incentivo para chegar sempre mais longe; e à minha irmã, o maior amor da minha vida, por me inspirar todos os dias a ser melhor.

Ao Professor Doutor Mário Simões, por toda a orientação, e por todos os conhecimentos transmitidos.

À Ana e à Bárbara por terem tornado Coimbra na melhor casa destes últimos cinco anos, e à Mariana, por se ter juntado a nós. Por esta amizade que tornou o caminho mais fácil de se fazer.

À Beatriz, que me acolheu naquela que já era a sua casa, por todo o carinho e apoio.

Ao Diogo, pela paciência, apoio e companheirismo que sempre me deram força para continuar.

À Rita, colega de estágio que se tornou amiga, pela companhia das horas de trabalho.

À Filipa, uma mentora e amiga, por toda a ajuda indispensável.

Ao Estabelecimento Prisional de Leiria (Jovens) e à sua equipa, por me terem aberto as portas neste último ano, e terem permitido a realização daquele que foi o meu projeto inicial, embora inacabado. Em especial à Dra. Célia Batista, pelo apoio e pelos conhecimentos transmitidos ao longo do estágio.

Índice

Introdução.....	1
1 – Estilos de Resposta.....	2
2 - Simulação e Exagero de Sintomas.....	4
3 - Testes de Validade de Desempenho (TVD) e Testes de Validade de Sintomas (TVS).....	10
4 - Testes de Validade de Sintomas (TVS).	15
5 – Questões éticas associadas ao diagnóstico de simulação.	23
6 - O Coaching – Simuladores instruídos e não instruídos.....	25
7 - O design de Simulação Análogo.....	28
8 - Estudos com design de simulação análogo em contextos forenses.....	32
8.1 - Estudos em contexto criminal.	32
8.2 - Estudos em contexto médico e médico-legal.	34
9 - Estudos com design de simulação análogo em contexto comunitário.	36
Conclusões	39
Bibliografia.....	41
Anexos.....	52

Lista de Anexos

Anexo A – Tabela 2.1 e Tabela 2.2.....	52
Anexo B – Tabela 3.1 e Tabela 3.2.....	54
Anexo C – Tabela 4.1 e Tabela 4.2.....	56

Lista de Tabelas

Tabela 1 - Análise dos estudos de avaliação do impacto do coaching na detecção de simulação.	27
Tabela 2.1 - Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto criminal.....	52
Tabela 2.2 - Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto criminal (continuação).....	53
Tabela 3.1 - Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto médico e médico-legal.....	54
Tabela 3.1 - Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto médico e médico-legal (continuação)....	55
Tabela 4.1 - Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto comunitário.....	56
Tabela 4.2 - Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto comunitário (continuação).....	58

Introdução

A validade dos sintomas reportados durante um processo de avaliação psicológica encontra-se dependente da honestidade e do esforço/motivação do sujeito para fornecer dados o mais realistas possível. No entanto, a existência de influências com ação sob essa genuinidade que geram, conseqüentemente, estilos de resposta enviesados, pode levar a um comprometimento dos resultados obtidos. Existe uma grande variedade de estilos de resposta que podem ser adotados pelos sujeitos, sendo que a sua utilização depende sempre dos objetivos pessoais específicos destes, em cada contexto (Merten et al., 2013; Merten & Merckelback, 2013b; Rogers, 2018a; Rogers & Bender, 2012).

Um dos estilos de resposta mais usualmente adotados em contextos forenses é a simulação ou o exagero de sintomas. De acordo com o DSM-5, este é um fenômeno que não é abrangido na categoria de doença mental, pertencendo a um grupo de condições que deve ser alvo de atenção clínica. É definido como uma produção de sintomas, de forma intencional e externamente motivada, podendo esses sintomas ser falsos ou exagerados (APA, 2014).

A prevalência dos comportamentos simulação em contextos forenses, incluindo meios prisionais, tem sido alvo de tentativas estimação em diversos estudos (Bass & Halligan, 2014). Embora não se tenha conseguido obter resultados consensuais, as taxas estimadas são consideráveis (e.g. McDermott et al., 2013; McDermott & Sokolov, 2009; Norris & May, 1998).

Os Testes de Validade de Sintomas (TVS) surgiram no sentido de ajudar na detecção da utilização dos comportamentos de simulação e exagero de sintomas pelos sujeitos durante os procedimentos de avaliação psicológica, e têm como base de construção o conhecimento limitado dos simuladores relativamente aos sintomas psicológicos genuínos. Estes podem ser divididos em duas categorias: os Testes de Validade de Desempenhos (TVD; *Performance Validity Tests*), que detetam o esforço insuficiente ou o baixo desempenho intencionais; e os Testes de Validade de Sintomas (TVS; *Symptom Validity Tests*), medidas de autorrelato que avaliam o exagero e a simulação de sintomas (Jelicic et al., 2007; Merten & Merckelback, 2013a; Merten et al., 2013; Merten et al., 2016)

Os TVS mais utilizados e mencionados pela literatura são o *Structured Inventory of Malingered Symptomatology* (SIMS; Widows, & Smith, 2005), o *Structured Interview of Reported Symptoms* (SIRS; Rogers et al., 1992; SIRS-2; Rogers, et al., 2010), o *Self-Report Symptom Inventory* (SRSI; Merten et al., 2016), e no contexto português, também a Escala de Validade de Sintomas (EVS-2, Cunha, 2011; Simões et al., 2015; como citados em Simões, Almiro, et al., 2017). Importa, ainda, referir a existência de escalas integradas em instrumentos de avaliação mais abrangentes, e que funcionam, também, como TVS. São as mais relevantes as escalas integradas no *Minnesota Multiphasic Personality Inventory – 2* (MMPI-2; Butcher et al., 2001, como citado em Wall et al., 2015) e no *Personality Assessment Inventory* (PAI; Morey, 2007, como citado em Merten et al., 2016).

Estudos com *design* de simulação análogo em Testes de Validade de Sintomas (TVS):

Uma revisão da literatura

Carolina Duarte de Almeida (carolinad Almeida2@gmail.com) 2020

A presente dissertação tem como objetivo fazer uma revisão dos estudos existentes, nos quais tenham sido utilizados TVS, e que tenham recorrido ao *design* de simulação análogo como estratégia de investigação/validação. Este tipo de *design* implica a distribuição aleatória de participantes por condições distintas sendo, habitualmente, formados dois grupos, um instruído a responder de forma honesta, e outro instruído a simular determinados sintomas (Merckelbach, et al., 2009; Rogers, 2018a).

1 – Estilos de Resposta.

Os resultados e as conclusões formuladas através das avaliações psicológicas, tanto em contextos clínicos como forenses, encontram-se fortemente dependentes da disponibilidade e vontade do examinado para responder de forma exata e aberta durante o processo de avaliação. Pode, então, dizer-se, que estas conclusões são altamente vulneráveis ao estilo de resposta adotado pelo mesmo. Sendo este estilo enviesado, leva a que o perfil de funcionamento obtido na avaliação psicológica, não corresponda ao seu nível de funcionamento psicológico real (Merten et al., 2013; Merten & Merckelbach, 2013b; Rogers & Bender, 2012).

Na verdade, as influências externas ou internas presentes, muitas vezes, nestes contextos podem comprometer a exatidão das respostas fornecidas pelos sujeitos. Embora os profissionais de saúde mental concentrem a sua atenção na obtenção de benefícios como motivação para a aplicação de estilos de resposta distorcidos nos autorrelatos, estas podem ser variadas. As influências externas podem ir desde as possíveis consequências adversas dos litígios, à pressão exercida pelos outros, como familiares ou advogados. Por outro lado, a existência de uma reação à credibilidade questionada, a existência de estigmas associados à doença mental e os efeitos de uma perturbação genuína, são algumas das possíveis influências internas à utilização de estilos de resposta nas avaliações psicológicas (Rogers & Bender, 2003, 2012).

Rogers (2018a) postula a existência de diversos tipos de estilos de resposta que podem surgir nos contextos psicológicos, e que são divididos em quatro categorias. A primeira diz respeito aos estilos *Não específicos* (“*Nonspecific Terms*”), que englobam comportamentos como a apresentação não fidedigna de informação (“*Unreliability*”); a omissão ou não divulgação de certas informações (“*Nondisclosure*”); o grau de abertura dos indivíduos na revelação de informação sobre si mesmos (“*Self-disclosure*”); a distorção ou apresentação deturpada no autorrelato dos indivíduos (“*Deception*”); e a dissimulação ou a apresentação de sintomas psicopatológicos distorcida (“*Dissimulation*”). A segunda categoria corresponde ao *Exagero de sintomas psicopatológicos* (“*Overstated Pathology*”) e engloba a simulação e falsificação de psicopatologia e sintomas, tendo subjacente uma motivação externa (“*Malingering*”) ou sem motivação especificada (“*Feigning*”), e a Perturbação Factícia, que consiste na produção de sintomas, com o objetivo de aparentar doença. No terceiro grupo denominado de *Ajustamento simulado* (“*Simulated Adjustment*”), são inseridos os comportamentos de defensividade (“*Defensiveness*”) que dizem respeito à minimização do número ou gravidade

Estudos com *design* de simulação análogo em Testes de Validade de Sintomas (TVS):

Uma revisão da literatura

Carolina Duarte de Almeida (carolinad Almeida2@gmail.com) 2020

dos sintomas apresentados; de desejabilidade social (“*Social Desirability*”) que corresponde a uma apresentação dos indivíduos socialmente mais favorável, com minimização das características negativas e enaltecimento das positivas; e de gestão de impressões (“*Impression Management*”) que consiste nos esforços para controlar a percepção dos outros sobre si, acontecendo de forma mais situacionalmente restrita do que o anteriormente referido. Por último, são agrupados numa categoria os estilos de resposta que, contrariamente aos anteriores, ainda não se encontram suficientemente compreendidos e categorizados. Entre eles encontram-se o estilo de resposta irrelevante (“*Irrelevant Responding*”), que surge como consequência da falta de envolvimento do sujeito no processo de avaliação; o estilo aleatório (“*Random Responding*”), caracterizado pelas respostas ao acaso; o estilo aquiescente (“*Acquiescent Responding*”), que corresponde à escolha maioritária das respostas positivas (i.e. “Sim”) e o estilo oposto a este, ou seja, respostas negativas indiscriminadas (“*Disacquiescent Responding*”). Ainda nesta última categoria, são incluídas a assunção de papéis (“*Role Assumption*”), no qual os indivíduos se comportam de acordo com uma personagem que assumem, e o estilo Híbrido (“*Hybrid Responding*”), em que são adotados estilos de resposta diferentes nas variadas situações.

Em conformidade com os resultados de um estudo de Ellison et al. (2013; como citados em Rogers, 2018a), em que foram avaliadas as escolhas de divulgação dos problemas de saúde mental de funcionários aos seus empregadores, Rogers (2018a) concluiu que as decisões respeitantes aos estilos de resposta utilizados pelos sujeitos são, muitas vezes, racionais e multideterminadas, e que correspondem a respostas individualizadas a variáveis interpessoais ou às exigências das situações. Ou seja, estes estilos são flexíveis e não são, ao contrário de algumas crenças, correspondentes a traços ou características pessoais. São, por isso, adaptáveis aos objetivos pessoais dos sujeitos, específicos de cada contexto (Rogers, 2018a).

Neste sentido, podendo as consequências desta influência ser substanciais, é importante a realização de uma avaliação cuidadosa da presença dos estilos de resposta nos autorrelatos do indivíduo, ou seja, uma avaliação da validade dos sintomas fornecidos. A validade de sintomas pode ser, então, definida como a precisão ou a exatidão da apresentação dos sintomas, comportamentos ou do desempenho de um sujeito nos testes psicológicos (Bass & Halligan, 2014). As investigações nesta área, iniciadas nos anos 90 (Merten et al., 2013), têm evidenciado a importância da avaliação da validade dos sintomas em contextos de avaliação psicológica e, em especial, nos contextos forenses nos quais os sujeitos têm, numa parte dos casos, uma motivação externa para recorrer à distorção das informações fornecidas. Atualmente, esta avaliação é já muito prevalente, sendo mesmo uma prática padrão em grande parte dos contextos forenses (Green & Merten, 2013, como citado em Martin et al., 2015; Sweet & Breting, 2013).

Rogers e Bender (2003) indicam a existência de três perspetivas principais no que respeita à avaliação da validade dos sintomas: a *Intuitiva* (“*Intuition*”), que se baseia na assunção de que a presença de um estilo de resposta distorcido é facilmente identificável, sem a necessidade de recorrer a

estratégias empíricas; a *Padronizada* (“*Standard*”), de acordo com a qual os estilos de resposta são avaliados através de testes e instrumentos de medida tradicionais; e a *Especializada* (“*Specialized*”), que apoia o reforço das medidas tradicionais com instrumentos específicos para a avaliação de validade de sintomas. À última perspectiva, embora com acréscimo no tempo despendido, corresponde um poder de precisão de classificação superior (Rogers & Bender, 2003).

2 - Simulação e Exagero de Sintomas.

O termo “simulação de doença/sintomas” foi incluído no DSM (*The Diagnostic and Statistical Manual of Mental Disorders/Manual de Diagnóstico e Estatístico das Perturbações Mentais*) pela primeira vez na sua terceira edição, tendo, desde essa altura, existido grandes discussões e controvérsias no que diz respeito ao termo utilizado e à sua definição (Ardic et al., 2019; Merten & Merckelback, 2013a). A simulação é, atualmente, referida como “a produção intencional de sintomas físicos ou psicológicos, falsos ou exagerados, motivados por incentivos externos” (American Psychological Association [APA], p.870, 2014). Na 5ª. edição do DSM, a simulação não é considerada uma perturbação mental, e encontra-se na categoria de “Outras condições que podem ser um foco de atenção clínica”. Ainda de acordo com o que é apresentado pelo DSM, a presença de simulação deve ser objeto de suspeita pelos clínicos quando existe uma combinação de qualquer das seguintes condições: (1) presença em contexto médico-legal; (2) discrepância entre os sintomas relatados e os sintomas observados; (3) falta de cooperação durante a avaliação ou falta de adesão ao tratamento; (4) diagnóstico de Perturbação da Personalidade Antissocial (APA, 2014; Ardic, et al., 2019; Impelen et al., 2017).

Embora apresentem uma definição semelhante, importa referir a existência de simulação ou exagero de sintomas de duas naturezas distintas: de perturbação mental propriamente dita, e de défices cognitivos. Estas duas dimensões do comportamento simulador implicam o recurso a diferentes estratégias e métodos de deteção (Bender & Frederick, 2018; Rogers & Bender, 2003).

Além disso, a simulação pode manifestar-se de três formas distintas: a *simulação pura*, que diz respeito à falsificação total ou à comunicação de sintomas inexistentes; a *simulação parcial* pode-se manifestar através do exagero de sintomas existentes ou da manutenção ou agravamento de sintomas que, embora existentes anteriormente, se atenuaram; e por último, a *falsa imputação de sintomas* a uma causa não etiologicamente relacionada (Simões, Almiro, Nunes, et al., 2017).

Este termo é, muitas vezes, comparado à “perturbação factícia”, pois ambos se caracterizam pela produção consciente de sintomatologia psicopatológica. No entanto, no caso de uma pessoa que sofre desta última, a produção ocorre devido a motivações internas de aparentar doença e se sentir tratada de modo especial por estar doente, enquanto na simulação a motivação é externa (APA, 2014; Conroy & Kwartner, 2006; Impelen et al., 2017; Merten & Merckelback, 2013a; Huhnt, 2013).

No entanto, diversas preocupações têm sido levantadas relativamente aos critérios da simulação presentes no DSM (Berry & Nelson, 2010; Huhnt, 2013). Em primeiro lugar, a distinção entre motivações internas e externas utilizadas para separar a simulação da perturbação factícia tem sido alvo de várias críticas, havendo autores (e.g. Bass & Halligan, 2007, como citados em Huhnt, 2013) que consideram mesmo que a exclusividade mútua destas duas classes não deveria existir, pois a diferenciação entre elas é pouco nítida (Huhnt, 2013). Berry e Nelson (2010), afirmam ainda que, embora seja possível detetar o exagero de sintomas de alguém, o mesmo não se pode afirmar para as motivações que o subjazem e para a sua natureza externa ou interna. Por esta mesma razão, com a 5ª Edição do DSM, dos critérios de diagnóstico da Perturbação Factícia deixou de fazer parte a existência necessária de incentivos internos para a apresentação do estilo de resposta distorcido, não se verificando, no entanto, o mesmo com o critério de existência de motivações externas para a categorização de simulação (Impelen et al., 2017). Adicionalmente, de acordo com vários autores (como citados em Berry & Nelson, 2010), deve existir um alerta para o critério obrigatório de intencionalidade para a categorização do comportamento de simulação, que torna este procedimento subjetivo e especulativo, na ausência de uma confissão do simulador.

Também a sugestão de suspeição da presença de simulação quando existe um diagnóstico de Perturbação da Personalidade Antissocial (PPAS), pode ser problemática. Kucharski et al. (2006, como citados em Huhnt, 2013), alertam para a possibilidade de surgirem erros de enviesamento confirmatório decorrentes deste critério, por poder levar os clínicos a considerar a simulação uma consequência inevitável da PPAS (Huhnt, 2013). Rogers e Cruise (2000, como citados em Huhnt (2013), afirmaram que não deve ser confundida a intenção de alguém de simular, com o sucesso com o que o faz, e que pacientes com Personalidade Antissocial não são, no geral, mais eficazes que outros na simulação de sintomas.

Uma outra preocupação de diversos investigadores reside na apresentação da simulação como sendo uma condição categórica, e não dimensional e contínua, como foi percebida num estudo de Walters et al. (2009, como citado em Berry & Nelson, 2010). Por último, Berry e Nelson (2010) apontam alguma apreensão causada pelo pressuposto do DSM de que os simuladores o serão em todos, ou quase todos, os procedimentos, assentes em sintomas de diversas naturezas (psiquiátricos, físicos/somáticos e cognitivos/neuropsicológicos). Isto já foi infirmado pelos resultados de alguns estudos (e.g. Alwes et al., 2007, Nelson et al., 2007; como citados em Berry & Nelson, 2010), comprovando assim alguma independência nos sintomas simulados.

Na verdade, a falta de atualização dos critérios presentes no DSM desde a sua 3ª edição, mesmo com o avanço das investigações na área, levam autores, como Berry e Nelson (2010) a considerar estes obsoletos. Ainda neste âmbito, de acordo com LoPiccolo et al. (1999, como citados em Rogers & Bender, 2003), estes critérios não são úteis, mesmo em situações de rastreio. Rogers (1990, como citado em Rogers & Bender, 2003) afirma, ainda, que a

taxa de precisão de categorização seria de apenas 20%, quando os índices do DSM são adotados em contextos criminais.

De acordo com Rogers (1990, citado por Pollock et al., 1997), podemos considerar a existência de três modelos explicativos da existência de simulação de sintomas psicopatológicos. O **Modelo Patogénico** afirma que a simulação é o resultado de uma tentativa mal conseguida de lidar com uma doença mental genuína, isto é, de forma a poderem ter alguma perceção de controlo na sua psicopatologia, os doentes passam a fabricar sintomas. No entanto, à medida que a perturbação mental agrava, o controlo dos doentes sobre a psicopatologia simulada diminui e torna-se involuntária. Deste modo, a motivação inerente à simulação é a doença mental. (Rogers et al., 1998; Rogers et al., 1994, como citados em Rogers, 2018a).

O **Modelo Criminológico**, na base dos índices de classificação do DSM, assenta na assunção de que uma pessoa com Perturbação da Personalidade Antissocial (PPAS) que demonstre uma atitude de fraca cooperação com a avaliação ou com o tratamento e apresente discrepâncias entre os relatos e o comportamento observável, tem elevada probabilidade de simular sintomas perante um contexto de avaliação psicológica forense. Portanto, a motivação para apresentar este comportamento de simulação seria criminológico/antissocial (Rogers, 2018a; Rogers et al., 1998; Rogers & Bender, 2003). Este é um modelo que não reúne evidência empírica e que não apresenta uma base teórica para a sua formulação (Pollock et al., 1997). De acordo com Rogers (2018a), o facto de apresentar o diagnóstico de PPAS e a presença em contextos de avaliação psicológica forense como critérios, é o principal problema deste modelo, por estas serem características comuns de grande parte da população criminal, não fazendo uma distinção entre os simuladores e os pacientes honestos com sintomas genuínos.

Por último, o **Modelo Adaptativo** é o mais aceite, e considera a simulação como um estilo de resposta que surge perante condições adversas, como é o caso médico-legal ou forense. Nestes contextos o simulador faz uma análise de custo-benefício das suas opções, influenciada pela estimativa que faz das suas capacidades para ter sucesso, e decide que a simulação é a opção mais viável para poder evitar consequências negativas para si. Deste modo, a simulação não é vista como uma dimensão estável ou uma característica de indivíduos que utilizam esta defesa, mas como uma resposta situacional baseada na avaliação das alternativas (McDermott et al., 2013; Rogers, 2018a; Rogers & Bender, 2003; Rogers et al., 2010).

No caso específico da simulação de disfunção neurocognitiva (“*Malingered Neurocognitive Disfunction*” ou MND), Slick e colaboradores (1999) propuseram um modelo de identificação, que foi considerado um marco importante na investigação na área. Este termo foi definido pelos autores como o exagero ou fabricação intencional de disfunção cognitiva, que tem subjacente uma motivação externa, como a obtenção de benefícios, ou o evitamento de deveres formais ou responsabilidades. Neste sentido, propuseram três categorias de diagnóstico: *MND Definitivo*, *MND Provável*, ou *MND Possível* (Slick et al., 1999). No entanto, ao longo dos anos de investigação, diversas limitações têm sido encontradas nos critérios

originalmente propostos (ver Rogers & Bender, 2012; Sherman et al., 2020 para mais detalhes). Por essa razão, foi recentemente desenvolvida pelos autores originais (Sherman et al., 2020), uma reconceptualização do modelo, que teve como base o original, com expansões e definições melhoradas. Neste sentido, a proposta passa por uma categorização multidimensional da simulação neurocognitiva, psiquiátrica e somática, definida por evidência clara com base em quatro critérios essenciais: (*Critério A*) Presença de um incentivo externo; (*Critério B*) Apresentação inválida na avaliação, que indica exagero ou fingimento, podendo esta surgir através de inconsistências ou de evidência psicométrica em TVS ou TVD; (*Critério C*) Discrepâncias marcadas entre os relato de sintomas e os resultados obtidos nos testes; (*Critério D*) Os comportamentos descritos no critério B não podem ser explicados totalmente por outra condição médica, desenvolvimental ou psiquiátrica. Além disso, são definidos quatro tipos de simulação através deste modelo, que variam de acordo com os critérios (e subcritérios) que são preenchidos. Estes são a simulação de disfunção cognitiva; a simulação de sintomas somáticos; a simulação de sintomas psiquiátricos; e a simulação com apresentação mista (Sherman et al., 2020).

Têm sido diversos os estudos a avaliar a taxa de simulação, sendo esta estimativa muito difícil de conseguir e muito variável entre contextos (Bass & Halligan, 2014). De acordo com uma resenha de autores (como citados em Merten & Merckelbach, 2013b), a prevalência de pacientes com um autorrelato não genuíno de sintomas pode atingir ou mesmo exceder os 50%. Embora em contexto clínico se espere a existência de uma relação terapêutica baseada na honestidade, Brooks et al. (2012, como citados em Merten & Merckelbach, 2013a) encontraram uma prevalência de 37% de pacientes com envesamento negativo de resposta na apresentação dos seus sintomas. No entanto, conseguiram identificar potenciais ganhos secundários em 58% da amostra.

No que respeita os desempenhos em Testes de Validade de Sintomas (TVS), uma resenha de estudos (como citados em Martin & Schroeder, 2020) concluíram que, quando estão presentes incentivos externos, a taxa base de invalidez de desempenho pode ser estimada entre 20 e 50% dos avaliados. Martin e Schroeder (2020), através de uma investigação junto de neuropsicólogos em contextos clínicos não forenses, perceberam que a maior prevalência de invalidez encontrada foi com pacientes com perturbações somatoforme, de conversão, ou com sintomas clinicamente inexplicáveis, sendo o valor estimado de cerca de 50%, sem especificação da presença ou ausência de incentivo externo. Este valor foi seguido por uma estimativa de 40% em pacientes que se propunham a avaliação para receber os benefícios de veterano, e em pacientes em processos de litígio civil. Em contextos nos quais estariam presentes possíveis ganhos secundários, as estimativas variaram entre os 25 e os 37,5%. O valor mais baixo encontrado foi em pacientes que se apresentavam para avaliações de demência, e em pacientes para os quais a invalidez nos resultados dos testes teria sido antecipada – cerca de 5% (ver Martin & Schroeder, 2020 para mais detalhes). Embora tenham sido apenas estimados pelos neuropsicólogos participantes, através de

uma avaliação da sua experiência, estes valores são genericamente congruentes com os resultados obtidos em investigações anteriores através de métodos empíricos (Martin & Schroeder, 2020).

Relativamente aos contextos forenses, a simulação de sintomas constitui a maior ameaça à validade das avaliações psicológicas, dado que aí existe uma maior evidência de ganhos secundários. Os indivíduos que recorrem a este comportamento podem ter como subjacentes as motivações de justificar, desculpar ou atenuar as consequências do comportamento criminal, como uma redução da pena aplicada, ilibação ou atenuação da culpa, ou mesmo obtenção de cuidados de saúde especiais (Resnick & Knoll, 2018).

De acordo com Rogers et al. (1994; como citado em McDermott et al., 2013) cerca de 16% dos pacientes sujeitos a avaliações em contextos forenses têm comportamentos de simulação de sintomas, enquanto em contextos não forenses se estima um valor de apenas 7%. Merten e Merckelbach (2013a) calculam que, em contextos médico-legais, cerca de um terço dos pacientes com o diagnóstico de perturbação somatoforme, dissociativa ou de dor, apresentem um estilo de resposta com enviesamento negativo.

Além disso, e mais especificamente em avaliações da competência para testemunhar, McDermott e Sokolov (2009) estimam que a prevalência de avaliados que simulam sintomas ronda os 18%. Já em avaliações de responsabilidade criminal, a estimativa para a existência de simulação sobe para 21% dos arguidos (Rogers et al., 1986; como citado em McDermott et al., 2013). Também Cornell e Hawk (1989) concluíram que 8% dos sujeitos que seriam presentes a julgamentos, estariam a simular traços psicóticos. Um estudo de Chafitz e Underhill (2013, como citado em Bass & Halligan, 2014) concluiu que a prevalência de exagero de sintomas nos Estados Unidos da América varia entre 46% a 60%, em avaliações de incapacidade da segurança social. Também os membros do *American Board of Clinical Psychology* apontam para estimativas de 8% de comportamento de exagero de sintomas em casos médicos ou psiquiátricos, 19% em casos criminais, 29% em casos de dano pessoal, e 30% em casos de incapacidade ou de compensação de trabalhadores (Mittenberg et al., 2002, como citados em Bass & Halligan, 2014). Por fim, um estudo de McDermott e colaboradores (2013) determinou uma prevalência de 17.5% de simulação provável numa amostra de sujeitos em internamento psiquiátrico e envolvidos em processos de incompetência para ser julgado, processos de inimputabilidade por insanidade, ou outros processos, ou com sujeitos ofensores com perturbações mentais. Estes autores encontraram diferenças estatisticamente significativas na gravidade do crime cometido, sendo que os sujeitos acusados de homicídio tenderiam a envolver-se mais em comportamentos de simulação (McDermott et al., 2013).

Em contexto prisional estão, frequentemente, presentes um ou mais incentivos externos, de modo que a simulação pode ocorrer devido a variadas razões. De forma mais comum, os reclusos podem simular psicopatologia de modo a cumprir a pena com acesso a determinadas regalias, como a colocação em celas mais confortáveis ou alas menos restritivas, ter acesso a medicação e serviços de saúde mental, ou apenas pela atenção ou divertimento que dali retiram. Além disso, este comportamento de simulação pode surgir como uma

resposta adaptativa para os reclusos com algum grau de doença mental, de forma a conseguirem acesso aos serviços de saúde mental, ou mesmo como “desculpa” para o comportamento criminal, passando o mesmo a ser atribuído à perturbação (Guy & Miller, 2004; Huhnt, 2013; McDermott & Sokolov, 2009; Resnick & Knoll, 2018; Walters, 2006, como citado em McDermott et al., 2013).

Neste âmbito, quando se trata de populações prisionais, a prevalência da simulação aumenta. Numa prisão de média segurança, de acordo com Pollock et al. (1997) a taxa de simulação é de cerca de 32%, com base em dados de avaliações psicológicas. Um estudo de Norris e May (1998) indicou que cerca de 56% dos reclusos que pediram acompanhamento psicológico ou que reportavam queixas, simulavam sintomas. Este valor foi estimado por Walters et al. (1988, como citado em McDermott & Sokolov, 2009) em 46%. Também Guy e Miller (2004) estimaram uma prevalência de simulação em 42% com uma população de reclusos de um estabelecimento prisional de alta segurança. Da amostra analisada que estava a receber tratamento psicológico na altura da avaliação, a estimativa para a simulação foi de 52%, enquanto para os participantes que não se encontravam a receber tratamento psicológico, a prevalência estimada foi de 32% (Guy & Miller, 2004). McDermott e Sokolov (2009) realizaram um estudo com uma amostra de 161 reclusos suspeitos de simular sintomatologia e 66.5% da amostra reuniu os critérios para a categorização de simulação. No entanto, esta estimativa foi calculada apenas através da aplicação do *Structured Interview of Reported Symptoms* (SIRS), o que significa que o número poderá ser mais elevado do que a realidade. Um outro estudo de McDermott et al. (2013) realizado com diversas amostras forenses, sendo uma delas composta por reclusos que procuravam apoio psiquiátrico encontrou uma prevalência de simulação de 53% para os participantes em internamento, e de 67% para a restante população prisional.

Relativamente às características dos simuladores em contexto criminal, vários estudos comprovam que os simuladores tendem a ser menos escolarizados comparativamente aos não simuladores (Cornell & Hawk, 1989; Guy & Miller, 2004; McDermott & Sokolov, 2009; Norris & May, 1998). Foram também encontrados efeitos significativos na idade, sendo que os simuladores tendem a ser mais novos (Norris & May, 1998), no historial de reclusão anterior (McDermott et al., 2013), e na gravidade da ofensa, tendo a probabilidade de simulação sido estimada como superior em casos de sujeitos acusados de homicídio (Cornell & Hawk, 1989; Taylor & Kopelman, 1984, como citados em McDermott et al., 2013). No entanto, este último fator, referente à influência da gravidade da ofensa nos comportamentos de simulação, não foi comprovado pelo estudo de McDermott e Sokolov (2009). Nos diversos estudos, não foram encontradas diferenças significativas ao nível do género, do historial psiquiátrico ou de hospitalizações prévias, entre simuladores e não simuladores (Cornell & Hawk, 1989; Guy & Miller, 2004; McDermott & Sokolov, 2009).

3 - Testes de Validade de Desempenho (TVD) e Testes de Validade de Sintomas (TVS).

A primeira era das estratégias de detecção de exagero de sintomas e de outros estilos de resposta era baseada em estudos de casos e métodos sem prova empírica de sucesso (Rogers, 2018b). Uma das falhas apontadas a este tipo de métodos diz respeito ao facto de se basear nas características encontradas em simuladores suspeitos, que se pensava serem salientes e distintivas, para identificar os simuladores, provocando, nesse sentido, erros de enviesamento confirmatório (Borum et al., 1991, como citado em Rogers, 2018b).

Mais recentemente, a definição de estratégia de detecção foi formulada de modo a combater os erros anteriormente referidos. Rogers (2018b) refere esta como um método estandardizado empiricamente validado e com base conceptual, utilizada para diferenciar estilos de resposta enviesados de outros. A detecção da simulação é um fenómeno complexo, devido às interações existentes entre o tipo de sintomatologia simulada, as diferentes populações e contextos, e a heterogeneidade das estratégias de detecção e de rastreio (Smith, 2018).

Os estudos nesta área têm destacado a existência de diversas estratégias que poderão ser úteis para reunir evidência de simulação ou exagero de sintomas. Estas podem ser divididas em estratégias de detecção de simulação de perturbação mental, ou de dificuldades cognitivas, e são importantes na compreensão do funcionamento dos instrumentos desenhados para a avaliação deste estilo de resposta (Conroy & Kwartner, 2006; Rogers & Bender, 2003).

Mais especificamente na área do exagero de sintomas da doença mental, Rogers (1997, como citado em Conroy & Kwartner, 2006; Rogers & Bender, 2003) postula a existência de oito estratégias principais, que se dividem em duas categorias: a detecção por padrões não expectáveis, e a detecção por apresentação exagerada.

Na primeira categoria pode ser encontrada uma das mais robustas estratégias, referida como *sintomas raros*, que indica que os simuladores são, muitas vezes, ingénuos relativamente aos sintomas que uma certa perturbação envolve, e reportam alguns raramente experienciados por populações clínicas. Também a captação do relato de *sintomas improváveis*, com algum grau de fantasia ou absurdez, pelos simuladores de sintomas, é referido como estratégia a utilizar. Embora ocorra apenas numa minoria de sujeitos, deve ser um sinal para questionar a credibilidade do relato. Adicionalmente, deve ser tida em atenção a *combinação dos sintomas* reportados, pois com a falta de conhecimentos avançados, os avaliados incorrem, diversas vezes, no erro de relatar sintomas que, embora possam ser comuns sozinhos, raramente surgem simultaneamente (Rogers & Bender, 2012; Rogers, 2018b).

Na categoria dos sintomas exagerados, pode ser considerada a existência de cinco estratégias validadas. Em primeiro lugar, tendo em conta que a experiência dos pacientes genuínos passa, apenas, por um número mínimo de sintomas reportados como intoleráveis, a *severidade dos sintomas* indicada pelos simuladores deve ser tida em conta pois estes tendem a sobrestimá-la. Além disso, um erro muito comum dos simuladores é o relato

de *sintomas indiscriminado*, não havendo da parte destes, seletividade na escolha dos sintomas que reportam. Uma outra estratégia apresentada por Rogers (1997, como citado em Conroy & Kwartner, 2006; Rogers & Bender, 2003) diz respeito à tendência de os simuladores apresentarem frequentemente um relato composto por diversos *sintomas óbvios*, deixando de parte os sintomas mais subtis, que não seriam tão facilmente associados à psicopatologia. Adicionalmente, deve ser considerada a existência de *estereótipos errôneos* relativamente aos sintomas associados à doença mental, e que podem revelar pistas para a deteção de simulação. Por último, as inconsistências e discrepâncias apresentadas entre *sintomas observados e relatados* devem ser um sinal de alerta importante. No entanto, a utilização desta última estratégia deverá ter em conta a possibilidade da existência de falta de *insight* de alguns pacientes genuínos relativamente à sua perturbação (Rogers, 1997, como citado em Conroy & Kwartner, 2006; Rogers & Bender, 2003; Rogers & Bender, 2012; Rogers, 2018b).

Adicionalmente, são referidas algumas estratégias que requerem a realização de mais estudos de validação para poderem ser usadas com confiança, como é o caso da estratégia de *aproximação dos sintomas genuínos* e a de *especificação excessiva de sintomas* (Rogers, 2018b).

A respeito da deteção de casos de simulação de problemas cognitivos, é, também, proposta a divisão das estratégias principais por duas categorias: deteção por dano excessivo, e deteção por padrões não expectáveis (Rogers et al., 1993, como citados em Rogers & Bender, 2003; Rogers, 2018b).

A primeira categoria inclui as estratégias de *efeito de teto* e de *desempenho significativamente abaixo do acaso*. O *efeito de teto* utiliza tarefas e questões simples que são corretamente concluídas pela maioria dos pacientes com danos cognitivos genuínos, mas que podem ser falhadas pelos simuladores, que realizam avaliações incorretas das reais capacidades comprometidas em casos de défices genuínos (Millis & Kler, 1995, como citado em Brennan & Gouvier, 2006; Rogers & Bender, 2003; Rogers, 2018b). Esta é uma estratégia que deve ser usada com precaução devido à sua sensibilidade ao *coaching* (isto é, as instruções dadas por uma terceira parte, com o objetivo de alterar os resultados de uma avaliação, favorecendo a pessoa instruída) e aos défices cognitivos reais (Rogers & Bender, 2003). A estratégia do *desempenho significativamente abaixo do acaso* passa por uma avaliação baseada numa taxa de falha improvável calculada através de uma probabilidade estatística. Grande parte dos testes de validade de sintomas era, inicialmente, composto por questões de escolha forçada com duas opções de resposta, baseando o seu método de deteção na assunção de que, mesmo os respondentes honestos, não devem apresentar um desempenho significativamente abaixo do acaso (Conroy & Kwartner, 2006; Rogers & Bender, 2003; Rogers, 2018b).

No domínio dos padrões não expectáveis inserem-se as estratégias de deteção da *magnitude do erro*, da *curva de desempenho* e de *violação dos princípios de aprendizagem*. De acordo com Rogers (1993, como citado em Rogers, 2018b), a primeira corresponde à deteção de padrões de respostas incorretas não expectáveis, podendo ser respostas aproximadas das corretas,

ou respostas erradas de forma consistente e clara. Esta é considerada por Bender (2000, como citado em Rogers & Bender, 2003), a estratégia mais eficaz na identificação de simuladores, mesmo dos que receberam avisos sobre a sua existência. A *curva de desempenho* assenta no pressuposto de que os sujeitos com danos genuínos apresentam desempenhos expectáveis, que variam de acordo com a dificuldade dos itens apresentados. Por sua vez, os simuladores não consideram o grau de dificuldade na escolha dos itens que erram propositadamente, e revelam um padrão de resposta indiscriminado (Conroy & Kwartner, 2006; Rogers, 2018b; Rogers & Bender, 2003). A estratégia de *violação dos princípios da aprendizagem* assenta no pressuposto de que os simuladores não se encontram conscientes dos princípios básicos de aprendizagem, e não consideram as diferenças de desempenho que são esperadas tendo em conta estes princípios. Exemplos disso serão as diferenças entre os ensaios de reconhecimento e de recordação e entre os de recordação livre e de recordação com pistas (Rogers, 2018b).

Embora em artigos mais antigos se faça referência a mais duas estratégias como fazendo parte destes conjuntos, Rogers (2018b) percebeu que a falta de bases conceptuais não deveria permitir a sua junção com as restantes. Estas passaram, então, a ser referidas como métodos não estratégicos para deteção de défices cognitivos, pois baseiam a sua análise em desempenhos mais baixos ou diferentes dos previstos por pacientes genuínos. Em primeiro lugar, surgem as *medidas de escolha forçada*, que se apoiam em dados normativos para avaliar os desempenhos dos indivíduos nos testes, envolvendo comparações entre grupos e estabelecimento de pontos de corte ótimos (Rogers & Bender, 2003; Rogers, 2018b). Este é, no entanto, um método bastante criticado devido aos seus baixos valores de sensibilidade (Haines & Norris, como citados em Brennan & Gouvier, 2006). Adicionalmente, surge o método de *padrão atípico em testes* que identifica as diferenças entre os desempenhos de simuladores em certas escalas, comparativamente aos dos pacientes com défices genuínos (Rogers, 2018b).

Além disso, são referidas na literatura algumas estratégias que precisam de mais validação para poderem ser usadas com confiança. São elas a *consistência entre itens comparáveis* e a *frequência de sintomas* (Rogers, 2018b).

Para avaliar se uma estratégia é suficientemente eficiente na deteção dos estilos de resposta, Roger e Bender (2013, como citados em Rogers, 2018b), postularam a necessidade de a mesma estar estabelecida e validada em diferentes medidas, produzir elevados tamanhos de efeito e ter bons valores de precisão de diagnóstico. Além disso, importa considerar que a deteção de diferentes estilos de resposta requer a utilização de diferentes estratégias de deteção, na medida em que estas últimas não são universais, mas sim específicas para determinado estilo de resposta (Rogers, 2018b).

Neste sentido, começaram a ser desenvolvidos instrumentos com o objetivo de detetar distorções de resposta, e que tinham como apoio as estratégias anteriormente mencionadas. A assunção principal na construção destes instrumentos é a de que os simuladores têm conhecimento limitado dos verdadeiros sintomas neuropsicológicos (Jelicic et al., 2007). Estes são

identificados através de investigação empírica, e as regras de tomada de decisão são partilhadas, de modo que os resultados podem ser facilmente comunicados e replicados. Isto significa que, com a criação destes testes, a simulação de sintomas entrou no campo de avaliação baseado na evidência (Merten & Merckelback, 2013a; Merten et al., 2016).

Embora no início fossem referidos como instrumentos de deteção da simulação de sintomas, o facto de os Testes de Validade de Sintomas (TVS) terem, apenas, a capacidade de detetar os estilos de resposta enviesados e não as motivações subjacentes a tal, levou a um abandono dessa conceção (Impelen et al., 2017; Merten & Merckelback, 2013a). Além disso, os TVS eram inicialmente concebidos como medidas de escolha forçada, nos quais era aplicada a estratégia de deteção do desempenho significativamente abaixo do acaso. Com os desenvolvimentos na investigação esta foi, também, uma conceção a ser progressivamente abandonada (Bender & Frederick, 2018).

Assim, os testes autónomos mais específicos de avaliação de simulação incluem os Testes de Validade de Desempenho (TVD; *Performance Validity Tests*), que detetam o esforço insuficiente ou o baixo desempenho intencionais e os Testes de Validade de Sintomas (TVS; *Symptom Validity Tests*), que avaliam o exagero e a simulação de sintomas (Merten & Merckelback, 2013a; Merten et al., 2013; Merten et al., 2016a). Tendo em conta a relativa independência e falta de correlação dos dois tipos de simulação e o facto de os testes terem mostrado diferentes taxas de efetividade nos diferentes domínios, esta divisão é de elevada relevância (Rogers et al., 2011, como citados em Bender & Frederick, 2018).

Os primeiros (TVD) incluem medidas independentes geralmente disfarçadas de medidas de avaliação de domínios neuropsicológicos, como a velocidade de processamento ou a memória. Exemplos destes são o *Test of Memory Malingering* (TOMM; Tombaugh, 1996, 1997; como citado em Simões, Sousa, et al., 2017) ou o Teste de Memória de Rey – 15 itens (15-IMT; Rey, 1964, como citado em Vilar et al., 2017), que funcionam como medidas independentes e incidem nos défices de memória simulados. Os TVD podem, ainda, ser escalas integradas em medidas mais abrangentes, que mesmo não tendo sido criadas com esse fim, acabaram por desempenhar essa função, ou mesmo ser uma avaliação derivada de desempenhos atípicos em testes padrão, como os desempenhos na Figura Complexa de Rey (Bender & Frederick, 2018, Larrabee, 2015).

O TOMM é um instrumento de reconhecimento visual composto por 50 itens, administrado em três ensaios – dois de aprendizagem e um último de reconhecimento. A sua utilidade na distinção entre défices cognitivos genuínos e comportamentos de esforço insuficiente ou de simulação de problemas mnésicos deve-se ao facto de parecer um instrumento mais complexo pelo seu elevado número de itens, mas ser, na verdade, bastante fácil de concluir corretamente (Simões, Sousa, et al., 2017). Para tentar combater os resultados fracos de sensibilidade encontrados em diversos estudos, foi proposta uma adaptação dos pontos de corte dos ensaios para interpretação dos resultados, levando a uma otimização na precisão do

diagnóstico por ele realizado (ver Martin et al., 2020 para mais detalhes; Rogers & Bender, 2012).

O 15-IMT é um instrumento constituído por apenas 15 itens, que tem como objetivo examinar a credibilidade das queixas cognitivas dos avaliados, bem como detetar desempenhos em que tenha sido aplicado esforço insuficiente ou reduzida motivação. Tem na sua base de construção a estratégia de efeito de teto e o objetivo da tarefa é, então, que sejam memorizados os 15 itens apresentados - que podem ser agrupados em ideias - e reproduzidos, de seguida (Ensaio de Evocação Imediata – EI). À semelhança do instrumento anteriormente apresentado, tem revelado valores modestos de sensibilidade (Rogers & Bender, 2012; Vilar et al., 2017). De modo a atenuar esta falha, Boone et al. (2002) desenvolveram um Ensaio de Reconhecimento (ER), no qual são apresentados ao sujeito os 15 itens anteriores, junto de outros 15 itens novos. A tarefa passará, então, por reconhecer os itens originais no conjunto total apresentado. Assim, é possível obter valores de falsos positivos, itens reconhecidos corretamente, e um resultado combinado do reconhecimento (RCR). Adicionalmente, é permitida uma análise do tipo de itens assinalados como falsos positivos (ver Boone et al., 2002 para mais detalhes). Os estudos realizados em Portugal demonstram uma necessidade de adaptação dos pontos de corte consonantes com a natureza da população a avaliar (Vilar et al., 2017).

Pode ser considerado ainda o *Word Memory Test* (Green, 2003; como citado em Denney, 2008) como relevante para a avaliação de desempenho. Este teste, com uma base empírica forte, é administrado via computador, e apresenta, de acordo com os resultados dos estudos realizados, valores de sensibilidade e especificidade muito bons (Denney, 2008).

Relativamente aos TVS, podem, também, ser escalas integradas em inventários clínicos ou da personalidade, ou medidas independentes (Giger & Merten, 2019; Merten & Merckelbach, 2013a; Merten et al., 2016). As escalas integradas mais comuns são as constituintes do *Minnesota Multiphasic Personality Inventory – 2* (MMPI-2; Butcher et al., 2001, como citado em Wall et al., 2015) e no *Personality Assessment Inventory* (PAI; Morey, 2007, como citado em Merten et al., 2016). Na categoria das medidas independentes, desenvolvidas para a avaliação do relato de sintomas distorcido, são as mais utilizadas o *Structured Inventory of Malingered Symptomatology* (Inventário Estruturado de Simulação de Sintomas; SIMS; Smith & Burger, 1997; Widows & Smith, 2005) e o *Structured Interview of Reported Symptoms* (SIRS; Rogers et al., 1992; SIRS-2; Rogers, et al., 2010). Importa, ainda, referir o *Self-Report Symptom Inventory* (SRSI; Merten et al., 2016), um instrumento recentemente desenvolvido para colmatar as falhas de outros, e a Escala de Validade de Sintomas (EVS-2, Cunha, 2011; Simões et al., 2015; como citados em Simões, Almiro, Nunes, et al., 2017), um instrumento desenvolvido em Portugal.

De uma forma geral, os TVS têm de possuir certas características para serem eficazes na deteção da simulação. Entre outras, é essencial que apresentem boas características psicométricas (como sensibilidade, especificidade, validade facial), insensibilidade a psicopatologia genuína,

resistência ao *coaching*, facilidade de administração e cotação, normas constantemente atualizadas, cobertura de sintomas mais prováveis de simular, e sensibilidade à prevalência diferencial. Este último requer que o instrumento produza pontuações mais altas para populações onde a simulação é comum, e pontuações mais baixas para as populações onde a simulação é mais rara (Impelen et al., 2014). Adicionalmente, importa perceber a apropriação das dimensões relevantes medidas pelo teste às queixas apresentadas pelo avaliado (Seron, 2014).

Kessel e Zimmerman (1993, como citados em Rogers et al., 1996), utilizaram, em adição à taxa de acertos, quatro medidas do rigor de diagnóstico que são críticas na avaliação da eficácia de uma medida utilizando determinado ponto de corte: a sensibilidade, a especificidade, o poder preditivo positivo (PPP) e o poder preditivo negativo (PPN). A sensibilidade diz respeito à proporção de simuladores identificados corretamente com uma pontuação elevada, enquanto a especificidade diz respeito à proporção de não simuladores identificados corretamente com uma pontuação baixa. Pelo contrário, o PPP refere-se à probabilidade de uma pessoa com uma pontuação elevada ser simulador, e o PPN à probabilidade de uma pessoa com uma pontuação baixa ser não simulador (Martin & Schroeder, 2020; Rogers et al., 1996; Rogers, 2018b). No entanto, é importante considerar que o cálculo destes dados para cada medida depende do ponto de corte utilizado (Rogers, 2018b) e da taxa de base de simulação na população em estudo, o que nem sempre é ponderado nas investigações com este objetivo (Rosenfeld et al., 2000, como citados em Martin & Schroeder, 2020; Wisdom et al., 2010).

A avaliação da validade de sintomas deve, então, ser sempre realizada recorrendo a diversos testes e múltiplas estratégias de deteção, devido à sua complexidade e às vastas consequências para os que são incorretamente avaliados como simuladores, como o bloqueio do acesso a serviços de que necessitariam (e.g., cuidados de saúde mental) ou aos benefícios por incapacidade. Também o estigma poderá ser uma consequência a longo prazo destes erros de identificação de simuladores, podendo ter implicações em diagnósticos, acesso a serviços e benefícios futuros (Bender & Rogers, 2004, como citados em Bender, 2018; Larrabee, 2012, como citado em Bender, 2018; Rogers et al., 1996; Widows & Smith, 2005, como citado em Cernovsky et al., 2019).

4 - Testes de Validade de Sintomas (TVS).

No domínio específico da avaliação do exagero ou simulação de sintomas psicopatológicos, podemos considerar a relevância de diversos testes.

Entre os TVS mais usados destaca-se o SIMS, cuja taxa de utilização por neuropsicólogos e profissionais de saúde mental foi estimada entre 10% (Martin et al., 2015) a 21% (de acordo com Hill & Demetriooff, 2019, como citados em Edens et al., 2020). É um teste constituído por uma lista de 75 itens que remetem para pseudossintomas, isto é, sintomas atípicos, bizarros e raros, mas que, aos olhos de pessoas leigas, parecem normais. Este instrumento foi, então, construído para detetar simulação em diferentes domínios

Estudos com *design* de simulação análogo em Testes de Validade de Sintomas (TVS):

Uma revisão da literatura

Carolina Duarte de Almeida (carolinad Almeida2@gmail.com) 2020

neuropsicológicos: *Inteligência Baixa* (LI, “*Low Intelligence*”); *Perturbações Afetivas* (AF, “*Affective Disorders*”); *Danos Neurológicos* (N, “*Neurologic Impairment*”), *Psicoses* (P, “*Psychosis*”) e *Perturbações Mnésicas* (AM “*Amnestic Disorders*”). A base deste instrumento reside na hipótese de que uma pessoa que simula ou exagera sintomas irá relatar um número considerável desses sintomas, de modo a parecer convincente na sua sintomatologia (Huhnt, 2013; Impelen et al., 2014; Merckelbach & Smith, 2003; Smith & Burger, 1997; Simões, Almiro, Mota, et al., 2017). Os resultados do estudo original de Smith e Burger (1997) com a primeira versão deste instrumento, indicaram níveis de consistência interna entre escalas aceitáveis, com valores de fiabilidade entre os 0.80 e os 0.88. Por outro lado, as correlações entre as escalas mostraram-se, no geral, baixas, com exceção da correlação entre as *escalas P e NI*. Além disso, os resultados encontrados através das análises estatísticas revelaram que todos os índices e os resultados da escala total eram significativamente capazes de diferenciar os simuladores dos participantes do grupo de controlo, como taxas de sensibilidade de 95% e de especificidade de 88% (Smith & Burger, 1997).

Adicionalmente, têm sido realizados vários estudos ao longo dos anos com o objetivo de avaliar as propriedades psicométricas deste instrumento, com os resultados a mostrarem-se promissores (e.g. Ardic et al., 2019; Dandachi-Fitzgerald et al., 2020; Jelicic et al., 2006; Parks et al., 2017). Também uma meta-análise de diversos estudos realizada por Impelen et al. (2014) revelou diversas vantagens deste instrumento, entre as quais o baixo nível de dificuldade de compreensão verbal e a independência dos resultados de condições como o género, a idade o nível de escolaridade.

Por outro lado, vários autores têm sido críticos relativamente à utilização do SIMS, tendo encontrado algumas limitações e desvantagens no mesmo. Impelen e colaboradores (2014) concluíram que os valores de especificidade encontrados nos diversos estudos analisados eram muito divergentes e com maior incidência para os valores mais baixos, e que os valores de consistência interna tinham tendência a variar consoante as características da amostra. Foram, também, encontradas falhas no índice de precisão de diagnóstico das subescalas do instrumento, sendo recomendado a utilização do resultado da escala completa. Adicionalmente, foi percebida a sensibilidade dos seus resultados a alguns tipos de psicopatologia genuína, podendo significar uma sobrestimativa da deteção de simulação pelo SIMS (Dandachi-Fitzgerald et al., 2020; Impelen et al., 2014; Parks et al., 2017). Cernovsky et al. (2019) apontam, também, algumas falhas no processo de construção deste instrumento, e alertam para a presença de itens que referem psicopatologia genuína, e de outros considerados inapropriados mesmo por leigos. Os autores concluem, mesmo, que este instrumento não foi validado de forma apropriada de acordo com as normas da APA, e que a sua utilização enquanto teste de deteção de simulação constitui uma prática profissional incorreta. De forma semelhante, a sua constituição exclusiva por pseudossintomas torna fraca a sua validade facial, na medida em que a probabilidade de ser facilmente reconhecido como instrumento de validade de sintomas se torna elevada. Por último, a sua precisão de diagnóstico é superior

em contextos de justiça criminal, relativamente a contextos de direito civil (Giger & Merten, 2019; Huhnt, 2013; Impelen et al., 2014; Merten et al., 2016).

Rogers e colaboradores (2014), de forma a tentar compensar o processo de construção das subescalas do instrumento (que não teve por base nenhuma estratégia de deteção), construíram duas subescalas que visavam avaliar o relato de *Sintomas Raros* (RS; “*Rare Symptoms*”) e de *Combinação de Sintomas* (SC; “*Symptom Combination*”). Os autores encontraram valores entre moderados e elevados de sensibilidade e de especificidade para ambas as escalas (Rogers et al. 2014). Edens et al. (2020), procuraram testar a precisão de classificação destas escalas, revelando resultados promissores na sua utilização, a um nível geral. Estes autores encontraram ainda problemas de generalização dos resultados com os pontos de corte originalmente propostos (Edens et al., 2020).

Considerando as conclusões dos diversos estudos, os resultados do SIMS não devem ser considerados evidência conclusiva da presença/ausência de psicopatologia simulada, sendo indispensável a utilização de diversos TVS. Também os autores originais do instrumento de avaliação (Smith & Burger, 1997) deixaram o aviso de que o diagnóstico de simulação deve ser feito com muita precaução, apenas após uma avaliação extensiva e quando os participantes são detetados por todos os instrumentos de avaliação de simulação utilizados. Isto permite aumentar a acuidade de deteção de simulação, reduzindo os falsos positivos.

O *Structured Interview of Reported Symptoms* (SIRS), referido em contextos clínicos e de investigação como sendo o “modelo de ouro” (“*gold standard*”) dos instrumentos de deteção de simulação, é uma entrevista estruturada administrada verbalmente. É constituído por 172 questões que combinam múltiplas estratégias de deteção, como os *sintomas raros*, *combinações de sintomas*, *sintomas improváveis*, e *inconsistências entre os sintomas relatados e os observados* (Rogers et al., 2009, como citado em Green & Rosenfeld, 2011). As suas pontuações são geradas com base no número de itens reportados em cada uma das oito escalas principais, sendo o teste constituído, ainda, por 4 escalas suplementares e uma quinta que avalia a defensividade. Os resultados de cada escala podem resultar em classificações de “*honesto*” (“*honest*”), “*indeterminado*” (“*indeterminate*”), “*simulação provável*” (“*probable malingering*”), ou “*simulação definitiva*” (“*definite malingering*”) (Green & Rosenfeld, 2011).

Green e Rosenfeld (2011), através de uma meta-análise de diversos estudos realizados com o SIRS, concluíram que este parece ser um bom instrumento na discriminação entre simuladores e respondentes honestos, com os resultados a apontarem para uma boa precisão de classificação. No entanto, os resultados de especificidade quase perfeitos obtidos no estudo de desenvolvimento do instrumento (Rogers et al., 1992, como citados em Green & Rosenfeld, 2011) não foram replicados, especialmente em amostras atípicas. Também os tamanhos de efeito calculados foram bastante variáveis entre diferentes tipos de *design* de estudo e entre diferentes tipos de comparações de grupos. Por outro lado, os valores de sensibilidade

encontrados por grande parte dos estudos analisados, foram superiores aos inicialmente estimados pelos autores originais (Green & Rosenfeld, 2011). Um outro estudo de McDermott e Sokolov (2009), realizado com uma população criminal, encontrou valores de precisão de diagnóstico de simulação com o SIRS bastante elevados. No entanto, este foi um estudo com diversas limitações metodológicas, nomeadamente a utilização de, apenas, uma amostra referida como simuladora após suspeitas fundadas em julgamento clínico, sem realização de comparação com grupo de controlo ou aplicação de outros TVS. Isto torna a interpretação destes resultados falaciosa (McDermott & Sokolov, 2009). Também Rogers e colaboradores (2010), num estudo realizado com uma amostra em internamento psiquiátrico numa unidade de saúde mental forense de alta segurança, concluíram que as escalas principais do SIRS revelaram consistência e precisão nos resultados entre diferentes avaliações. Considerando estas conclusões, apesar de ser um instrumento eficaz na discriminação entre simuladores e respondentes honestos, o SIRS deve ser usado com precaução (Green & Rosenfeld, 2011). Além disso, apresenta as desvantagens de exigir um longo tempo de administração e a necessidade de treino prévio para os seus utilizadores (Smith & Burger, 1997).

Recentemente, os autores originais desenvolveram uma segunda versão do instrumento (SIRS-2), com modificações ao nível do exame das classificações obtidas. Esta mantém as escalas primárias originais, com adição de novas escalas (*Rare Symptoms Total Index - RS-Total* e *Modified Total Index - MT*), e conta com um novo modelo de tomada de decisão. As pontuações obtidas através do índice MT podem resultar em duas novas classificações: *indeterminado a avaliar* (“*indeterminate-evaluate*”) e *indeterminado geral* (“*indeterminate-general*”). No caso de o sujeito ser categorizado com a primeira, a sua avaliação para deteção de simulação deverá continuar. O último caso indica baixa probabilidade de se tratar de um caso em que ocorre simulação de sintomas (Correa, 2018).

No estudo original de validação do SIRS-2, Rogers et al. (2010, como citados em Rogers, 2018d) encontraram valores de sensibilidade e de especificidade elevados (0.80 e 0.975, respetivamente), e um valor de 2.5% para a taxa de falsos positivos. Uma resenha de estudos (citados em Rogers, 2018d) revelou valores elevados de fiabilidade interavaliadores, bem como níveis de confiança de classificação altos, com baixa variabilidade desse valor nas escalas principais. Por outro lado, a utilização do modelo de decisão proposto no SIRS-2 revelou vantagens relativamente ao modelo original. De acordo com Rogers (2018d) na segunda edição do SIRS são minimizados problemas com a literacia dos pacientes na resposta ao instrumento, além de fornecer sugestões importantes respeitantes à distinção entre a simulação de sintomas e a perturbação factícia. Green e colaboradores (2013) realizaram um estudo que permitiu concluir que a taxa de especificidade obtida com o SIRS-2 superou o valor obtido com a versão original, com um número inferior de falsos positivos, com recurso ao novo modelo de decisão. Por outro lado, a sensibilidade de ambas as versões revelou ser relativamente baixa, principalmente em pacientes clínicos suspeitos de simulação ou exagero de

sintomas, com valores da primeira versão a surgirem superiores. Algumas dúvidas se colocaram, ainda, na relevância da utilização do índice MT enquanto diferenciador de pacientes genuínos e simuladores (Green et al., 2013). Também um estudo de Glassmire et al. (2016) revelou uma correlação superior entre os resultados obtidos com o M-FAST (*Miller-Forensic Assessment of Symptoms Test*) e os resultados obtidos com a versão original do SIRS, quando comparados com os obtidos com o SIRS-2. Adicionalmente, Rubenzer (2010) e DeClue (2011) apresentam nos seus trabalhos algumas críticas ao desenvolvimento do SIRS-2, incluindo a falha existente por parte dos autores a respeito do cálculo das estimativas de precisão de classificação, bem como a ausência de dados importantes no manual (ver Rubenzer, 2010; DeClue, 2011 para mais detalhes). Não obstante, Rubenzer (2010) apoia a valorização e a utilização do SIRS-2 por parte dos profissionais que não devem, no entanto, apoiar exclusivamente as suas decisões nos resultados deste instrumento.

Em Portugal deve ser considerada, ainda, a Escala de Validade de Sintomas (EVS-2, Cunha, 2011; Simões, et al., 2015; como citados em Simões, Almiro, Nunes, et al., 2017). Este é um questionário breve de autorrelato que pode ser administrado a sujeitos com mais de 18 anos. O objetivo é que seja usado como complemento aos instrumentos anteriormente referidos, tendo o princípio de construção usado sido o mesmo que o do SIMS. Neste sentido, é constituído por 48 itens de resposta dicotómica (Verdadeiro/Falso), distribuídos por três subescalas: a *Subescala Perturbações Emocionais* (PE), que avalia o exagero de sintomas relativos aos estados depressivo e de ansiedade, e que incide em sintomas de humor, apetite e sono pouco comuns; a *Subescala Psicose* (P), relativa ao exagero de sintomas relacionados com as perturbações psicóticas, e que incide em sintomas bizarros e combinações pouco comuns; e a *Subescala Problemas Cognitivos* (PC), que avalia o exagero ou fabrico de sintomas relacionados com a memória, concentração e atenção recorrendo a itens que apresentam sintomas pouco frequentes e itens fáceis, geralmente acertados por pessoas com perturbações genuínas. Os estudos realizados com este instrumento recorreram a diversos grupos amostrais da comunidade e forenses. As amostras forenses encontravam-se presentes em contextos médico-legais, no âmbito de processos de litígio laboral, de regulação das responsabilidades parentais, ou de diversas perícias psicológicas. Os resultados dos diversos estudos mostraram valores de consistência interna a oscilar entre muito bons e respeitáveis nas subescalas, com exceção da subescala PC, que revelou um resultado inferior. Na Escala Total, os valores encontrados variaram entre excelentes e respeitáveis, nos diferentes estudos. Estes valores revelaram-se superiores nas amostras forenses, comparativamente às amostras da comunidade. Da mesma forma, os resultados ao nível da validade interna e de validade convergente com outros TVS revelaram-se promissores. Por outro lado, limitações surgem ao nível da validade discriminante, na distinção entre psicopatologia simulada e genuína, à semelhança do que acontece com o SIMS. Considerando estes resultados, a EVS-2 parece ser um instrumento com potencialidades de utilização em casos de rastreio de comportamentos de

simulação, quando integrado num protocolo mais vasto (Simões, Almiro, Nunes, et al., 2017).

O *Self-Report Symptom Inventory* (SRSI; Merten, et al., 2016) é um instrumento composto por 107 itens, que surgiu com o objetivo de responder às necessidades cada vez maiores da clínica e da forense, e de colmatar as falhas identificadas nos outros instrumentos de deteção do exagero de sintomas. Foi desenvolvido em 2006, e combina uma Escala de *Sintomas Genuínos* com uma Escala de *Pseudossintomas*, permitindo proceder ao cálculo do resultado total, dos resultados de cada uma das escalas, e do resultado das subescalas presentes nestas. O que se espera com a utilização deste instrumento é que pacientes honestos com sintomatologia genuína reportem sintomas clínicos, mas não pseudossintomas – ou em baixa frequência (Giger & Merten, 2019; Geurten et al., 2018; Merten et al., 2016).

A adição da escala de *Sintomas Genuínos* a este instrumento permitiu diminuir a probabilidade de identificação deste como teste de validade de sintomas, melhorando a sua validade facial quando comparado com outros que não a incluem, e torná-lo, também, mais resistente ao *coaching*. O desenvolvimento da escala de *Pseudossintomas* utilizou um princípio semelhante àquele referido no SIMS (e EVS-2), ou seja, recorrendo a itens que, aos olhos de pessoas leigas, passam por sintomas genuínos. No entanto, esta escala contém itens diferentes e menos extremos do que os apresentados no SIMS, de modo a poder servir de alternativa ou complemento ao mesmo. O SRSI é um instrumento que cobre um espectro de sintomas tipicamente reportados nos campos de litígio civil, lei pública e administrativa, compensação de trabalhadores, e outros ramos de segurança social, adequando-se mais à utilização nestes contextos. Pelo contrário, de acordo com os sintomas que apresenta, o SIMS adequa-se mais a contextos criminais (Geurten et al., 2018; Merten et al., 2016).

Os estudos realizados com este instrumento têm revelado resultados promissores. Em primeiro lugar, tem sido apoiada a sua capacidade de discriminação entre participantes de grupos de controlo e simuladores instruídos (Boskovic et al., 2018; Boskovic, Merckelback, et al., 2019; Geurten et al., 2018; Giger & Merten, 2019; Helvoort et al., 2019). Os valores de consistência interna das escalas têm sido estimados entre muito bons e excelentes (Geurten et al., 2018; Huhnt, 2013; Merten et al., 2016), e os resultados para a estabilidade temporal foram, também, positivos (Rafdal, 2013, como citado em Merten et al., 2016; Schlicht & Merten, 2014, como citados em Merten et al., 2016). Relativamente à sensibilidade e especificidade das escalas, os valores apresentados pelos investigadores não se têm revelado tão consistentes. Embora alguns dos estudos tenham reportado resultados elevados nestes valores (e.g. Geurten et al., 2018; Giger & Merten, 2019), outros têm encontrado uma certa fragilidade na sua especificidade (e.g. Akca et al., 2020; Boskovic et al., 2018; Boskovic, Hope, et al., 2019; Huhnt, 2013), ou sensibilidade (e. g. Akca et al., 2020; Boskovic, Merckelbach, et al., 2019) em casos de simulação de perturbação da dor. Por outro lado, um estudo de Helvoort e colaboradores (2019) veio contradizer os baixos valores de especificidade, revelando resultados que indicam que o SRSI não produz erros

de falsos-positivo em casos de pacientes com sintomatologia genuína.

Em Portugal, foram realizados estudos com o SRSI no âmbito de dissertações de mestrado. Dwarkadas (2018) procedeu à administração da versão experimental do SRSI a uma amostra de estudantes, até obter a versão final. De seguida, foram utilizados grupos de estudantes universitários e da comunidade para iniciar o processo de validação do instrumento para a população portuguesa, com análise das suas propriedades psicométricas e determinação de dados normativos para contexto comunitário. Para isso, recorreu à utilização de outros instrumentos de validade de sintomas (SIMS e EVS-2), bem como de escalas clínicas, e de uma escala de avaliação da desejabilidade social, que foram administrados, por adição ao SRSI. Os resultados de análise das propriedades psicométricas do TVS revelaram valores de consistência interna abaixo daqueles encontrados por Merten e colaboradores (2016) para as subescalas e escala de *Pseudossintomas*, variando entre inaceitáveis e respeitáveis. Por outro lado, na escala e subescalas de *Sintomas Genuínos*, os valores revelaram-se semelhantes aos encontrados no estudo supramencionado. Adicionalmente, importa considerar a presença de algumas correlações problemáticas. As correlações elevadas entre as pontuações da escala *Pseudossintomas* e os resultados obtidos no SIMS e EVS-2 seriam expectáveis, mas não as destes últimos com a escala *Sintomas Genuínos*. Também os resultados nos instrumentos de avaliação de psicopatologia aplicados revelaram correlações com ambas as escalas do SRSI, o que pode significar, ao contrário do encontrado anteriormente, alguma sensibilidade à psicopatologia genuína. Relativamente aos dados normativos, conclui-se um baixo número de participantes com pontuações superiores aos pontos de corte sugeridos por Merten et al. (2016) e um maior relato de sintomas pertencentes à escala de *Sintomas Genuínos*, comparativamente aos da escala de *Pseudossintomas* (Dwarkadas, 2018).

Pinheiro (2019) iniciou o trabalho de validação do SRSI em contexto prisional, recorrendo, para isso, à administração do instrumento em análise, do SIMS, e de outras escalas clínicas e de desejabilidade social, a uma amostra de 85 reclusos do sexo masculino. Foram encontrados valores de consistência interna bons, e correlações entre ambas as escalas do SRSI e o SIMS. Relativamente às correlações com os instrumentos de avaliação de psicopatologia, estas revelaram-se superiores com a escala e subescalas de *Sintomas Genuínos*, mas surgiram, também significativas com a escala de *Pseudossintomas*. A natureza dos sintomas psicopatológicos avaliados por cada um dos instrumentos utilizados teve, também, peso na subescala do SRSI com a qual obtiveram correlação superior. Pinheiro (2019) procedeu, também, à análise da influência das características jurídico-penais e demográficas da amostra nos resultados obtidos. Os resultados apontam para a influência da toma de medicação nos valores obtidos pelos reclusos avaliados, com tendência para um maior número de pseudossintomas relatados no caso de sujeitos que se encontravam medicados. Além disso, também o número de pseudossintomas relatados se revelou maior com um menor nível de escolaridade dos reclusos (Pinheiro, 2019).

Por último, Domingues (2019) recorreu a um *design* de simulação

análogo para realizar a validação do SRSI em contexto universitário. Neste âmbito, os participantes foram instruídos a responder de um modo honesto na primeira aplicação do SRSI, rematando com a resposta ao SIMS. Seguiu-se a apresentação de dois cenários de simulação, distribuídos de forma aleatória, que instruíam os participantes a responder de novo ao SRSI de modo a fingir sofrer de ansiedade ou de dor crónica. Os resultados obtidos para a consistência interna do SRSI assemelharam-se aos obtidos no estudo de Dwarkadas (2018), variando entre inaceitáveis e muito bons para as subescalas *Sintomas Genuínos* e entre inaceitáveis e aceitáveis para as subescalas de *Pseudossintomas*, no caso da condição “honesto”. Na condição “simulação”, a consistência das subescalas *Sintomas Genuínos* apresenta variações qualitativas semelhantes às referidas anteriormente, enquanto nas escalas de *Pseudossintomas* as consistências são todas muito boas. Uma análise das estatísticas descritivas apuradas através deste estudo indica que o número de *Sintomas Genuínos* reportado foi superior ao de *Pseudossintomas* na condição “honesto”, mas também na condição de “simulação”. É ainda claro o aumento do número de sintomas relatados entre a primeira e a segunda aplicação do SRSI, o que comprova o empenho dos participantes no exagero e/ou simulação de sintomas. O relato de *Pseudossintomas* em menor número poderá ter sido devido à sua aparência atípica, que levava a alguma relutância dos participantes em assinalá-los (Domingues, 2019).

De um modo geral, os resultados obtidos nos estudos com o SRSI não permitem a utilização deste em alternativa a nenhum dos anteriores, como o SIMS, mas sim como forma de complemento, devido à existência de limitações claras em ambos. É, também, visível a necessidade de proceder a estudos futuros com este instrumento recente, de modo a melhorar a confiança com que pode ser utilizado.

O *Minnesota Multiphasic Personality Inventory – 2* (MMPI-2) é um dos instrumentos de avaliação mais utilizado por profissionais da saúde mental, tanto em contextos clínicos como forenses (de acordo com diversos autores, como citados em Sellbom et al., 2010). Tem sido, também, um dos instrumentos mais estudados na área da validade de sintomas, pela eficácia das suas escalas *Infrequência* (“*Infrequency*”) na deteção da simulação e exagero de sintomas (Rogers et al., 2003, como citados em Wall et al., 2015). Na sua composição, o MMPI-2 contém diversas escalas que avaliam a validade de sintomas, tanto na área do exagero como da defensividade. No exagero ou simulação de sintomas, a escala *Respostas Infrequentes* (“*Infrequent Responses*”, F) utiliza os sintomas raramente assinalados por pacientes genuínos como método de deteção; as escala *Respostas Infrequentes de Psicopatologia* (“*Infrequent Psychopathology Responses*”, Fp) e *Respostas Somáticas Infrequentes* (“*Infrequent Somatic Responses*” Fs) detetam o relato de sintomas psicopatológicos e somáticos severos, respetivamente, sendo que a última inclui, também, a presença de itens correspondentes a sintomas raramente reportados por pacientes genuínos; a escala *Validade de Sintomas* (“*Symptom Validity*”, *Fake Bad Scale*, FBS) é utilizada para detetar a presença de queixas somáticas e neurocognitivas não credíveis em contextos médico-legais e civis; por último, a escala *Enviesamento Negativo de*

Resposta (“*Negative Response Bias*”, RBS) é utilizada na avaliação do exagero de problemas cognitivos (de acordo com uma resenha de autores, como citados em Wall et al., 2015). Diversos estudos com o MMPI-2 e com a sua forma reestruturada (MMPI-2-*Restructured Form*) têm sido realizados, para avaliar a capacidade de discriminação destas escalas entre participantes de controlo e simuladores, e os resultados têm-se revelado, no geral, positivos (e.g. Lewis et al., 2002; Marion et al., 2011; Sellbom et al., 2010; Sullivan & Elliot, 2014; Wall et al., 2015; Wygant et al., 2009).

Por último, é importante referir o *Personality Assessment Inventory* (PAI), um instrumento constituído por 344 itens distribuídos por 22 escalas não sobrepostas de avaliação clínica, de tratamento, interpessoais, e de validade de sintomas. Relativamente às últimas, podemos considerar a escala *Negative Impression Management* (NIM), que utiliza a estratégia dos sintomas raros para avaliar a existência de uma apresentação exageradamente desagradável do sujeito de si mesmo; a *Positive Impression Management* (PIM) que, pelo contrário, avalia a tentativa do sujeito de se apresentar de forma favorável; a escala *Inconsistency* (ICN), que deteta a falta de atenção ou de concentração; e, por último, a *Infrequency* (INF), através da qual é avaliado um padrão de resposta randomizado ou idiossincrático (Morey & Lowmaster, 2010; Rogers & Bender, 2012; Young, 2014a). Considera-se, ainda, a existência da escala *Malingering Index* (MAL) que consiste na deteção de padrões inconsistentes de psicopatologia; e da escala *Rogers Discriminant Function* (RDF), que se considera ter uma falha ao nível da base conceptual, pois consiste num método estatisticamente derivado, baseado no peso das pontuações de outras escalas e subescalas. As propriedades psicométricas do instrumento relativas à validade e fiabilidade são consideradas adequadas, embora não exista um indicador único relativo aos pontos de corte recomendados, sendo esses variáveis de acordo com a população a ser avaliada (ver Young, 2014a, para mais detalhes).

Tendo estas informações em consideração, é de constatar que, para que haja um diagnóstico bem-sucedido, é essencial a conjugação convergente de resultados de múltiplos Testes de Validade de Sintomas, de preferência com baixas correlações entre eles (Impelen et al., 2014).

5 – Questões éticas associadas ao diagnóstico de simulação.

Tendo em consideração todos os avanços da investigação na área da simulação de sintomas, diversos dilemas éticos têm sido levantados pelos autores no que respeita o uso do conceito de “simulação” (i.e., “*Malingering*”) para categorização. Importa, em primeiro lugar, e como já foi feito atrás, mencionar que uma categorização deste tipo pode ter para alguém, implicações graves em termos de estigmatização futura. Este conceito é, muitas vezes, utilizado como equivalente ao de decepção/fraude (“*deception*”), embora não o possam ser considerados. O primeiro, como supracitado, implica a presença de benefícios secundários e de intenção, ou seja, motivação externa, o que nem sempre pode ser estabelecido através de uma avaliação (Sahoo et al., 2020). Uma das falhas apontadas aos critérios de diagnóstico apresentados na DSM-5 recai precisamente na inexistência de um guia para

Estudos com *design* de simulação análogo em Testes de Validade de Sintomas (TVS):

Uma revisão da literatura

Carolina Duarte de Almeida (carolinadualmeida2@gmail.com) 2020

estabelecer a natureza externa das motivações subjacentes à simulação de sintomas (Merten & Rogers, 2017). É preciso considerar sempre que um determinado resultado pode ser devido a outras causas - como perturbação de transtorno factício, frustração, aborrecimento, cansaço, resistência, instabilidade emocional, pedido de ajuda, ou mesmo pela falta de *insight* da pessoa relativo ao seu estado real (Impelen et al., 2014; Young, 2014b). Também Merckelbach et al. (2019) alertam para a complexidade do fenómeno que é o exagero de sintomas, e que pode surgir devido a variadas razões, não apenas individuais, mas também situacionais. Exemplos disso são a sequência dos itens e dos testes apresentados, a desinformação de sintomas, a desatenção a responder, e certas características de personalidade ou dificuldades pessoais (como altos níveis de afeto negativo, ansiedade ou alexitimia; Merckelbach et al., 2019). Além disso, é de extrema importância que se atente para o facto de que a evidência de apresentação de sintomas imprecisa por parte de um paciente, não exclui a existência de psicopatologia genuína, embora exista a probabilidade da recusa de tratamento quando essa ocorre (Sahoo et al., 2020).

É nesse sentido que Sahoo e colaboradores (2020) alertam para que a categorização de simulação seja realizada apenas em casos em que existem certezas de que disso se trata, mas sugerem que, em substituição deste último, sejam utilizados conceitos mais inócuos, como é o caso de apresentação “imprecisa” (“*inaccurate*”) ou “incerta” (“*unreliable*”) de sintomas. O mesmo foi feito por Boone (2007), através de uma análise dos critérios de Slick et al. (1999; como citados por Boone, 2007) para o diagnóstico de simulação de disfunção neurocognitiva (MND). De acordo com o autor, deve ser atribuída importância à credibilidade da apresentação de sintomas realizada pelo sujeito, que é apenas o que pode ser avaliado pelos TVS disponíveis. As razões pela qual esta ocorreu, ou não, de forma credível, não podem ser determinadas de forma objetiva, de modo que não devem ser foco de atenção. Assim sendo, apresentam como alternativa conceptual as expressões “desempenho/apresentação de sintomas não credível” (“*noncredible symptoms performance*”), ou “enviesamento negativo de resposta” (“*negative response bias*”; Boone, 2007).

A crescente preocupação dos profissionais de saúde mental com a escolha dos termos a utilizar aquando da devolução de informação relativa à validade de sintomas é notável através dos resultados de um estudo de Martin et al. (2015). Estes autores, perceberam que a utilização do termo *simulação* (i.e., “*malingering*”) em relatórios psicológicos se tornou cada vez menos recorrente. Entre os neuropsicólogos inquiridos, cerca de 90% afirmou nunca (ou quase nunca) utilizar este termo. Em alternativa, a expressão mais comumente adotada pelos profissionais parece ser “os resultados dos testes são inconsistentes com a severidade dos danos” (Martin et al., 2015).

Young (2014c) apresenta a sugestão aos profissionais de saúde mental para que, na ausência de evidência definitiva, apresentem, nas suas conclusões, declarações de que a apresentação de sintomas terá sido enviesada, mas sem que se possa determinar a presença ou ausência de simulação propriamente dita.

O tipo e a natureza das investigações realizadas no âmbito do exagero

ou simulação de sintomas têm, também, diversas implicações éticas. Em primeiro lugar, importa considerar os estudos com simuladores instruídos, nos quais se incorre no risco de revelar demasiada informação relativa aos testes utilizados. Nestas investigações, que têm como objetivo a análise do impacto das informações fornecidas aos respondentes antes da avaliação no seu desempenho (i.e., o *coaching*, discutido no ponto seguinte), o desafio será não pôr em casa a segurança dos TVS, de modo a não invalidar os resultados obtidos com a sua aplicação. Neste ponto, convergem dois princípios éticos dos profissionais de saúde mental: o de proteger a integridade dos testes que utilizam; e o de basear as suas conclusões em medidas válidas e confiáveis, devendo os resultados ser sempre interpretados com consideração dos diversos fatores presentes na avaliação, incluindo os fatores situacionais com potencial influência nesta interpretação (Victor & Abeles, 2004). Portanto, as investigações relativas ao *coaching* surgem com grande relevância para a validação e aumento da confiança depositada nos resultados dos instrumentos de avaliação utilizados, que devem, para isso, ser relativamente resistentes às informações fornecidas por terceiros. A dificuldade será a de conseguir estudar de forma eficiente se esta resistência está presente, revelando aos participantes apenas a informação necessária para este objetivo, sem fornecer detalhes específicos que possam comprometer a integridade dos testes utilizados. Adicionalmente, nas publicações destes estudos, surge um dilema semelhante: conseguir apresentar informação suficiente que oriente as decisões dos clínicos e futuras investigações na área, sem, no entanto, revelar demasiada informação relativa às estratégias de *coaching* utilizadas que possam ser utilizadas por terceiros, colocando, mais uma vez, em causa a integridade dos testes (Ben-Porath, 1994; Berry et al., 1994; Springman & Vandenberg, 2009).

Importa, ainda, considerar o chamado “paradoxo das investigações de simulação” (Rogers & Cavanaugh, 1983; como citados em Gillard, 2010) presente nos estudos que utilizam o *design* de simulação análogo como metodologia principal. Este conceito faz referência ao facto de ser requerido aos participantes das condições experimentais de simulação que sigam as instruções de modo a simular sintomas, com o objetivo de estudar os sujeitos que simulam quando lhes é pedido que cooperem de forma honesta. Deste modo, os simuladores em contextos reais encontram-se a fazer o contrário daquilo que lhes é pedido, ao contrário dos participantes experimentais, que seguem instruções. Além de colocar em causa a validade externa das investigações, poderá ser considerado um problema ético por ser pedido aos participantes que façam o contrário daquilo que lhes seria pedido em situações verídicas (Gillard, 2010).

6- O *Coaching* – Simuladores instruídos e não instruídos.

O *coaching* consiste no fornecimento de instruções por uma terceira parte, tendo como objetivo a alteração dos resultados de uma avaliação, favorecendo a pessoa instruída (Gorny & Merten, 2006; Springman & Vandenberg, 2009).

Apesar de o *coaching* poder advir de diversas fontes, um estudo de

Estudos com *design* de simulação análogo em Testes de Validade de Sintomas (TVS):

Uma revisão da literatura

Carolina Duarte de Almeida (carolinadualmeida2@gmail.com) 2020

Wetter e Corrigan (1995, como citados em Springman & Vandenberg, 2009) com diversos advogados e estudantes de direito, concluiu que grande parte dos inquiridos acreditam que o *coaching* faz parte das obrigações dos advogados para com os clientes, e que cerca de 63% está disposto a fornecer informações acerca das escalas de validade de sintomas presentes nas avaliações psicológicas. Foi também demonstrado que esta é uma prática recorrente dos mesmos, e que põe em causa os resultados das avaliações psicológicas realizadas (Lees-Haley, 1997, como citado em Gorny & Merten, 2006). Também um estudo de Essig et al. (2001; como citado em Gorny & Merten, 2006), revelou que os advogados entrevistados passavam, em média, uma hora a preparar os seus clientes para a avaliação neuropsicológica a que seriam sujeitos. Um resultado semelhante a este último foi encontrado num estudo de Victor e Abeles (2004), que concluiu, ainda, que cerca de 44% dos respondentes indicou vontade em saber quais os instrumentos psicológicos que seriam utilizados durante a avaliação dos seus clientes, e que a maioria destes tinham acesso a essa informação quando a requeria aos psicólogos responsáveis pela avaliação. Mais recentemente, Spengler e colaboradores (2020) realizaram um estudo que determinou que cerca de 72.3% dos advogados inquiridos acredita ser seu dever aconselhar os seus clientes sobre o que envolve a avaliação psicológica. Adicionalmente, dos 205 advogados que participaram no estudo, cerca de 52.7% concorda que deve fornecer informações específicas sobre as escalas de validade de sintomas presentes no MMPI-2, das quais os clientes se poderiam valer para evitar a deteção do exagero ou simulação de sintomas psicopatológicos. A comparação destes resultados com os apurados no estudo de Wetter e Corrigan (1995, como citados em Spengler et al., 2020) – 52.7% vs 63% - mostram que as atitudes dos advogados relativamente a este tema poderão estar a mudar, mas com um tamanho de efeito baixo (Spengler et al., 2020).

Por outro lado, também a sujeição repetida às avaliações psicológicas durante um processo judicial pode, por si mesma, constituir uma fonte de *coaching*. Da mesma forma, as informações encontradas na internet acerca dos sintomas mais comuns de perturbações específicas, e sobre como responder aos testes de modo a criar uma exposição convincente de uma perturbação, é uma outra fonte de *coaching* referida na literatura (Suhr & Gunstad, 2007).

Os estudos têm referido a existência de três categorias dentro das quais podem ser inseridos os diferentes tipos de *coaching* (Suhr & Gunstad, 2007): (a) informação sobre os sintomas típicos da perturbação a ser simulada; (b) detalhes sobre os testes que serão administrados e fornecimento de estratégias de resposta aos mesmos, de modo a evitar a deteção das respostas não credíveis; e (c) aviso sobre a presença de métodos de avaliação da validade de sintomas no protocolo de avaliação.

As conclusões dos estudos realizados (e.g. Lees-Haley, 1997; como citado em Gorny & Merten, 2006) enfatizaram a necessidade da criação de estratégias de correção da influência do *coaching*, de modo que os resultados das avaliações psicológicas realizadas possam ser utilizados com confiança. Neste sentido, as investigações realizadas têm como principal objetivo a

deteção do grau de resistência dos TVS's ao *coaching* (Gorny & Merten, 2006), e têm como base diversos tipos de *design*, como a Comparação de Grupos Conhecidos (*Known-Group Comparison* - KGC), ou o *design* de Simulação Análogo, sendo este último o mais comum, e a base dos estudos analisados na presente dissertação. Os participantes mais comumente utilizados são os estudantes universitários, de modo a minimizar o risco de revelar informações a arguidos verdadeiros que os ajudem a evitar a deteção da simulação de sintomas. Com a utilização destes participantes deve ser tido em conta a utilização de um cenário que se aproxime de alguma situação suscetível de se realizar (e.g. Rüsseler et al., 2008; Springman & Vandenberg, 2009, que utilizam cenários com envolvimento em acidentes de viação). Adicionalmente, devem ser efetuadas verificações de manipulação para confirmar a compreensão, a complacência e a recordação das instruções por parte dos participantes, e deve ser utilizado um incentivo para aumentar a sua motivação (Rogers, 1997, como citado em Springman & Vandenberg, 2009). Ao longo dos anos, diversos estudos têm sido conduzidos com o objetivo de perceber o impacto do *coaching* na deteção da simulação pelos TVS's. Os resultados encontrados apresentam diferenças, que podem, em grande parte, ser devidas ao tipo de informação fornecida aos grupos experimentais e aos restantes aspetos metodológicos das investigações (e.g., N e tipo de amostra; grupos instruídos). A Tabela 1 apresenta as informações essenciais de cada uma das investigações analisadas, no que respeita a amostra e o tipo de *coaching* utilizado, e os resultados obtidos relativos ao impacto do mesmo na deteção de simulação.

Tabela 1

Análise dos estudos de avaliação do impacto do coaching na deteção de simulação.

Estudo	Amostra	Método	Impacto
Jelicic et al. (2006)	EU (60).	IS; AP; AE.	Não significativo.
Jelicic et al. (2007)	EU (90).	IS; AR.	Significativo apenas no desempenho do ASTM.
Gorny & Merten (2006)	PC (65).	IS; AE; DA.	Significativo apenas em grupos em que existiu AE e DA.
Rüsseler et al. (2008)	EU (116); RP (7); PI (33).	IS; AE; DS; AP.	Significativo apenas no grupo em que todos os métodos foram aplicados.
Springman & Vandenberg (2009)	EU (101).	DS; AP; AE.	Não significativo.
Merten et al. (2010)	PC (80).	IS; AE; AP.	Significativo no desempenho do SIMS, mas não do MMPI-2; maior impacto com combinação de métodos.

Estudo	Amostra	Método	Impacto
Weiborn et al. (2012)	EU (113); PC (10).	IS; DA; AE; DS.	Mais significativo apenas quando usada a combinação dos métodos.

Nota. EU – Estudantes universitários; PC – Participantes da comunidade; RP – Recém-profissionais; PI – Pacientes em internamento; IS- Informação sobre os sintomas; AP - Aviso sobre presença de TVS's; AE - Aviso para não incorrer em exagero de sintomas relatados; DS – Dicas sobre como simular uma perturbação de forma eficiente; DA - Detalhes sobre os testes de avaliação da validade de sintomas; ASTM – *Amsterdam Short-Term Memory Test*.

No geral, os resultados parecem apontar para a existência de um impacto significativo apenas com formas mais extensivas e elaboradas de *coaching* (Gorny & Merten, 2006; Rüsseler et al., 2008; Weinborn et al., 2012). Por outro lado, estudos de Jelicic e colaboradores (2006, 2007) nos quais foram utilizadas formas mais subtis de *coaching*, ou seja, em que lhes foram fornecidas poucas informações, mostraram alguma resistência do SIMS a este fenómeno, sendo que as diferenças na deteção entre os grupos com e sem *coaching* não se mostraram significativas. Além disso, têm sido vários os estudos (como citados em Springman & Vandenberg, 2009) a apontar para um impacto mais significativo do *coaching* por orientação na resposta aos testes de avaliação psicológica, por contraste à informação dos sintomas de psicopatologias, que se revelou uma forma menos impactante. Também um estudo de Springman e Vandenberg (2009) não encontrou diferenças significativas na deteção da simulação com instrumentos de avaliação da competência para ser julgado que incluem escalas de validade de sintomas (*Georgia Court Competency Test* – GCCT; e *Evaluation of Competency to Stand Trial – Revised* – ECST-R), com o grupo sujeito a uma forma relativamente mais sofisticada de *coaching*, mas não extensiva. Neste sentido, os autores alertam para a possibilidade da existência de resultados diferentes, no caso de os respondentes serem instruídos por advogados ou uma terceira parte com conhecimentos detalhados de psicopatologia (Jelicic et al., 2006; 2007). Por outro lado, os resultados podem variar, também, entre instrumentos, independentemente do tipo de *coaching* que é implementado. Em Jelicic et al. (2007) os resultados obtidos com o SIMS não foram replicados com o *Amsterdam Short Term Memory* (ASTM), o que demonstra falta de resistência deste instrumento ao *coaching*, mesmo com formas menos extensivas do fenómeno. Este foi, também, o instrumento que mostrou maior sensibilidade ao *coaching* no estudo de Gorny and Merten (2006). Em Merten e colaboradores (2010), contrariamente aos estudos referidos anteriormente, o SIMS mostrou-se pouco resistente ao fornecimento de informações e avisos aos simuladores experimentais, principalmente perante métodos combinados. O mesmo não aconteceu com o MMPI-2, que se mostrou relativamente resistente (Merten et al., 2010).

7 - O design de Simulação Análogo.

A avaliação dos estilos de resposta e dos instrumentos de validade de sintomas encontra-se dependente da realização de investigações estruturadas.

São quatro os *designs* usados na grande parte dos estudos de simulação: o *design* de simulação análogo (“*Simulation Design*”, SD); a comparação de grupos conhecidos (“*Known-Groups Comparisons*”, KGC); o *design* de prevalência diferencial (“*differential prevalence design*”); e o *design* de critério parcial (“*partial criterion design*”; Rogers, 2018a).

O primeiro, base de análise da presente dissertação e o mais comum nas investigações existentes, baseia-se na distribuição aleatória de participantes por condições distintas, isto é, condições experimentais de simulação e de controlo. Adicionalmente, poderão ser usadas comparações com grupos clínicos. As falhas apontadas a este *design* encontram-se ao nível da validade externa, devido ao risco de os participantes utilizados não se confrontarem verdadeiramente com as consequências do insucesso. Por outro lado, a sua validade interna, tendo em conta a utilização de instruções estandardizadas, verificações de manipulação, condições operacionalizadas, e incentivos, é, no geral, indicada como forte (Rogers, 2018a).

O *design* de comparação de grupos conhecidos passa pelo estabelecimento de classificações independentes (e.g., pacientes com perturbações genuínas; simuladores), sendo de seguida realizadas comparações entre os grupos estabelecidos. Estas categorizações têm como base as pontuações em medidas validadas sendo que, por forma a aumentar a precisão, são eliminados os participantes cuja classificação é indeterminada, e são considerados apenas os participantes cujas pontuações se encontram significativamente abaixo da performance ao acaso. Contrariamente ao anterior, a validade externa deste *design* é considerada forte, devido às estimações de participantes, contextos e incentivos consistentes com a realidade. Por outro lado, o fraco controlo dos investigadores sobre a distribuição experimental ou sobre o empenho dos participantes no seu desempenho, tornam a sua validade interna fraca. No entanto, esta falha pode ser minimizada através da utilização de procedimentos estandardizados com regras de decisão à priori, fornecendo dados sistemáticos (Rogers, 2018a; Rogers & Bender, 2012).

No *design* de prevalência diferencial é realizada uma assunção de pertença dos indivíduos a um grupo de simuladores suspeitos ou de respondentes genuínos, com base nos incentivos e nas características dos contextos em que estes se inserem (e.g. grupo de simuladores suspeitos presentes em contextos de litígios; e grupo de controlo, considerado genuíno por não estar perante contextos adversos). Esta é uma metodologia fortemente criticada, e estima-se ter uma validade interna fraca, e uma validade externa entre fraca a moderada. Adicionalmente, a sua precisão de classificação não pode ser determinada (Rogers, 2018a).

Por último, o *design* de critério parcial, anteriormente designado por comparação por *bootstrapping* (“*bootstrapping comparison*”; no qual era utilizada apenas uma medida para melhorar a precisão de classificação de outra; Rogers, 1997, como citado em Rogers, 2018a) consiste na utilização de diversas escalas ou indicadores para identificar um grupo com fortes probabilidades de recorrer a um estilo de resposta específico. A utilização de diversas medidas, por contraste ao que ocorria na metodologia de comparação

por *bootstapping* serviu para aumentar o rigor experimental não conseguido com esta estratégia. Ainda assim, o *design* de critério parcial apresenta falhas ao nível da validade interna, considerada fraca, embora mantendo uma validade externa moderadamente forte quando são utilizadas amostras clínicas apropriadas (Rogers, 2018a). Contrariamente ao que acontece no método de comparação dos grupos conhecidos, este utiliza escalas apenas moderadamente validadas que são convenientemente selecionadas, e não faz remoção de participantes cujo desempenho não permitiu realizar determinações de pertença a um grupo com confiança (Rogers, 2018c).

Tendo estas características em consideração, Rogers (1997, como citado em Rogers, 2018c) sugere que uma validação bem-sucedida de medidas implica a combinação de estudos que utilizem o *design* de simulação e o de comparação de grupos conhecidos, de forma que se tire proveito dos pontos fortes de ambos os métodos.

Tal como referido anteriormente, o *design* de simulação análogo - será a base para os estudos apresentados na presente dissertação. A metodologia utilizada na maioria das investigações desta natureza passa pela atribuição aleatória de participantes não clínicos por uma condição de controlo, na qual são instruídos para responder de modo honesto ao protocolo de avaliação, e por uma ou mais condições experimentais, nas quais os participantes recebem indicações para simular. Em adição à comparação realizada entre os desempenhos das condições de controlo e experimentais, pode ser realizada uma comparação com grupos clínicos relevantes (Rogers, 2018c). Embora não seja prática padrão em todos os estudos, existindo ainda várias publicações em que não foram considerados os grupos clínicos, Rogers e Gillard (2011, como citados em Rogers, 2018c), indicam que, sem incluir esta comparação na investigação, os resultados obtidos apresentam relevância limitada. Este *design* permite, ainda, realizar a investigação numa vertente intra-sujeitos (“*within-subjects*”). Isto é, os mesmos sujeitos respondem aos protocolos seguindo as instruções de ambas as condições, permitindo fazer-se comparações de padrões para as respostas que fornecem de forma honesta e em condição de simulação (Rogers, 2018c).

As investigações de simulação desta natureza implicam a utilização de instruções explícitas que, de modo a conseguir uma participação bem-sucedida na condição experimental, deverão provocar envolvimento e motivação interna genuína dos participantes na atividade (Rogers, 2018c). Neste sentido, Rogers e Cruise (1998) e Rogers (2018c) referem a existência diversos cuidados a ter aquando da redação das instruções experimentais a apresentar numa investigação de simulação. Em primeiro lugar, estas devem ser compreensíveis, devendo ter em conta a população a que são dirigidas, para que todos os participantes possam compreender aquilo que lhes é pedido. Devem, também, ser específicas, na medida em que os investigadores devem conseguir fazer uma interpretação clara dos resultados. Adicionalmente, deve ser considerada a adequabilidade do cenário ou do contexto apresentado à população respondente ao estudo, que sendo percebido como relevante por estes, aumenta o seu envolvimento na tarefa. A relevância da investigação percebida pelos sujeitos, para os próprios ou para os seus familiares, é também

um importante fator de envolvimento. A estimulação da motivação interna dos participantes é, também, necessária para o seu envolvimento genuíno na tarefa. A par da motivação, deve existir um incentivo para que o desempenho dos participantes seja bem-sucedido, podendo esse ser negativo ou positivo. A credibilidade é outro dos elementos a ter em conta, isto é, a motivação dos participantes para serem credíveis pode ser tomado como um incentivo com muito poder. É importante que, ao proceder-se à aplicação dos instrumentos de avaliação numa investigação deste *design*, seja incluída uma sugestão para que os simuladores respondam de forma mais credível possível e um aviso de que o protocolo de avaliação contém indicadores de simulação. Isto permite que os resultados obtidos nas investigações se aproximem da realidade, sendo que as investigações que não contemplam estas indicações correm o risco de produzir resultados deturpados devido à falta de monitorização da credibilidade dos simuladores desinformados. Esta última indicação serve também para que os participantes percebam que a sua credibilidade não é um construto abstrato, mas que pelo contrário pode ser medido (Rogers & Cruise, 1998; Rogers, 2018c).

Por último, uma investigação com este *design* implica a realização de uma verificação de manipulação, i.e., uma avaliação do empenho que os participantes depositaram na tarefa. Este é, no entanto, um elemento diversas vezes desconsiderado pelos investigadores. Este tipo de verificação deve envolver: confirmação de recordação e de compreensão das instruções, e avaliação do esforço reportado (Rogers, 2018c).

Rogers e Cruise (1998) concluíram, através de um estudo realizado com estudantes universitários, que as diferenças metodológicas ao nível das instruções dadas aos participantes provocam grandes diferenças nos resultados das investigações. A diferença mais significativa ocorreu ao nível do incentivo fornecido, na medida em que um tipo de incentivo negativo (i.e., em que eram apresentadas consequências do insucesso da prestação) originou uma apresentação mais focada dos participantes (Rogers & Cruise, 1998). Também Merckelbach e colaboradores (2009) realizaram um estudo no qual utilizaram cenários experimentais de diversas naturezas, de modo a avaliarem o impacto das diferenças instrucionais no desempenho dos simuladores. As conclusões apontaram para formas de simulação mais exageradas em cenários de contexto criminal, e menos intensas em casos de litígio civil. Esta última revelou-se, conseqüentemente, mais difícil de detetar. Os autores apresentaram uma hipótese explicativa, que aponta para uma crença nos participantes de que apenas apresentações sintomatológicas mais dramáticas poderão resultar na desresponsabilização criminal de alguém (Merckelbach et al., 2009). Também um estudo de Gillard (2010) encontrou diferenças significativas nas pontuações de diversas escalas do SIMS quando manipulada a familiaridade dos participantes com o cenário que lhes era fornecido. De acordo com os resultados, a uma maior familiaridade com o contexto do cenário corresponde uma maior motivação e envolvimento na tarefa (Gillard, 2010).

Este é um método que tem sido alvo de críticas ao nível da validade externa, pois as diferenças motivacionais entre os simuladores experimentais

e os simuladores reais tornam incerta a generalização dos resultados obtidos (Rogers & Cavanaugh, 1983; como citados em Rogers & Cruise, 1998). Na verdade, alguns estudos têm demonstrado que quando é requerido que simulem uma perturbação específica, os simuladores experimentais tendem a generalizar e a reportar sintomas variados (e.g. Merten et al., 2010; Giger et al., 2010). No entanto, outros estudos (e.g. Brennan & Gouvier, 2006; Clegg et al., 2009) têm encontrado resultados que indicam desempenhos semelhantes entre simuladores experimentais e simuladores reais, concluindo que os primeiros podem ser usados com confiança para representar os últimos.

8 - Estudos com *design* de simulação análogo em contextos forenses.

Ao longo dos anos, as investigações com este *design* têm surgido em diferentes contextos, com o objetivo de perceber a eficácia dos TVS na detecção dos estilos de resposta enviesados, e mais especificamente da simulação, utilizados pelos sujeitos. Nos contextos forenses, devido à elevada prevalência deste fenómeno e aos custos associados à simulação de sintomas bem-sucedida, a realização destes estudos ganha especial importância. No entanto, o acesso dos investigadores às populações em contextos forenses tem alguns constrangimentos, o que limita o número de estudos realizados. De forma a contornar esses obstáculos, grande parte dos estudos realizados com este *design* recorre a populações de acesso facilitado, ou seja, da comunidade, e utiliza, na grande maioria das vezes, cenários de contextos forenses. Deste modo, será apresentada, de seguida, uma análise das principais investigações realizadas, em contexto criminal e em contexto médico ou médico-legal, embora em número limitado. Para isso, serão considerando as principais características metodológicas e instrucionais das mesmas. O objetivo será perceber até que ponto os elementos referidos por Rogers e Cruise (1998) e por Rogers (2018c), como essenciais para a formulação de instruções experimentais bem conseguida, estão presentes. Importa assinalar a data de alguns dos artigos analisados, por anteceder a publicação destas diretrizes.

8.1 - Estudos em contexto criminal.

Após extensiva pesquisa, é possível concluir, tal como referido, que a disponibilidade de estudos em contexto criminal com este tipo de *design* é limitada. Deste modo, a análise será incidente em apenas sete artigos - Rogers et al. (1996, 2004); Edens et al. (2007); Steffan e Morgan (2008); Wall et al. (2015); Marchi e Balboni (2018); e Helvoort et al. (2019). As tabelas 2.1 e 2.2 (Anexo A) contêm uma síntese das informações apresentadas de seguida.

Em primeiro lugar, relativamente aos instrumentos de avaliação da validade de sintomas utilizados nestas investigações, com maior número de utilizações encontram-se o SIMS, que surge nos estudos de Rogers et al. (1996), Edens et al. (2007) e Marchi e Balboni (2018), e o MMPI-2, utilizado em Steffan e Morgan (2008) e Marchi e Balboni (2018). A sua versão reestruturada (MMPI-2-RF) surge em Wall et al. (2015), e o MMPI-A, a versão específica para adolescentes foi validada por Rogers et al. (1996). Este último utilizou, ainda, o SIRS, que foi, também, escolhido por Rogers et al. (2004), juntamente com o *Miller Forensic Assessment of Symptoms Test* (M-FAST),

Estudos com *design* de simulação análogo em Testes de Validade de Sintomas (TVS):

Uma revisão da literatura

Carolina Duarte de Almeida (carolinad Almeida2@gmail.com) 2020

quando procedeu à avaliação do *Evaluation of Competency to Stand Trial – Revised* (ECST-R), um instrumento para avaliar a competência para ser julgado. Adicionalmente, as escalas do PAI surgiram como método de avaliação e validação em estudos como o de Edens e colaboradores (2007), que contou ainda com o SIRS, e em Marchi e Balboni (2018). Por último, os estudos de validação do SRSI neste contexto foram iniciados por Helvoort e colaboradores (2019).

Embora todos os estudos referidos tivessem utilizado o *design* de simulação análogo (SD) como estratégia de investigação, Edens et al. (2007) e Marchi e Balboni (2018) recorreram, também, ao método de comparação de grupos conhecidos (KGC), enquanto Rogers et al. (1996, 2004) e Helvoort et al. (2019) decidiram maximizar o uso de participantes disponíveis e avaliar a eficácia das instruções disponibilizadas, através do *design* intra-sujeitos.

De um modo geral, os investigadores decidiram recorrer, tanto a reclusos integrados na zona comum, como a reclusos da área psiquiátrica, fosse em internamento (Steffan & Morgan, 2008), ou apenas em regime de tratamento (Edens et al., 2007; Wall et al., 2015; Marchi & Balboni, 2018). No estudo de Rogers et al. (1996) e Helvoort et al. (2019) os participantes recrutados foram, exclusivamente, reclusos em internamento psiquiátrico. No caso do primeiro, seriam adolescentes internados numa unidade de tratamento psiquiátrico após serem referenciados pelo tribunal. A segunda investigação contou com a colaboração de pacientes internados num hospital psiquiátrico prisional de máxima segurança. Em Rogers et al. (2004), o grupo de pacientes psiquiátricos, utilizado para fazer comparação clínica, foi excepcional por não serem reclusos. Considerando a validade limitada dos resultados de uma investigação sem realização das comparações com grupos clínicos, pode concluir-se, através de uma análise dos estudos mencionados que, essa comparação foi realizada, então, em 4 (Rogers et al., 2004; Edens et al., 2007; Steffan & Morgan, 2008; Wall et al., 2015) dos estudos analisados. As exceções surgiram nos casos em que a amostra utilizada foi exclusivamente constituída por reclusos em internamento (Rogers et al., 1996 e Helvoort et al., 2019), e no estudo de Marchi e Balboni (2018), no qual a amostra clínica foi utilizada, apenas, para a comparação de grupos conhecidos.

Relativamente às características das instruções fornecidas, são visíveis grandes variações entre os estudos analisados. Apenas duas das investigações podem ser destacadas, tendo em conta o cumprimento dos requisitos postulados por Rogers e Cruise (1998) e Rogers (2018c). Em Rogers et al. (2004) não podem ser apontadas falhas perceptíveis, de acordo com o que é apresentado no texto, enquanto a investigação de Helvoort e colaboradores (2019) desrespeitou a falta de adequação do contexto do cenário apresentado à amostra. Os autores utilizaram, ao invés, um cenário em situação de incapacidade de trabalho simulada, embora com descrição e instruções completas. Por outro lado, Steffan e Morgan (2008) desrespeitaram a disponibilização de um incentivo à participação, bem como a realização de questionário de verificação da manipulação para avaliação da compreensão e empenho dos participantes. Na verdade, é possível verificar que este apenas surge nos estudos de Rogers et al. (2004) e de Helvoort et al. (2019). Também

as instruções de simulação apresentadas ficam muito aquém daquilo que seria ideal para uma investigação com *design* de simulação análogo, com inexistência de um cenário adequado em três das investigações (Edens et al., 2007; Marchi & Balboni, 2018; Rogers et al., 1996). Nestes apenas era requerido que simulassem determinada psicopatologia. Adicionalmente, Rogers et al. (1996) incluíram uma descrição das características da perturbação em causa. Relativamente ao cenário disponibilizado por Wall e colaboradores (2015), a informação foi muito vaga, embora com apresentação de uma motivação para a simulação ser bem-sucedida, e adequada ao contexto em questão. Por último, Rogers et al. (2004) e Steffan e Morgan (2008) conseguiram apresentar um cenário relevante para o contexto, com instruções para simular e sugestão para evitar a deteção. O último elemento analisado foi o incentivo disponibilizado, que faltou nas investigações de Steffan e Morgan (2008), Wall et al. (2015), e de Marchi e Balboni (2018). Nas restantes investigações o incentivo foi de natureza monetária e de valor simbólico. Em adição, nas investigações de Edens et al. (2007) e de Rogers et al. (1996, 2004) foi indicada a existência de um bónus extra no caso de a simulação ser bem-sucedida. O aviso da presença de testes de validade de sintomas no protocolo é, também, uma sugestão de Rogers (2018c), mas que não foi considerada nesta análise devido ao facto de essa informação não estar contida em todos os artigos.

8.2 - Estudos em contexto médico e médico-legal.

Decorrentes em contexto médico-legal, serão analisadas apenas sete investigações (Dearth et al., 2005; Hopwood et al., 2010; McGuire et al., 2001; Puente-López et al., 2020; Rogers et al., 1995; Sellbom et al., 2011; Wygant et al., 2009), cujas informações se encontram nas tabelas 3.1 e 3.2 (Anexo B).

As amostras utilizadas nestas investigações são representativas de populações presentes em contextos de avaliação clínica ou médico-legal, embora a participação nos estudos dependesse de cumprirem o requisito de não estarem perante situações de litígio, de modo a não existirem motivações externas para exagerar sintomas, que não as apresentadas em investigação. Rogers et al. (1995) recorreram a pacientes de instituições de serviços de saúde mental, com diagnóstico de diversas perturbações mentais. Por sua vez, McGuire e colaboradores (2001) e Hopwood et al. (2010) utilizaram amostras de pacientes com queixas de dor crónica. Na primeira investigação, os pacientes foram instruídos a responder de modo honesto ou a exagerar as suas queixas, e como grupo de comparação foi requisitado um grupo clínico sem queixas de dor, mas instruído a simular esse tipo de sintomas. Por outro lado, Hopwood e colaboradores (2010) utilizaram um grupo de estudantes universitários como grupo de controlo. Dearth et al. (2005) recorreram a comparações de desempenhos entre pacientes que sofreram de ferimentos na cabeça e de participantes de controlo da comunidade. Wygant et al. (2009) realizaram três estudos, recorrendo à utilização de diferentes amostras. No entanto, para este âmbito, não será considerado o último, que utilizou a comparação de grupos conhecidos como método de investigação. No estudo 1, utilizaram, também, participantes que sofreram de ferimentos na zona da

cabeça, e no segundo a amostra, proveniente de um centro médico de veteranos, sofria de incapacidades físicas decorrentes de atividades ao serviço. Em Sellbom et al. (2011), foram realizadas comparações entre um grupo de estudantes universitários, instruído a simular, um grupo de veteranos com incapacidades físicas, e um grupo de pacientes com diagnósticos variados de perturbações somatoforme. Por último, a investigação de Puente-López e colaboradores (2020) contou com o recrutamento de amostras de duas naturezas: em primeiro, estudantes universitários que, divididos em dois grupos, foram instruídos a responder de forma honesta ou a simular sintomas; e, em segundo, pacientes de uma clínica, que haviam sofrido um acidente de viação. Este último grupo foi, ainda, dividido em dois, consoante fossem considerados, através de uma avaliação, pacientes com sintomatologia genuína, ou pacientes que se encontravam a exagerar os sintomas que reportavam. Tendo em conta a utilização de amostras clínicas com sintomatologia genuína em todos estes estudos, não será considerada a realização de comparação com grupo clínico como critério de avaliação dos mesmos, pois acabou por ser sempre realizada.

Relativamente ao *design* utilizado, apenas Rogers et al. (1995) e Wygant et al. (2009), no segundo estudo, recorreram à utilização dos mesmos participantes em ambas as condições, beneficiando, assim, de um *design* de simulação análogo intra-sujeitos.

Os instrumentos utilizados para a avaliação da validade de sintomas não foram muito variados entre os diversos estudos. Rogers et al. (1995) e Dearth et al. (2005) recorreram ao MMPI-2. Por sua vez, a forma reestruturada deste instrumento (MMPI-2-RF) foi escolhida por Wygant et al. (2009) e por Sellbom et al. (2011). McGuire et al. (2001) utilizaram um instrumento específico para avaliação da dor com uma escala de validade de sintomas integrada, enquanto em Hopwood et al. (2010) foram utilizadas as escalas de validade de sintomas do PAI. Em Puente-López et al. (2020) foi avaliada a capacidade de diagnóstico de simulação do SIMS.

Através de uma análise dos cenários disponibilizados nas investigações, é possível destacar o facto de todos terem utilizado contextos relevantes para o tipo de amostra em questão, sendo pedido para simular ou exagerar sintomas de acordo com o tipo de perturbação de que cada amostra sofria. Destaca-se, também, a presença de aviso para a credibilidade da apresentação de sintomas em todas as investigações, com exceção de Hopwood et al. (2010). Estes autores optaram por não disponibilizar as instruções utilizadas, mas referiram a utilização do aviso de estarem presentes TVS no protocolo, com um dos grupos de simuladores instruídos. Este aviso surgiu, também, em Rogers et al. (1995). Por último, no estudo 1 de Wygant et al. (2009), os participantes foram, também, presenteados com dicas para que a simulação fosse realizada de modo mais eficiente. É de destacar a investigação de Puente-López et al. (2020) pela presença de todos os avisos referidos.

A disponibilização de incentivo para a participação esteve ausente apenas em McGuire et al. (2001) e nos grupos clínicos usados em Hopwood et al. (2010), Sellbom et al. (2011) e Puente-López et al. (2020). Os grupos de estudantes universitários que participaram nestas investigações beneficiaram

de créditos académicos ou pontos extra na avaliação final. Nas restantes investigações, a motivação era estimulada com compensações monetárias pela participação. Adicionalmente, estes investigadores prometiam bónus no caso de as instruções serem seguidas corretamente.

Por último, é de referir a ausência de questionário de verificação da manipulação em parte das investigações analisadas. Apenas em McGuire et al. (2001), Sellbom et al. (2011), no estudo 2 de Wygant et al. (2009) e em Puente-López et al. (2020) foi, então, avaliado o nível de cooperação dos participantes e a sua compreensão das instruções fornecidas.

Em conclusão, é de destacar a metodologia utilizada por Sellbom et al. (2011), tendo ficado apenas ausente o incentivo nos grupos clínicos e o aviso da presença de TVS no protocolo, de acordo com os padrões de avaliação seguidos na presente dissertação. A ausência aviso foi, também, uma das falhas apontadas em Wygant et al. (2009), com adição da falta do questionário de verificação de manipulação na segunda amostra do estudo. Adicionalmente, merece realce a metodologia adotada por Puente-López e colaboradores (2020) que, ao longo do artigo de investigação, informam mesmo ter usado o trabalho de Rogers e Cruise (1998) e Rogers (2018c) como guia. Apesar disso, pode ser apontada a falha de distribuição de incentivo pelas amostras clínicas utilizadas.

9 - Estudos com *design* de simulação análogo em contexto comunitário.

Os estudos em contexto comunitário com o *design* de simulação análogo representam a grande maioria dos estudos disponíveis realizados com esta metodologia. Na análise apresentada nesta secção são incluídos os seguintes estudos: Smith e Burger (1997); Edens et al. (1999); Merckelbach e Smith (2003); Schoenberg et al. (2003); Jelicic et al. (2006, 2007); Springman & Vandenberg (2009); King e Sullivan (2009); Sullivan e King (2010); Marion et al. (2011, 2013); Sullivan e Elliot (2012); Christiansen e Vincent (2012); Dhilon et al. (2016); Impelen et al. (2017); Geurten et al. (2018); Boskovic et al. (2018); Boskovic, Merckelbach et al. (2019); Giger e Merten (2019); Domingues (2019). Marion et al. (2013) e Impelen et al. (2017) realizaram dois estudos, publicados ambos no mesmo artigo. Para o primeiro, apenas será considerado o Estudo 1, pois o segundo foi realizado com a metodologia de comparação de grupos conhecidos. Para Impelen et al. (2017), será feita a distinção quando assim for necessário, pois os Estudos 1 e 2 partilham muitos elementos metodológicos. Uma síntese das informações retiradas das investigações analisadas pode ser encontrada nas Tabelas 4.1 e 4.2 (Anexo C).

À semelhança do que foi referido na análise das investigações com o *coaching* (sendo que algumas destas investigações são apresentadas em ambas as secções), a população requisitada em maior grau para amostragem dos estudos com este *design* são os estudantes universitários. Nos 20 estudos que comportaram a análise realizada, esta população correspondeu à maioria amostral em 18. Devido à facilidade de acesso aos estudantes, e na utilização de incentivos (como o caso de créditos nas unidades curriculares) que

Estudos com *design* de simulação análogo em Testes de Validade de Sintomas (TVS):

Uma revisão da literatura

Carolina Duarte de Almeida (carolinadualmeida2@gmail.com) 2020

umentam a motivação da participação desta população na resposta aos testes, esta é uma amostra muito requerida. Os restantes estudos (Geurten et al., 2018; Giger & Merten, 2019) solicitaram participações na comunidade geral. Por outro lado, o Estudo 2 de Impelen et al. (2017) requereu à participação de ambas as populações referidas acima. Adicionalmente, diversos tipos de população psiquiátrica fizeram parte de seis estudos (Geurten et al., 2018; Impelen et al., 2017; Marion et al., 2011; Merckelbach & Smith, 2003; Schoenberg et al., 2003), surgindo estas amostras com o objetivo de realizar comparação com grupo clínico. Por último, Marion et al. (2011) requisitou a participação de estudantes de anos avançados e de profissionais de psicologia, de forma a conseguir uma amostra com conhecimentos específicos de saúde mental, que serviriam de simuladores sofisticados.

Relativamente aos TVS utilizados nas investigações, o SIMS continua a ser muito escolhido para grande parte dos protocolos. Neste conjunto de estudos, o instrumento esteve presente em 8 (Domingues, 2019; Edens et al., 1999; Giger & Merten, 2019; Impelen et al., 2017; Jelicic et al., 2006, 2007; Merckelbach & Smith, 2003; Smith & Burger, 1997). As escalas de validade de sintomas do MMPI-2 foram utilizadas por Sullivan e Elliot (2012), enquanto Marion et al. (2011, 2013) e Dhillon et al. (2016) recorreram a este instrumento na sua versão reestruturada (MMPI-2-RF). King e Sullivan (2009) e Sullivan e King (2010) utilizaram as escalas de validade de sintomas presentes no PAI, enquanto Schoenberg et al. (2003) procuraram estudar a utilidade das escalas do *Millon Clinical Multiaxial Inventory – Third Edition* (MCMI-III) para avaliação dos estilos de resposta. Por sua vez, Springman e Vandenberg utilizaram dois instrumentos desenhados para avaliar a capacidade de ser julgado (ECST-R – *Evaluation of Competency to Stand Trial – Revised*; GCCT – *Georgia Court Competence Test*). Também o *Miller Forensic Assessment of Symptoms Test* (M-FAST) foi utilizado em Christiansen e Vincent (2012) e em Impelen et al. (2017; Estudo 1). Em Impelen et al. (2017) esteve presente, também, um instrumento recente de validade de sintomas, que tem como objetivo a distinção entre a natureza interna e externa dos incentivos para simular (*Symptom and Dispositive Interview*; SDI). Por último, os estudos mais recentes analisados (Boskovic et al., 2018; Boskovic, Merckelbach et al., 2019; Domingues, 2019; Geurten et al., 2018; Giger & Merten, 2019) utilizaram o SRSI como integrante do protocolo de avaliação. Importa referir que, nesta análise foram apenas apresentados os TVS (ou os instrumentos cujo principal objetivo não é fazer avaliação da validade de sintomas, mas com escalas para esse fim) utilizados nos estudos, embora em alguns tenham sido administrados, adicionalmente, escalas de avaliação clínica, ou mesmo TVD.

Dos estudos analisados, apenas Edens et al. (1999), Dhillon et al. (2016), Boskovic et al. (2018), Giger e Merten (2019) e Domingues (2019) recorreram a um *design* intra-sujeitos, de modo a tirar maior proveito da amostra e a fazer comparações de desempenhos dos mesmos sujeitos sob as diferentes instruções disponibilizadas. Além disso, tal como referido previamente, a comparação com grupo de relevância clínica tem um papel muito importante neste tipo de investigações. Nesta análise, este elemento surgiu em apenas

seis, listadas acima por terem recorrido a populações psiquiátricas.

Da mesma forma, o incentivo disponibilizado, que surge de natureza monetária ou em forma de créditos acadêmicos, tem um enorme peso na motivação que os sujeitos aplicam na sua participação no estudo. Este elemento serve, também, para conseguir aumentar a validade externa das investigações com este *design*, apesar da comparabilidade limitada com as motivações presentes nos contextos das avaliações psicológicas da vida real. Apesar disso, o incentivo esteve em falta nas investigações de Merckelback e Smith (2003), Marion et al. (2013), Dhilon et al. (2016), Guerten et al. (2018) e Domingues (2019).

A realização do questionário de verificação de manipulação, para avaliação da compreensão das instruções e do esforço disponibilizado pelo participante, é, também, alvo de análise. A ausência desta prática fez-se notar nas investigações de Smith e Burger (1997), Merckelbach e Smith (2003), Jelicic et al. (2006, 2007), Sullivan e King (2010), Marion et al. (2011), Dhilon et al. (2016), Geurten et al. (2018), e Boskovic, Merckelbach, et al. (2019).

Numa análise mais específica das características das instruções disponibilizadas aos grupos de simuladores, é possível destacar a ausência de cenário em quatro das investigações (King & Sullivan, 2009; Marion et al., 2011, 2013; Sullivan & King, 2010). Nestas, as instruções fornecidas incluíram apenas um pedido para simularem sintomas psicopatológicos. Em adição, King e Sullivan (2009) e Sullivan e King (2010) disponibilizaram uma lista dos sintomas que poderiam ser utilizados pelos simuladores, e avisaram um dos grupos da presença de TVS no protocolo. Por outro lado, na investigação de Marion et al. (2011) os simuladores receberam avisos para serem credíveis na sua apresentação, e da presença de TVS no protocolo. Esta investigação teve, ainda, a particularidade de não contar com um grupo de controlo da população normal, sendo que as instruções para responder honestamente aos instrumentos serviu apenas para o grupo clínico. Esta prática foi utilizada, também, no Estudo 2 de Impelen et al. (2017).

A relevância e adequação do cenário apresentado à população utilizada é, também, um ponto importante. É de destacar o esforço da maioria das investigações analisadas nesse sentido. A irrelevância dos cenários foi registada em estudos em que os pedidos eram que se colocassem no papel de arguidos à procura de desresponsabilização criminal por homicídio (Jelicic et al., 2006), de doentes à procura de internamento hospitalar psiquiátrico (Schoenberg et al., 2003), ou de polícias que assistiam a um evento traumático (cenário 1 de Boskovic, Merckelbach, et al., 2019 e de Domingues, 2019). Devido ao distanciamento existente, salvo exceções pontuais, entre os papéis desempenhados pelas amostras na sua vida diária, e aqueles que eram pedidos que adotassem, a tarefa de simulação de sintomas poderia ser comprometida nestas investigações. Por outro lado, situações em que se deveriam imaginar vítimas em acidentes de viação e trabalho, atropelamentos, assaltos, e conseqüentemente, em contexto de avaliação psicológica para obtenção de compensação, foram os cenários mais escolhidos para as investigações em análise (ver Tabela 4.2, para mais detalhes). Também o envolvimento em situações de agressão (Impelen et al., 2017 [Estudo 2]; Smith & Burger,

1997), ou de sujeição a avaliação psicológica por parte do staff da universidade (Boskovic et al., 2018; Dhilon et al., 2016; Impelen et al., 2017 [Estudo 1]) fizeram parte da lista de cenários analisados. Estes últimos correspondem, então, aos únicos que utilizam contextos não forenses nas situações apresentadas nas instruções. Tendo em conta as especificidades da investigação, Springman e Vandenberg (2009) utilizaram a incompetência para ser julgado, após acidente de viação fatal, como cenário. No estudo de Merckelbach e Smith (2003), o cenário foi indicado pelos autores como sendo adequado e relacionado com avaliação forense, mas não foi apresentado.

Por último, importa considerar o aviso para a apresentação de sintomas ser credível, que foi mencionado como presente nas instruções fornecidas em metade dos estudos apresentados em análise (Boskovic, Merckelbach, et al., 2019; Domingues, 2019; Dhilon et al., 2016; Edens et al., 1999; Giger & Merten, 2019; Jelicic et al., 2006, 2007; Marion et al., 2011; Smith & Burger, 1997; Springman & Vandenberg, 2009). Em investigações como Edens et al. (1999), Jelicic et al. (2006), Springman e Vandenberg (2009), King e Sullivan (2009), e Sullivan e King (2010) o aviso da presença de TVS no protocolo foi, também, um elemento encontrado nas instruções. Adicionalmente, Schoenberg et al. (2003), Jelicic et al. (2006, 2007), King e Sullivan (2009), e Sullivan e King (2010) forneceram aos participantes das suas investigações uma lista com os sintomas típicos das simulações que pretendiam que fossem simuladas. Importa referir que, de acordo com os objetivos específicos de cada uma das investigações, a presença destes elementos de aviso ou informativos era controlada e manipulada entre as diferentes condições, podendo não estar presente para todos os grupos de simuladores.

Em conclusão, podemos considerar a existência de algumas investigações que se aproximaram daquilo que seria indicado como ideal neste tipo de *design*, embora nenhuma tenha conseguido cumprir todos os requisitos propostos. Neste sentido, podem ser referidos como exemplo, os estudos de Impelen et al. (2017), que apenas falharam na disponibilização de avisos de credibilidade, e de avisos da inclusão de testes de deteção de exagero de sintomas. O erro de Schoenberg et al. (2003) foi ao nível da adequação do cenário à população utilizada, que por serem estudantes universitários, mais dificilmente se colocariam em situação de requerimento de internamento psiquiátrico. Os avisos referidos anteriormente também se notaram ausentes nesta investigação (ver Tabelas 4.1 e 4.2 para mais detalhes).

Conclusões

Os estilos de resposta enviesados têm-se revelado alvo de atenção por parte dos investigadores, principalmente em contextos forenses, devido às elevadas implicações que podem ter, quando adotados pelos sujeitos que se apresentam para avaliação tendo algo a ganhar, ou pelo contrário, a perder. Embora os critérios apresentados pelo DSM para o termo “simulação de doença/sintomas” sejam fortemente criticados, e careçam de atualização congruente com os avanços da investigação, este fenómeno não pode ser ignorado pelos investigadores, nem pelos clínicos, que baseiam as suas decisões nos relatos dos pacientes. O que se verifica, e que pode ser

comprovado através das estimativas apresentadas, é que esses relatos nem sempre refletem a realidade, podendo ter implicações graves nas conclusões dos profissionais. Foi nesse sentido que surgiu, junto dos clínicos e investigadores, a necessidade de desenvolvimento dos Testes de Validade de Sintomas (TVS), que permitem perceber se a qualidade dos autorrelatos fornecidos não se apresenta comprometida. No entanto, também estes instrumentos, cuja base de construção são diversas estratégias de deteção, necessitam de validação empírica para poderem ser utilizados com confiança. Através da análise dos diversos estudos apresentados com os mais relevantes TVS, é possível notar a existência de limitações em quase todos eles, quando utilizados em certos tipos de amostra ou contexto. Neste sentido, as recomendações passam sempre por: em primeiro lugar, e para os investigadores - continuar os trabalhos de validação dos vários TVS, de modo a permitir uma utilização mais confiante, com conhecimento das suas potencialidades e limitações nos vários contextos e populações, pelos clínicos; e, em segundo lugar, e para os clínicos – recorrer à utilização de mais que um TVS num mesmo paciente, e com análise prévia das investigações realizadas com cada um, com ponderação das suas potencialidades e limites, para que a escolha seja a mais adequada e proveitosa possível considerando o sujeito avaliado.

Adicionalmente, através de uma análise das informações apresentadas ao longo da presente dissertação, é possível concluir que a área da validade de sintomas e, mais especificamente, da simulação e exagero de sintomas, carece ainda de unificação e concordância entre as mais diversas investigações realizadas até ao momento. Desde as estimativas relativas à prevalência deste fenómeno, obtidas nos vários contextos, até à resistência dos testes de validade de sintomas a este comportamento, a falta de consenso nas conclusões formuladas pelos investigadores pode ser problemática. Neste sentido, e embora seja previsível que nunca se atinja total consenso numa área tão complexa, surge a importância extrema de se continuar a investir em investigação, por forma a poder orientar, o mais possível, as decisões dos clínicos, que podem comprometer as vidas dos sujeitos no caso de realizarem a categorização errada do sujeito como sendo simulador ou honesto.

Por último, e seguindo o objetivo principal do presente trabalho, é de destacar as falhas metodológicas de uma grande parte dos estudos realizados com o *design* de simulação análogo. Embora as limitações presentes nos diversos contextos de investigação utilizados possam não ter permitido seguir todas as recomendações de Rogers e Cruise (1998) e de Rogers (2018c), a verdade é que a confiabilidade das conclusões formuladas nos estudos está sempre dependente da metodologia adotada para as atingir. A validade externa fraca é, já, uma das limitações apontadas aos estudos realizados com este *design*, e o contorno da utilização das recomendações que poderiam levar à atenuação dessa mesma limitação, não deveria ser prática padrão nas investigações. O peso maior coloca-se na qualidade das instruções fornecidas, que comprometem a qualidade dos comportamentos de simulação levados a cabo pelos participantes. Nesse ponto inclui-se, principalmente, a adequação dos cenários às populações que, positivamente, conseguiu ser um dos

princípios mais cumpridos pelos investigadores. Pelo contrário, a inexistência de cenário ou a inadequação do mesmo ao contexto, que tornam mais difícil a colocação dos participantes no papel do outro, são práticas a evitar em futuras investigações. Adicionalmente, é de relevar a importância da realização de mais estudos que recorram a contextos forenses, tanto criminais como civis ou médico-legais, pela diferença notória entre o número de investigações nestes e em contextos comunitários, com utilização de participantes da comunidade. Nestas últimas, deve ser destacado que a utilização de grupo de comparação clínica é, também, um padrão a ser seguido, sob pena de se estar a trabalhar com resultados sem validade empírica.

Bibliografia

- Akca, Y. E. A., Slootmaekers, L., & Boskovic, I. (2020). Verifiability and symptom endorsement in genuine, exaggerated and malingered pain. *Psychological Injury and Law*. <https://doi.org/10.1007/s12207-020-09375-w>.
- APA (2014). *DSM 5. Manual de Diagnóstico e Estatístico das Perturbações Mentais, 5ª Edição*. Climepsi Editores.
- Ardic, F. C., Kose, S., Solmaz, M., Kulacaoglu, F., & Balcioglu, Y. H. (2019). Reliability, validity, and factorial structure of the Turkish version of the Structured Inventory of Malingered Symptomatology (Turkish SIMS). *Psychiatry and Clinical Psychopharmacology*, 29(2), 182-188. <http://doi.org/10.1080/24750573.2019.1599237>.
- Bass, C. & Halligan, P. (2014). Factitious disorders and malingering: Challenges for clinical assessment and management. *Lancet*, 383(9926), 1422-1432. [http://dx.doi.org/10.1016/S0140-6736\(13\)62186-8](http://dx.doi.org/10.1016/S0140-6736(13)62186-8).
- Bender, S. D. & Frederick, R. (2018). Neuropsychological Models of Feigned Cognitive Deficits. In R. Rogers & S. D. Bender (Eds.), *Clinical Assessment of Malingering and Deception* (4th ed, pp. 42-60). The Guilford Press.
- Bender, S. D. (2018). Malingered Traumatic Brain Injury. In R. Rogers & S. D. Bender (Eds.), *Clinical Assessment of Malingering and Deception* (4th ed, pp. 122-150). The Guilford Press.
- Ben-Porath, Y. S. (1994). The ethical dilemma of coached malingering research. *Psychological Assessment*, 6(1), 14–15. <https://doi.org/10.1037/1040-3590.6.1.14>.
- Berry, D. T. R. & Nelson, N. W. (2010). DSM-5 and Malingering: A modest proposal. *Psychological Injury and Law*, 3(4), 295-303. <http://dx.doi.org/10.1007/s12207-010-9087-7>.
- Berry, D. T. R., Lamb, D. G., Wetter, M. W., Baer, R. A., & Widiger, T. A. (1994). Ethical considerations in research on coached malingering. *Psychological Assessment*, 6(1), 16–17. <https://doi.org/10.1037/1040-3590.6.1.16>.
- Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) Criteria for Malingered Neurocognitive Dysfunction. In K. B. Boone (Ed.),

- Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective* (pp. 29-49). The Guilford Press.
- 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. <http://dx.doi.org/10.1076/jcen.24.5.561.1004>.
- Boskovic, I., Biermans, A. J., Merten, T., Jelacic, M., Hope, L., & Merckelbach, H. (2018). The Modified Stroop Task is susceptible to feigning: Stroop performance and symptom over-endorsement in feigned test anxiety. *Frontiers in Psychology*, *9*, 1195. <http://doi.org/10.3389/fpsyg.2018.01195>.
- Boskovic, I., Hope, L., Ost, J., Orthey, R., & Merckelbach, H. (2019). Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task. *Journal of Behavior Therapy and Experimental Psychiatry*, *65*, 101483. <https://doi.org/10.1016/j.jbtep.2019.101483>.
- Boskovic, I., Merckelbach, H., Merten, T., Hope, L., & Jelacic, M. (2019). The Self-Report Symptom Inventory as an instrument for detecting over reporting: An exploratory study with instructed simulators. *European Journal of Psychological Assessment*. <https://doi.org/10.1027/1015-5759/a000547>.
- Brennan, A. M. & Gouvier, W. D. (2006) Are we honestly studying malingering? A profile comparison of simulated and suspected malingerers. *Applied Neuropsychology*, *13*(1), 1-11. https://doi.org/10.1207/s15324826an1301_1
- Bush, S. S., Heibronner, R. L., & Ruff, R. M. (2014). Psychological Assessment of Symptom and Performance Validity, Response Bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law*, *7*(3), 197-205. <https://doi.org/10.1007/s12207-014-9198-7>.
- Cernovsky, Z. Z., Ferrari, J. J. R., & Mendonça, J. D. (2019). Pseudodiagnoses of malingering of neuropsychological symptoms in survivors of car accidents by the Structured Inventory of Malingered Symptomatology. *Archives of Psychiatry and Behavioral Sciences*, *2*(1), 55-65.
- Christiansen, A. K. & Vincent, J. P. (2012). Assessment of litigation context, suggestion, and malingering measures among simulated personal injury litigants. *Journal of Forensic Psychology Practice*, *12*(3), 238-258. <https://doi.org/10.1080/15228932.2012.674470>.
- Clegg, C., Fremouw, W., & Mogge, N. (2009). Utility of the Structured Inventory of Malingered Symptomatology (SIMS) and the Assessment of Depression Inventory (ADI) in screening for malingering among outpatients seeking to claim disability. *The Journal of Forensic Psychiatry & Psychology*, *20*(2), 293-254. <https://doi.org/10.1080/14789940802267760>.

- Conroy, M. A. & Kwartner, P. P. (2006). Malingering. *Applied Psychology in Criminal Justice*, 2(3), 29-51.
- Cornell, D. G. & Hawk, G. L. (1989). Clinical presentation of malingerers diagnosed by experienced forensic psychologists. *Law and Human Behavior*, 13(4), 375–383. <http://doi.org/10.1007/bf01056409>.
- Correa, A. A. (2018). Beyond Borders: Cultural and transnational perspectives of feigning and other response styles. In R. Rogers & S. D. Bender (Eds.), *Clinical Assessment of Malingering and Deception* (4th ed, pp. 61-80). The Guilford Press.
- Dandachi-FitzGerald, B., Duits, A. A., Leentjens, A. F. G., Verhey, F. R. J., & Ponds, W. H. M. (2020). Performance and symptom validity assessment in patients with apathy and cognitive impairment. *Journal of the International Neuropsychology Society*, 26(3), 314-321. <http://doi.org/10.1017/S1355617719001139>.
- Dearth, C. S., Berry, D. T. R., Vickery, C. D., Vagnini, V. L., Baser, R. E., Orey, S. A., & Cragar, D. E. (2005). Detection of feigned head injury symptoms on the MMPI-2 in head injured patients and community controls. *Archives of Clinical Neuropsychology*, 20(1), 95-110. <https://doi.org/10.1016/j.acn.2004.03.004>.
- DeClue, G. (2011). Harry Potter and the Structured Interview of Reported Symptoms? *Open Access Journal of Forensic Psychology*, 3, 1-17.
- Denney, R. L. (2008). Negative response bias and malingering during neuropsychological assessment in forensic settings. In R. L. Denney & J. P. Sullivan, *Clinical Neuropsychology in the Criminal Forensic Setting* (pp. 91-134). The Guilford Press.
- Dhilon, S., Bagby, R. M., Kushner, S. C., & Burchett, D. L. (2016). The impact of underreporting and overreporting on the validity of the Personality Inventory for DSM-5 (PID-5): A simulation analog design investigation. *Psychological Assessment*, 29(4), 1-6. <https://doi.org/10.1037/pas0000359>.
- Domingues, M. V. (2019). *Self-Report Symptom Inventory (SRSI): Validação através de um estudo de simulação análogo* (Dissertação de Mestrado). Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Dwarkadas, A. K. (2018). *Self-Report Symptom Inventory (SRSI): Adaptação, Validação e Normas com base em Estudos na Comunidade* (Dissertação de Mestrado). Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Edens, J. F., Otto, R. K., & Dwyer, T. (1999). Utility of the Structured Inventory of Malingered Symptomatology in identifying Persons Motivated to Maligner Psychopathology. *Journal of the American Academy of Psychiatry and Law*, 27(3), 387-396. <http://doi.org/10.1037/lhb0000361>.
- Edens, J. F., Poythress, N. G., & Watkins-clay, M. M. (2007). Detection of malingering in psychiatric unit and general population prison inmates: A comparison of the PAI, SIMS, and SIRS. *Journal of Personality*

- Assessment*, 88(1), 33-42.
<https://doi.org/10.1080/00223890709336832>.
- Edens, J. F., Truong, T. N., & Otto, R. K. (2020). Classification accuracy of the Rare Symptoms and Symptom Combination Scales of the Structured Inventory of Malingered Symptomatology in tree archival samples. *Law and Human Behavior*, 44(2), 167-177.
<http://dx.doi.org/10.1037/lhb0000361>.
- Geurten, M., Meulemans, T., & Seron, X. (2018) Detecting over-reporting of symptoms: the French version of the self-report symptom inventory. *The Clinical Neuropsychologist*, 32(1), 164-181.
<https://doi.org/10.1080/13854046.2018.1524027>.
- Giger, P. & Merten, T. (2019). Equivalence of the German and the French versions of the Self-Report Symptom Inventory. *Swiss Journal of Psychology*, 78(2), 5-13. <https://doi.org/10.1024/1421-0185/a000218>.
- Giger, P., Merten, T., Merckelbach, H., & Oswald, M. (2010). Detection of feigned crime-related amnesia: A multi-method approach. *Journal of Forensic Psychology Practice*, 10(5), 440-463.
<https://doi.org/10.1080/15228932.2010.489875>.
- Gillard, N. D. (2010). Methodological issues in malingering research: The use of simulation designs (Master thesis). University of North Texas, North Texas, United States of America.
- Glassmire, D. M., Tarescavage, A. M., & Gottfried, E. D. (2016). Likelihood of obtaining Structured Interview of Reported Symptoms (SIRS) and SIRS-2 elevations among forensic psychiatric inpatients with screening elevations on the Miller Forensic Assessment of Symptoms Test. *Psychological Assessment*, 28(2), 1586-1596.
<https://doi.org/10.1037/pas0000289>.
- Gorny, I. G. D. & Merten, T. (2006). Symptom information—warning—coaching, *Journal of Forensic Neuropsychology*, 4(4), 71-97.
https://doi.org/10.1300/J151v04n04_05.
- Green, D. & Rosenfeld, B. (2011). Evaluating the Gold Standard: A Review and Meta-Analysis of the Structured Interview of Reported Symptoms. *Psychological Assessment*, 23(1), 95-107.
<https://doi.org/10.1037/a0021149>.
- Green, D., Rosenfeld, B., & Belfi, B. (2013). New and improved? A comparison of the original and revised versions of the Structured Interview of Reported Symptoms. *Assessment*, 20(2), 210-218.
<https://doi.org/10.1177/1073191112464389>.
- Guy, L. S. & Miller, H. A. (2004). Screening for malingered psychopathology in a correctional setting: Utility of the Miller-Forensic Assessment of Symptoms Test (M-FAST). *Criminal Justice and Behavior*, (31)5, 695-716.
<https://doi.org/10.1177/0093854804268754>
- Helvoort D. van, Merckelbach, H., & Merten, T. (2019). The Self-Report Symptom Inventory (SRSI) is sensitive to instructed feigning, but not to genuine psychopathology in male forensic inpatients: An initial

- study. *The Clinical Neuropsychologist*, 33(6), 1069-1082.
<https://doi.org/10.1080/13854046.2018.1559359>
- Hopwood, C. J., Orlando, M. J., & Clark, T. S. (2010). The detection of malingered pain-related disability with the Personality Assessment Inventory. *Rehabilitation Psychology*, 55(3), 307-310.
<https://doi.org/10.1037/a0020516>.
- Huhnt, B. (2013). *Stone Cold Manipulation or Just an Excuse: Psychopathy and blame attribution as possible explanations for faking symptoms in a forensic setting* (Dissertação de Mestrado). Universidade de Maastricht, Maastricht, Países Baixos.
- Impelen, A. van, Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist*, 28 (8), 1336-1365. <https://doi.org/10.1080/13854046.2014.984763>.
- Impelen, A. van, Merckelbach, H., Jelicic, M., Niesten, I. J. M., & Campo, J. à (2017). Differentiating factitious from malingered symptomatology: The development of a psychometric approach. *Psychological Injury and Law*, 10 (4), 341-357. <https://doi.org/10.1007/s12207-017-9301-y>.
- Jelicic, M., Hessels, A., & Merckelbach, H. (2006). Detection of feigned psychosis with the Structured Inventory of Malingered Symptomatology (SIMS): A study of coached and uncoached simulators. *Journal of Psychopathology and Behavioral Assessment*, 28(1), 19-22. <https://doi.org/10.1007/s10862-006-4535-0>.
- Jelicic, M., Merckelbach, H., Candel, I., & Geraerts, E. (2007). Detection of feigned cognitive dysfunction using special malingering tests: A simulation study in naïve and coached malingerers. *International Journal of Neuroscience*, 117(8), 1185-1192.
<https://doi.org/10.1080/00207450600934697>.
- King, J. & Sullivan, K. A. (2009). Deterring malingered psychopathology: The effect of warning simulating malingerers. *Behavioral Sciences and the Law*, 27(1), 35-49. <https://doi.org/10.1007/10.1002/bsl.839>.
- Larrabee, G. (2015). The multiple validities of neuropsychological assessment. *American Psychologist*, 70(8), 779-788. <https://doi.org/10.1037/a0039835>.
- Lewis, J. L., Simcox, A. M., & Berry, D. T. R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment*, 14(2), 170-176.
<https://doi.org/10.1037/1040-3590.14.2.170>.
- Marchi, B. De & Balboni, G. (2018). Detecting malingering mental illness in forensics: Known-Group Comparison and Simulation Design with MMPI-2, SIMS and NIM. *Peer J*, 6:e5259.
<http://doi.org/10.7717/peerj.5259>.
- Marion, B. E., Sellbom, M., & Bagby, R. M. (2011). The detection of feigned psychiatric disorders using the MMPI-2-RF overreporting validity scales: An analogue investigation. *Psychological Injury and Law*, 4(1), 1-12. <https://doi.org/10.1007/s12207-011-9097-0>.

- Marion, B. E., Sellbom, M., Salekin, R. T., Toomey J. A., Kucharski, T., & Duncan, S. (2013). An examination of the association between psychopathy and dissimulation using the MMPI-2-RF validity scales. *Law and Human Behavior, 37*(4), 219-230. <https://doi.org/10.1037/lhb0000008>.
- Martin, P. K. & Schroeder, R. W. (2020). Base rates of invalid test performance across clinical non-forensic contexts and settings. *Archives of Clinical Neuropsychology, 00*. <https://doi.org/10.1093/arclin/acia017>.
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of north American professionals. *The Clinical Neuropsychologist, 29*(6), 741-776. <https://doi.org/10.1080/13854046.2015.1087597>.
- Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernest, N., & Okut, H. (2020). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception. *The Clinical Neuropsychologist, 34*(1), 88-119. <https://doi.org/10.1080/13854046.2019.1637027>.
- McDermott, B. E. & Sokolov, G. (2009). Malingering in a correctional system: The use of the Structured Interview of Reported Symptoms in a jail sample. *Behavioral Sciences and the Law, 27*(5), 753-765. <https://doi.org/10.1002/bsl.892>.
- McDermott, B. E., Dualan, I. V., & Scott, C. L. (2013). Malingering in the correctional system: Does incentive affect prevalence? *International Journal of Law and Psychiatry, 36*(3-4), 287-292. <https://doi.org/10.1016/j.ijlp.2013.04.013>.
- McGuire, B. E., Harvey, A. G., & Shores, E. A. (2001). Simulated malingering in pain patients: A study with the Pain Patient Profile. *British Journal of Clinical Psychology, 40*(1), 71-79. <https://doi.org/10.1348/014466501163490>.
- Merckelbach, H. & Smith, G. P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting malingering. *Archives of Clinical Neuropsychology, 18*(2), 145-152. [https://doi.org/10.1016/S0887-6177\(01\)00191-3](https://doi.org/10.1016/S0887-6177(01)00191-3).
- Merckelbach, H., Dandachi-FitzGerald, B., Helvoort, D. van, Jelacic, M., & Otgaar, H. (2019). When patients overreport symptoms: More than just malingering. *Current Directions in Psychological Science, 28*(3), 321-326. <https://doi.org/10.1177/0963721419837681>.
- Merckelbach, H., Smeets, T., & Jelacic, M. (2009). Experimental simulation: Type of malingering scenario makes a difference. *The Journal of Forensic Psychiatry & Psychology, 20*(3), 378-386. <http://dx.doi.org/10.1080/14789940802456686>.
- Merten, T. & Marckelbach, H. (2013a). Symptom validity testing in Somatoform and Dissociative Disorders: A critical review. *Psychological Injury and Law, 6*(2), 122-137. <https://doi.org/10.1007/s12207-013-9155-x>.

- Merten, T. & Rogers, R. (2017). An International perspective on feigned mental disabilities: Conceptual issues and continuing controversies. *Behavioral Sciences and the Law*, 35(2), 97-112. <https://doi.org/10.1002/bsl.2274>.
- Merten, T. & Merckelbach, H. (2013b). Introduction to malingering research and symptom validity assessment. *Journal of Experimental Psychopathology*, 4(1), 3–5. <https://doi.org/10.1177/204380871300400102>.
- Merten, T., Dandachi-FitzGerald, B., Hall, V., Schmand, 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. [https://doi.org/10.1016/S1130-5274\(13\)70014-8](https://doi.org/10.1016/S1130-5274(13)70014-8).
- Merten, T., Lorenz, R., & Schlatow, S. (2010). Posttraumatic Stress disorder can easily be faked, but faking can be detected in most cases. *German Journal of Psychiatry*, 13(3), 140-149.
- Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A new instrument for the assessment of distorted symptom endorsement. *Psychological Injury and Law*. <https://doi.org/10.1007/s12207-016-9257-3>
- Morey, L. C. & Lowmaster, S. E. (2010). Personality Assessment Inventory. In I.B. Weiner & W. E. Craighead (Eds), *Corsini Encyclopedia of Psychology* (4th Edition, pp. 1200-1203). John Wiley.
- Parks, A. C., Gfeller, J., Emmert, N., & Lammert, H. (2017). Detecting feigned postconcussional and posttraumatic stress symptoms with the structured inventory of malingered symptomatology (SIMS). *Applied Neuropsychology: Adult*, 24(5), 429-438. <https://doi.org/10.1080/23279095.2016.1189426>
- Pinheiro, D. A. da S. (2019). Self-Report Symptom Inventory (SRSI): Estudo de validação em uma amostra de reclusos (Dissertação de Mestrado). Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.
- Pollock, P. H., Quigley, B., Worley, K. O., & Bashford, C. (1997). Feigned mental disorder in prisoners referred to forensic mental health services. *Journal of Psychiatric and Mental Health Nursing*, 4(1), 9-15. <https://doi.org/10.1111/j.1365-2850.1997.tb00171.x>.
- Puente-López, E., Pina, D., Ruiz-Hernández, J. A., & Llor-Esteban, B. (2020). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in motor vehicle accident patients. *The Journal of Forensic Psychiatry & Psychology*. <https://doi.org/10.1080/14789949.2020.1833073>
- Resnick, P. J. & Knoll, J. L. (2018). Malingered psychosis. In R. Rogers & S. D. Bender (Eds.), *Clinical Assessment of Malingering and Deception* (4th ed, pp. 98-121). The Guilford Press.
- Rogers, R. & Bender, S. (2012). Evaluation of malingering and related response styles. In I. B. Weiner (Ed.), *Handbook of Psychology – Vol.11: Forensic Psychology* (2nd Ed., pp. 517-540). Wiley.

- Rogers, R. & Bender, S. D. (2003). Evaluation of malingering and deception. In A. M. Goldstein & I. B. Weiner (Eds.), *Handbook of Psychology – Vol.11: Forensic Psychology* (pp. 109-129). Wiley.
- Rogers, R. & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, 22(3), 273-285. <https://doi.org/10.1023/a:1025702405865>.
- Rogers, R. (2018a). An Introduction to Response Styles. In R. Rogers & S. D. Bender (Eds.), *Clinical Assessment of Malingering and Deception* (4th ed, pp. 3-17). The Guilford Press.
- Rogers, R. (2018b). Detection Strategies for Malingering and Defensiveness. In R. Rogers & S. D. Bender (Eds.), *Clinical Assessment of Malingering and Deception* (4th ed, pp. 18-41). The Guilford Press.
- Rogers, R. (2018c). Researching Response Styles. In R. Rogers & S. D. Bender (Eds.), *Clinical Assessment of Malingering and Deception* (4th ed, pp. 592-614). The Guilford Press.
- Rogers, R. (2018d). Structured Interviews and Dissimulation. In R. Rogers & S. D. Bender (Eds.), *Clinical Assessment of Malingering and Deception* (4th ed, pp. 422-448). The Guilford Press.
- Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment*, 67(2), 244-257. https://doi.org/10.1207/s15327752jpa6702_2.
- Rogers, R., Jackson, R. L., Sewell, K. W., & Harrison, K. S. (2004). An Examination of the ECST-R as a Screen for Feigned Incompetency to Stand Trial. *Psychological Assessment*, 16(2), 139–145. <https://doi.org/10.1037/1040-3590.16.2.139>.
- Rogers, R., Robinson, E. V., & Gillard, N. D. (2014). The SIMS screen for feigned mental disorders: The development of detection-based scales. *Behavioral Sciences and the Law*, 32(4), 455-466. <https://doi.org/10.1002/bsl.2131>.
- Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior*, 22(4), 353-367. <https://doi.org/10.1023/A:1025714808591>
- Rogers, R., Sewell, K. W., & Ustad, K. L. (1995). Feigning among chronic outpatients on the MMPI-2: A systematic examination of fake-bad indicators. *Assessment*, 2(1), 81-89. <https://doi.org/10.1177/1073191195002001008>.
- Rogers, R., Vitacco, M. J., & Kurus, S. J. (2010) Assessment of malingering with repeat forensic evaluations: Patient variability and possible misclassification on the SIRS and other feigning measures. *Journal of the American Academy of Psychiatry and Law*, 38(1), 109-114.
- Rubenzler, S. (2010). Review of the Structured Interview of Reported Symptoms-2 (SIRS-2). *Open Access Journal of Forensic Psychology*, 2, 273-286.

- Rüsseler, J., Brett, A., Klaue, U., Sailer, M., & Münte, T. (2008). The effect of coaching on the simulated malingering of memory impairment. *BMC Neurology*, 8(37). <https://doi.org/10.1186/1471-2377-8-37>.
- Sahoo, S., Kumar, R., & Oomer, F. (2020). Concepts and controversies of malingering: A re-look. *Asian Journal of Psychiatry*, 50, 101952. <https://doi.org/10.1016/j.ajp.2020.101952>.
- Schoenberg, M., Dorr, D., & Morgan, C. D. (2003). The ability of the Millon Clinical Multiaxial Inventory – Third Edition to detect malingering. *Psychological Assessment*, 15(2), 198-204. <https://doi.org/10.1037/1040-3590.15.2.198>.
- Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T., & Duncan, S. (2010). Utility of the MMPI-2-RF (Restructured Form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment*, 22(1), 22-31. <https://doi.org/10.1037/a0018222>.
- Sellbom, M., Wygant, D., & Bagby, M. (2011). Utility of the MMPI-2-RF in detecting non-credible somatic complaints. *Psychiatric Research*, 197(3), 295-301. <https://doi.org/10.1016/j.psychres.2011.12.043>
- Seron, X. (2014). Lying in neuropsychology. *Neurophysiologie Clinique/Clinical Neurophysiology*, 44(4), 389-403. <http://dx.doi.org/10.1016/j.neucli.2014.04.002>.
- Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the Neuropsychological Dysfunction Criteria. *Archives of Clinical Neuropsychology*, 00, 1-30. <https://doi.org/10.1093/arclin/acia019>.
- Simões, M. R., Almiro, P. A., Mota, M., Oliveira, C. S., Maior, F. S., Cunha, C., Domingues, A. F., Pinho, J. I., Soares, D. M., Almeida, D., Barreto, M., Duarte, F., & Dias, I. T. (2017). Inventário Estruturado de Simulação de Sintomas (SIMS). In M. R. Simões, L. S. Almeida, & M. M. Gonçalves (Coord.), *Psicologia Forense: Instrumentos de Avaliação* (pp. 397-415). Factor.
- Simões, M. R., Almiro, P. A., Nunes, A. P., Cunha, C., Almeida, D., Domingues, A. F., Ferreira, S., Paiva, J., Dias, I. T., & Cruz, I. (2017). Escala de Validade de Sintomas – Versão 2. In M. R. Simões, L. S. Almeida, & M. M. Gonçalves (Coord.), *Psicologia Forense: Instrumentos de Avaliação* (pp. 417-435). Factor.
- Simões, M. R., Sousa, L. B., Fonseca, M. S., Garcia, S. G., Pinho, J. I., Soares, D. M., Silva, D. C., Oliveira, C. S., Mota, M., Maior, F. S., Jesus, S. S., Fernandes, S. P., Faustino, J. P., Cruz, A. P., Barreto, M., Dias, I. T., Perdiz, C., Douro, D., Santana, I., & Pinho, M. S. (2017). Teste de Simulação de Problemas de Memória (TOMM). In M. R. Simões, L. S. Almeida, & M. M. Gonçalves (Coord.), *Psicologia Forense: Instrumentos de Avaliação* (pp. 353-375). Factor.
- Slick, D., Sherman, E., & Iverson, G. L. (1999). Diagnostic criteria for Malingered Neurocognitive Dysfunction: Proposed standards for

- clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 545-561. [http://doi.org/10.1076/1385-4046\(199911\)13:04;1-Y;FT545](http://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545).
- Smith, G. P. (2018). Brief measures for the detection of feigning and impression management. In R. Rogers & S. D. Bender (Eds.), *Clinical Assessment of Malingering and Deception* (4th ed, pp. 449-472). The Guilford Press.
- 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*, 25(2), 183-189.
- Springman, R. E. & Vandenberg, B. R. (2009). The effects of teste strategy coaching on measures of competency to stand in trial. *Journal of Forensic Psychology Practice*, 9(3), 179-198. <http://doi.org/10.1080/15228930902809856>.
- Steffan, J. S. & Morgan, R. (2008). Diagnostic accuracy of the MMPI-2 Malingering Index in the detection of malingering among inmates. *Journal of Personality Assessment*, 90(4), 392-398. <https://doi.org/10.1080/00223890802108204>.
- Suhr, J. A. & Gunstad, J. (2007). Coaching and malingering: A review. In G. J. Larrabee (Ed.), *Assessment of Malingered Neuropsychological Deficits* (pp. 287-311). Oxford University Press.
- Sullivan, K. & King, J. (2010). Detecting faked psychopathology: a comparison of two tests to detect malingered psychopathology using a simulation design. *Psychiatry Research*, 176(1), 75-81. <https://doi.org/10.1016/j.psychres.2008.07.013>.
- Sullivan, K. A. & Elliot, C. (2012). An investigation of the validity of the MMPI-2 Response Bias Scale using an analogue simulation design. *The Clinical Neuropsychologist*, 26(1), 160-176. <https://doi.org/10.1080/13854046.2011.647084>.
- Sweet, J. & Breting, L. M. G. (2013). Symptom validity test research: Status and clinical implications. *Journal of Experimental Psychopathology*, 4(1), 6-19. <https://doi.org/10.5127/jep.022311>.
- Victor, T. L. & Abeles, N. (2004). Coaching clients to take psychological and neuropsychological tests: A clash of ethical obligations. *Professional Psychology: Research and Practice*, 35(4), 373-379. <http://doi.org/10.1037/0735-7028.35.4.373>.
- Vilar, M., Sousa, L. B., Fonseca, M. S., Lages, H., Garcia, S. G., Pinho, J. I., Soares, D. M., Oliveira, C. S., Jesus, S. S., Fernandes, S. P., Barreto, M., Dias, I. T., Perdiz, C., Douro, D., Santana, I., Firmino, H., Pinho, M. S., & Simões, M. R. (2017). Teste de Memória de Rey – 15 Itens (15 IMT). In M. R. Simões, L. S. Almeida, & M. M. Gonçalves (Coord.), *Psicologia Forense: Instrumentos de Avaliação* (pp. 377-396). Pactor.
- Wall, T. D., Wygant, D. B., & Gallagher, R. W. (2015). Identifying overreporting in a correctional setting: Utility of the MMPI-2 Restructured form validity scales. *Criminal Justice and Behavior*, 42(6), 610-622. <https://doi.org/10.1177/0093854814556881>.

- Weinborn, M., Woods, S. P., Nulsen, C., & Leighton, A. (2012). The effects of coaching on the verbal and nonverbal symptom validity tests. *The Clinical Neuropsychologist*, *26*(5), 832-849.
<http://doi.org/10.1080/13854046.2012.686630>.
- Wisdom, N. M., Callahan, J. L., & Shaw, T. G. (2010). Diagnostic utility of the Structured Inventory of Malingered Symptomatology to detect malingering in a forensic sample. *Archives of Clinical Neuropsychology*, *25*(2), 118-125.
<http://doi.org/10.1093/arclin/acp110>.
- Wygant, D. B., Ben-Porath, Y., Arbisi, P. A., Berry, D. T. R., Freeman, D. B., & Heilbronner, R. L. (2009). Examination of the MMPI-2 Restructured Form (MMPI-2-RF) validity scales in civil forensic settings: Findings from simulation and known group studies. *Archives of Clinical Psychology*, *24*(7), 671-680.
<http://doi.org/10.1093/arclin/acp073>.
- Young, G. (2014a). PTSD and Malingering: Tests, Diagnostics, Cut Scores, and Cautions. In G. Young, *Malingering, feigning, and response bias in psychiatric/psychological injury: Implications for practice and court* (pp. 855-880). Springer Science & Business Media.
- Young, G. (2014b). Assessing Psychological Injuries and Malingering: Disability and Report Writing. In G. Young, *Malingering, feigning, and response bias in psychiatric/psychological injury: Implications for practice and court* (pp. 361-384). Springer Science & Business Media.
- Young, G. (2014c). Malingering: Definitional and Conceptual Ambiguities and Prevalence Base Rates. In G. Young, *Malingering, feigning, and response bias in psychiatric/psychological injury: Implications for practice and court* (pp. 25-52). Springer Science & Business Media.

Anexos

Anexo A – Tabela 2.1 e Tabela 2.2

Tabela 2.1

Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto criminal.

Estudo	Amostra	TVS	Design	CGC	Condições
Rogers et al. (1996)	RPP (53).	MMPI-A; SIRS; SIMS.	SDi.	Não se aplica.	RPP – SI (3); RPP – C. PP – C;
Rogers et al. (2004)	RZC (96); PP (56).	ECST-R; M-FAST; SIRS.	SDi.	Sim.	RZC – C; RZC – SI. RZC – C;
Edens et al. (2007)	RZC (60); RPP (30); RS - 26.	PAI; SIMS; SIRS.	SD; KGC.	Sim.	RZC – SI; RPP – C; RS - C.
Steffan & Morgan (2008)	RZC (45); RPP (46).	MMPI-2.	SD.	Sim.	RZC – SI; RPP – C.
Wall et al. (2015)	RZC (196); RPP (73).	MMPI-2- RF.	SD.	Sim.	RZC – C; RZC – SI; RPP – C.
Marchi & Balboni (2018)	RZC (71); RPP (57).	MMPI-2; SIMS; PAI (NIM).	SD; KGC.	Não.	RZC – C; RZC – SI.
Helvoort et al. (2019)	RPP (41).	SRSI.	SDi.	Não se aplica.	RPP – C; RPP – SI.

Nota. CGC – Comparação com Grupo Clínico; RZC – Reclusos da zona comum; RPP – Reclusos pacientes psiquiátricos; PP – Pacientes psiquiátricos; RS – Reclusos suspeitos de simulação de sintomas psiquiátricos; SD – Design de simulação análogo; SDi - Design de simulação análogo intra-sujeitos; KGC – Comparação de grupos conhecidos; ECST-R – *Evaluation of Competency to Stand Trial – Revised*; M-FAST – *Miller Forensic Assessment of Symptoms Test*; SI – Instrução para simular; C – Instrução para responder honestamente.

Tabela 2.2

Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto criminal (continuação).

Estudo	Incentivo	QVM	Características das instruções
Rogers et al. (1996)	Sim.	Não.	Sem cenário; simulação de uma de três perturbações; Incluiu: DS.
Rogers et al. (2004)	Sim.	Sim.	Cenário relevante: simulação de perturbação mental que provoca incapacidade para ser julgado de forma justa; Incluiu: AE.
Edens et al. (2007)	Sim.	Não.	Sem cenário; simulação de perturbação mental. Incluiu: AE.
Steffan & Morgan (2008)	Não.	Não.	Cenário relevante, mas pouco explícito: simulação de sintomas, de modo a ser admitido em instituição prisional psiquiátrica regional; Incluiu: AE.
Wall et al. (2015)	Não.	Não.	Cenário relevante, mas pouco elaborado: simulação de sintomas, de modo a poderem ter cela própria e não trabalharem.
Marchi & Balboni (2018)	Não.	Não.	Sem cenário; preenchimento dos testes como pacientes psiquiátricos.
Helvoort et al. (2019)	Sim.	Sim.	Cenário irrelevante: simulação de incapacidade para exercer atividade laboral, perante avaliação psicológica da companhia de seguros; Incluiu: AE.

Notas. QVM – Questionário de Verificação de Manipulação; AE - Aviso para não incorrer em exagero de sintomas relatados/ser credível na apresentação; DS – Dicas sobre como simular uma perturbação de forma eficiente.

Anexo B – Tabela 3.1 e Tabela 3.2

Tabela 3.1

Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto médico e médico-legal.

Estudo	Amostra	TVS	Design	Condições
Rogers et al. (1995)	PP (42)	MMPI-2	SDi.	PP - C; PP - SI.
McGuire et al. (2001)	PCD (88); PP (34).	P3	SD.	PCD - C; PCD - SI; PP - SI.
Dearth et al. (2005)	PP (49); PC (53)	MMPI-2	SD.	PP - C; PP - SI; PC - C; PC - SI.
Wygant et al. (2009)	Estudo 1: PP (46); Estudo 2: PFF (90)	MMPI-2-RF.	Estudo 1: SD; Estudo 2: SDi.	Estudo 1: PP - C; PP - SI; Estudo 2: PFF - C; PFF - SI.
Hopwood et al. (2010)	PCD (317); EU (152).	PAI.	SD.	PCD - C; EU - SI (2).
Sellbom et al. (2011)	EU (114); PFF (72); PP (69).	MMPI-2-RF.	SD.	EU - SI; PFF - C; PP - C.
Puente-López et al. (2020)	EU (401); PVA (195); PVAS (54).	SIMS	SD.	EU - C; EU - SI; PVA - C; PVAS - C.

Notas. CGC – Comparação com Grupo Clínico; SD – *Design* de simulação análogo; SDi - *Design* de simulação análogo intra-sujeitos; PP - Pacientes psiquiátricos; PCD – Pacientes com problemáticas de dor crónica; PFF - Pacientes com incapacidades/danos a nível físico; PC – Participantes da comunidade; EU – Estudantes universitários; PVA – Pacientes vítimas de acidente; PVAS – Pacientes vítimas de acidente com relato de sintomas exagerado; P3 – *Pain Patient Profile*; SI – Instrução para simular; C – Instrução para responder honestamente.

Tabela 3.2

Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto médico e médico-legal (continuação).

Estudo	Incentivo	QVM	Características das instruções
Rogers et al. (1995)	Sim.	Não.	Cenário relevante: simulação ou exagero de sintomas psiquiátricos, de modo a conseguir internamento; Incluiu: AE; AP.
McGuire et al. (2001)	Não.	Sim.	Cenário relevante: Simulação ou exagero de sintomas relacionados com dor, em situação de avaliação para incompetência; Incluiu: AE.
Dearth et al. (2005)	Sim.	Não.	Cenário relevante: simulação de incompetência para retornar a funções laborais, após ferimento na cabeça. Incluiu: AE.
Wygant et al. (2009)	Sim.	Estudo 1: Não; Estudo 2: Sim.	Cenários relevantes; Estudo 1: Simulação de sintomas decorrentes de danos cerebrais, de modo a ser declarada incapacidade para trabalhar; Incluiu: AE; DS. Estudo 2: Simulação de sintomas físicos e emocionais decorrentes de serviço militar para obtenção de compensação; Incluiu: AE.
Hopwood et al. (2010)	Sim (EU).	Não.	Cenário relevante (indisponível para consulta): simulação de sintomas relacionados com dor, em avaliação de incompetência; Incluiu: AP (apenas num grupo).
Sellbom et al. (2011)	Sim (EU)	Sim.	Cenário relevante: simulação/exagero de sintomas, para receber compensação por acidente de trabalho; Incluiu: AE.
Puente-López et al. (2020)	Sim.	Sim.	Cenário relevante: simulação de sintomas físicos e psicológicos, consequência de um acidente de viação, para obter compensação; Incluiu: AE; AP; DS.

Estudos com *design* de simulação análogo em Testes de Validade de Sintomas (TVS):
Uma revisão da literatura
Carolina Duarte de Almeida (carolinadalmeida2@gmail.com) 2020

Notas. QVM – Questionário de Verificação de Manipulação; AE - Aviso para não incorrer em exagero de sintomas relatados/ser credível na apresentação; DS – Dicas sobre como simular uma perturbação de forma eficiente; AP – Aviso sobre a presença de TVS.

Anexo C – Tabela 4.1 e Tabela 4.2

Tabela 4.1

Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto comunitário.

Estudo	Amostra	TVS	Design	CGC	Condições
Smith & Burger (1997)	EU (476).	SIMS.	SD.	Não.	EU – C; EU (7)
Edens et al. (1999)	EU (196).	SIMS.	SDi.	Não.	EU – C; EU – SI (3).
Merckelbach & Smith (2003)	EU (82); PP (10).	SIMS.	SD.	Sim.	EU – C; EU – SI (3); PP – C.
Schoenberg et al. (2003)	EU (236); PP (181).	MCMI- III.	SD.	Sim.	EU – C; EU – SI; PP – C.
Jelicic et al. (2006)	EU (60).	SIMS.	SD.	Não.	EU – C; EU – SI (3).
Jelicic et al. (2007)	EU (90).	SIMS.	SD.	Não.	EU – C; EU – SI.
Springman & Vandenberg (2009)	EU (101).	ECST- R; GCCT.	SD.	Não.	EU – C; EU – SI (2).
King & Sullivan (2009)	EU (67).	PAI.	SD.	Não.	EU – C; EU – SI (2).
Sullivan & King (2010)	EU (41).	PAI.	SD.	Não.	EU – C; EU – SI.
Marion et al. (2011)	EU (157); PCSM (159); PP (237).	MMPI- 2-RF.	SD.	Sim,	PP – C; EU – SI; PCSM – SI.
Sullivan & Elliot (2012)	EU (81).	MMPI- 2.	SD.	Não.	EU – C; EU – SI.

Estudo	Amostra	TVS	Design	CGC	Condições
Christiansen & Vincent (2012)	EU (264).	M-FAST.	SD.	Não.	EU – C (3); EU – SI.
Marion et al. (2013)	Estudo 1: EU (465).	Estudo 1: MMPI-2-RF.	Estudo 1: SD.	Não.	EU – C; EU – SI; EU – DI.
Dhilon et al. (2016)	EU (228).	MMPI-2-RF.	SDi.	Não.	EU -C; EU – SI; EU – DI.
Impelen et al. (2017)	Estudo 1: EU (96); PP (34); Estudo 2: EU (64); PC (24); PP (40).	Estudo 1: SIMS; SDI; M-FAST; Estudo 2: SIMS; SDI.	SD.	Sim.	Estudo 1: EU – C; PP – C; EU – SI (2); Estudo 2: PP – C; EU + PC – SI (3) SI.
Guerten et al. (2018)	PC (643); PP (13).	SRSI.	SD.	Sim.	PC – C; PP – C; PC – SI (2).
Boskovic et al. (2018)	EU (22).	SRSI.	SDi.	Não.	EU – C; EU – SI.
Boskovic, Merckelbach, et al. (2019)	EU (158).	SRSI.	SD.	Não.	EU – C; EU – SI (2).
Giger & Merten (2019)	PC (40).	SRSI; SIMS.	SDi.	Não.	EU – C; EU – SI.
Domingues (2019)	EU (70).	SRSI; SIMS.	SDi;	Não.	EU – C; EU – SI (2).

Notas. CGC – Comparação com Grupo Clínico SD – *Design* de simulação análogo; SDi - *Design* de simulação análogo intra-sujeitos; EU - Estudantes universitários; PP - Pacientes psiquiátricos; PC – Participantes da comunidade; PCSM – Participantes com conhecimento de saúde mental; ECST-R – *Evaluation of Competency to Stand Trial – Revised*; GCCT – *Georgia Court Competence Test*; MCMI-III – *Millon Clinical Multiaxial Inventory – 3ª Edição*; FBS – *Fake Bad Scale*; SDI – *Symptom and Disposition Interview*; M-FAST – *Miller Forensic Assessment of Symptoms Test*; SI – Instrução para simular; C – Instrução para responder honestamente; DI – Instrução para responder com estilo de resposta defensivo.

Tabela 4.2

Análise da metodologia utilizada nas investigações com design de simulação análogo em contexto comunitário (continuação).

Estudo	Incentivo	QVM	Características das instruções de simulação
Smith & Burger (1997)	Sim.	Não.	Cenário relevante, pouco explícito: simulação de diferentes perturbações, para evitar sentença por agressão; Incluiu: AE.
Edens et al. (1999)	Sim.	Sim.	Cenário relevante, pouco explícito: simulação de sintomatologia após envolvimento em processo judicial civil. Incluiu: AE; AP.
Merckelbach & Smith (2003)	Não.	Não.	Cenário indicado como relevante (forense), mas não revelado.
Schoenberg et al. (2003)	Sim.	Sim.	Cenário não relevante, pouco explicitado: simulação de problemas psicológicos e emocionais severos, de modo que cumprissem requisitos para internamento hospitalar; Incluiu: IS.
Jelicic et al. (2006)	Sim.	Não.	Cenário não relevante: pouco explícito; simulação de psicose para evitar responsabilização por homicídio; Incluiu: grupo A – IS; grupo B – IS; AP; AE.
Jelicic et al. (2007)	Sim.	Não.	Cenário relevante, pouco explícito: simulação e exagero de sintomas decorrentes de um acidente; Incluiu, num grupo: IS; AE.
Springman & Vandenberg (2009)	Sim.	Sim.	Cenário relevante: simulação de incompetência para ser julgado, após acidente de viação fatal; Incluiu: AE; DS; AP.
King & Sullivan (2009)	Sim.	Sim.	Sem cenário; Incluiu: IS; num grupo adicionalmente: AP.
Sullivan & King (2010)	Sim.	Não.	Sem cenário; Incluiu: IS; num grupo adicionalmente: AP.

Estudo	Incentivo	QVM	Características das instruções de simulação
Marion et al. (2011)	Sim.	Não.	Sem cenário; Incluiu: AP; AE. Cenário relevante:
Sullivan & Elliot (2012)	Sim.	Sim.	simulação/exagero de sintomas resultantes de atropelamento, para conseguir compensação monetária. Cenário relevante, sem indicações explícitas: manipulação da sugestão para simular sintomas e da presença em condição de litígio após acidente de viação.
Christiansen & Vincent (2012)	Sim.	Sim.	Sem cenário: apresentação de forma negativa. Cenário relevante, mas pouco explícito: simulação de sintomas de modo a conseguir benefícios no curso.
Marion et al. (2013)	Não.	Sim.	Incluiu: AE. Cenários relevantes (todos os participantes, excluindo PP, recebiam o mesmo cenário de contextualização, sendo depois manipulados detalhes, de acordo com as condições): Estudo 1: Avaliação psicológica realizada por um conselheiro de estudantes, na unidade de apoio psicológico da universidade (condições perturbação factícia, ou simulação de ansiedade); Estudo 2: Envolvimento em episódio violento numa festa (condições de perturbação factícia, simulação em processo civil, ou criminal). Cenário relevante: exagero de sintomas resultantes de acidente de viação de modo a conseguir compensação;
Dhilon et al. (2016)	Não.	Não.	Incluiu: IS (apenas num grupo).
Impelen et al. (2017)	Sim.	Sim.	
Geurten et al. (2018)	Não.	Não.	

Estudo	Incentivo	QVM	Características das instruções de simulação
Boskovic et al. (2018)	Sim.	Sim.	Cenário relevante, mas complexo: simulação de problemas de ansiedade em situações de avaliação, na universidade, de modo a ajudar um amigo. Dois cenários completos, distribuídos aleatoriamente: Cenário 1 não relevante: simulação de sintomatologia pós-traumática, depois de assistência a um acidente, imaginando-se no papel de um polícia;
Boskovic, Merckelbach, et al. (2019)	Sim.	Não.	Cenário 2 relevante: simulação de sintomatologia de dor crónica, depois de um incidente na estrada, para conseguir compensação; Incluiu: AE;
Giger & Merten (2019)	Sim.	Sim.	Cenário relevante: simulação de persistência dos sintomas resultantes de acidente de viação, para obter compensação. Incluiu: AE.
Domingues (2019)	Não.	Sim.	Cenário e instruções adaptadas de Boskovic, Merckelbach, et al. (2019).

Notas. QVM – Questionário de Verificação de Manipulação; IS- Informação sobre os sintomas; AP - Aviso sobre presença de TVS; AE - Aviso para não incorrer em exagero de sintomas relatados/ser credível na apresentação; DS – Dicas sobre como simular uma perturbação de forma eficiente.