



**UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO
CURSO DE PÓS-GRADUAÇÃO EM PSICOLOGIA**

RAÍZA LORENA SANDES SOUZA

**PROPRIEDADES PSICOMÉTRICAS DA ESCALA DE SUGESTIONABILIDADE DE
GUDJONSSON (GSS): uma revisão sistemática e metanálise.**

PETROLINA

2024

RAÍZA LORENA SANDES SOUZA

**PROPRIEDADES PSICOMÉTRICAS DA ESCALA DE
SUGESTIONABILIDADE DE GUDJONSSON (GSS): uma revisão
sistemática e metanálise.**

Dissertação apresentada a Universidade Federal do Vale do São Francisco – UNIVASF, Campus Petrolina, como requisito para obtenção do título de mestre em psicologia.

Orientadora: Prof^a. Dr^a. Marina Pereira Gonçalves.
Coorientador: Prof. Prof. Dr^o Ricardo Freitas Dias

**PETROLINA
2024**

S729p Souza, Raíza Lorena Sandes
Propriedades Psicométricas da Escala de Sugestionabilidade de Gudjonsson (GSS): uma revisão sistemática e metanálise / Raíza Lorena Sandes Souza. – Petrolina - PE, 2024.
xiii, 103 f. : il.

Dissertação (Mestrado em Psicologia) - Universidade Federal do Vale do São Francisco, Campus Petrolina-PE, 2024.

Orientadora: Profª. Drª. Marina Pereira Gonçalves.
Banca examinadora: João Carlos Alchieri, Debora Alves De Amorim.

Bibliografia.

1. Psicometria. 2. Propriedades psicométricas 3. Processos cognitivos. I. Título. II. Gonçalves, Marina Pereira. III. Universidade Federal do Vale do São Francisco.

CDD 152.3

**UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO CURSO DE
PÓS-GRADUAÇÃO EM PSICOLOGIA**

FOLHA DE APROVAÇÃO


RAÍZA LORENA SANDES SOUZA

**PROPRIEDADES PSICOMÉTRICAS DA ESCALA DE
SUGESTIONABILIDADE DE GUDJONSSON (GSS): uma revisão
sistemática e metanálise**


Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Psicologia, pela Universidade Federal do Vale do São Francisco – UNIVASF.

Aprovado em: 31 de julho de 2024.


Banca Examinadora

Documento assinado digitalmente
 **MARINA PEREIRA GONCALVES**
Data: 31/07/2024 16:07:03-0300
Verifique em <https://validar.iti.gov.br>

(Prof.^a Dr^a. Marina Pereira Gonçalves da Universidade Federal do Vale do São Francisco – UNIVASF).

Documento assinado digitalmente
 **JOAO CARLOS ALCHIERI**
Data: 31/07/2024 16:33:17-0300
Verifique em <https://validar.iti.gov.br>

(Prof^o. Dr^o João Carlos Alchieri da Universidade Federal do Rio Grande do Norte – UFRN).

Documento assinado digitalmente
 **DEBORA ALVES DE AMORIM**
Data: 14/08/2024 07:16:16-0300
Verifique em <https://validar.iti.gov.br>

(Prof^a. Dr^a Debora Alves De Amorim da Universidade Federal do Vale do São Francisco - UNIVASF)

AGRADECIMENTOS

À Deus pela possibilidade da existência...

À minha família, pelo apoio incondicional mesmo quando eu mesma não acreditava;

À todas as pessoas vítimas e testemunhas de violência que dividiram suas histórias comigo durante meus anos de prática na psicologia jurídica, minha busca por conhecimento também é por vocês;

À Prof^a. Dr^a. Marina Pereira Gonçalves, pela excelente orientação, toda a imensa paciência, sabedoria e gentileza com as quais conduziu todo o processo, nunca terei como agradecer...a você todo meu respeito e admiração.

Ao Prof. Dr^o Ricardo Freitas Dias pela gentileza de compartilhar todo seu conhecimento e experiência na área da revisão sistemática sempre presando pelo bom andamento do processo;

Aos colegas de equipe (Felipe Negreiros dos Santos e Jânio Luiz Correia Júnior) que contribuíram em todo o processo da revisão sistemática, minha gratidão.

" *São precisamente as perguntas para as quais não existem respostas que marcam os limites das possibilidades humanas e que traçam as fronteiras da nossa existência.* " " . – Milan Kundera.

RESUMO

A sugestionabilidade interrogativa é entendida como o grau em que os indivíduos aceitam e incorporam informação durante um interrogatório formal. As Escalas de Sugestionabilidade de Gudjonsson (GSS 1 e 2) são instrumentos utilizados em alguns países para a avaliação da sugestionabilidade em contextos forenses, clínicos e acadêmicos. Este estudo objetivou realizar uma Revisão Sistemática e Metanálise das propriedades psicométricas da GSS. Seguindo as recomendações do PRISMA, foi realizada busca em 11 bases de dados. Das 470 publicações encontradas, excluindo os duplicados restaram 194 publicações; após a leitura dos títulos e resumos dos artigos, considerando critérios de elegibilidade (apresentar as propriedades psicométricas das escalas; serem originais de pesquisa envolvendo seres humanos e estarem indexados nas bases), restaram 43 publicações, as quais foram avaliadas na íntegra, sendo incluídos ao final 17 artigos na Revisão Sistemática. Todas as fases foram realizadas por dois avaliadores, de forma independente, treinados para garantir maior confiabilidade. Os resultados da Metanálise indicaram que no geral a escala GSS apresenta evidências de fidedignidade com base no alfa de Cronbach com valores classificados como moderados (0,75 – 0,79) e justos (0,70 – 0,74), com correlações fortes no modelo de efeitos aleatórios. Porém, foram verificados no fator Yields um estudo com alfa de Cronbach abaixo do aceitável ($\alpha = 0,43$) com uma alta heterogeneidade nesse fator ($I^2 = 94,5\%$). Já no fator Shift, o alfa dos estudos metanalisados foi um pouco abaixo do recomendado pela literatura [$\alpha = 0,67$; IC 95% [0,65; 0,69], $p < 0,001$], com uma baixa heterogeneidade entre os estudos ($I^2 = 36,5\%$). A análise de subgrupos revelou diferenças significativas entre os tipos de escala GSS 1 e GSS 2, especialmente para o fator Yield. Conclui-se que apesar desses resultados indicarem índices aceitáveis de fidedignidade, é necessário ainda desenvolver mais estudos para esclarecer a estrutura interna da GSS (1 e 2), para que esta escala apresente contribuições práticas na entrevista investigativa e na psicologia do testemunho.

Palavras-chave: testemunho; sugestionabilidade; psicomетria.

ABSTRACT

Interrogative suggestibility is understood as the degree to which individuals accept and incorporate information during a formal interrogation. The Gudjonsson Suggestibility Scales (GSS 1 and 2) are instruments used in some countries to assess suggestibility in forensic, clinical, and academic contexts. This study aimed to conduct a Systematic Review and Meta-Analysis of the psychometric properties of the GSS. Following PRISMA recommendations, a search was conducted in 11 databases. Of the 470 publications found, excluding duplicates, 194 publications remained; after reading the titles and abstracts of the articles, considering eligibility criteria (presenting the psychometric properties of the scales; being original research involving humans and indexed in the databases), 43 publications were evaluated in full, and finally, 19 articles were included in the Systematic Review. All phases were carried out by two independent reviewers trained to ensure greater reliability. The results of the meta-analysis indicated that, in general, the GSS scale presents evidence of reliability based on Cronbach's alpha, with values classified as moderate (0.75–0.79) and fair (0.70–0.74), with strong correlations in the random effects model. However, one study was found in the Yields factor with a Cronbach's alpha below the acceptable level ($\alpha = 0.43$) and high heterogeneity in this factor ($I^2 = 94.5\%$). In the Shift factor, the alpha of the meta-analyzed studies was slightly below that recommended by the literature [$\alpha = 0.67$; 95% CI [0.65; 0.69], $p < 0.001$], with low heterogeneity between studies ($I^2 = 36.5\%$). The subgroup analysis revealed significant differences between the GSS 1 and GSS 2 scale types, especially for the Yield factor. It is concluded that although these results indicate acceptable reliability rates, it is still necessary to develop further studies to clarify the internal structure of the GSS (1 and 2), so that this scale presents practical contributions in investigative interviews and in the psychology of testimony.

Keywords: witness; suggestibility; psychometric.

LISTAS DE FIGURAS

Figura 1 -	Esquemas com os tipos e subtipos de evidências de validade	34
Figura 2 -	Flowchart da seleção dos estudos.	47
Figura 3 -	Avaliação da qualidade metodológica dos estudos incluídos	60
Figura 4 -	Comparação das evidências de consistência interna do fator Yield entre os estudos da escala GSS, do modelo bifatorial.	62
Figura 5 -	Comparação das evidências de consistência interna do fator Shift entre os estudos da escala GSS, do modelo bifatorial.	62
Figura 6 -	Comparação das evidências de consistência interna do fator Total entre os estudos da escala GSS, do modelo bifatorial.	63
Figura 7 -	Comparação das evidências de consistência interna do fator Yield, dividido pelos subgrupos GSS1 e GSS 2.	64
Figura 8 -	Comparação das evidências de consistência interna do fator Shift, dividido pelos subgrupos GSS1 e GSS 2.	64
Figura 9 -	Comparação das evidências de consistência interna do fator Shift, dividido pelos subgrupos GSS1 e GSS 2.	65

LISTAS DE TABELAS

Tabela 1 -	Esquema com os tipos de medidas possíveis com o uso das Escala de Sugestionabilidade de Gudjonsson (GSS 1 e 2)	29
Tabela 2 -	Esquema estratégia de busca	41
Tabela 3 -	Síntese e características dos resultados dos estudos da GSSS.	50

LISTA DE ABREVIATURAS E SIGLAS

GSS	Escala De Sugestionabilidade De Gudjonsson
DEPAJ	Diretoria de Estudos e Pesquisas de Acesso à Justiça
SI	Sugestionabilidade Interrogativa
DPE/RJ	Defensoria Pública do Estado do Rio de Janeiro
AERA	American Educational Research Association, e
APA	American Psychological Association
NCME	National Council on Measurement in Education
CFI	Comparative Fit Index
RMSEA	Root-Mean-Square Error of Approximation
SRMR	Standardized Root Mean Square Residual
AVE	Average Variance Extracted
ICC	Intraclass Correlation Coefficients
EA	Efeitos aleatórios
PROSPERO	International Prospective Register of Systematic Reviews
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

LISTA DE SÍMBOLOS

%	Porcentagem
τ^2	<i>tau squared</i>
α	alfa de <i>Cronbach</i>

SUMÁRIO

1. INTRODUÇÃO	7
2. REFERENCIAL TEÓRICO	9
2.1. PSICOLOGIA E TESTEMUNHO	10
2.2. SUGESTÃO E SUGESTIONABILIDADE	16
2.2.1 Sugestionabilidade interrogativa	18
2.3. ESCALA DE SUGESTIONABILIDADE DE GUDJONSSON	26
2.4. PROPRIEDADES PSICOMÉTRICAS	32
2.4.1 Validade	32
2.4.2 Evidências de validade baseadas no conteúdo	34
2.4.3 Evidências de validade baseada na estrutura interna	35
2.4.4 Evidências de validade baseadas nas relações com medidas exter- nas	36
2.4.5 Evidências de validade baseadas no padrão de resposta	36
2.4.6 Validade Consequencial	37
2.4.7 Fidedignidade ou precisão	37
3. MÉTODO	38
3.1. CRITÉRIOS DE ELEGIBILIDADE	39
3.2. ESTRATÉGIA DE BUSCA	40
3.3. SELEÇÃO DOS ESTUDOS	40
3.4. EXTRAÇÃO DE DADOS	41
3.4. AVALIAÇÃO DO RISCO DE VIÉS	42
3.5. ANÁLISE DE DADOS	43
4. RESULTADOS	46
4.1. CARACTERÍSTICAS DOS ARTIGOS	47

4.2. ANÁLISE DE RISCO DE VIÉS	61
4.3. SÍNTESE DAS EVIDÊNCIAS DAS PROPRIEDADES PSICOMÉTRI- CAS	64
4.3.1 Análise de subgrupo	66
5. DISCUSSÕES	69
6. CONCLUSÕES	86
7. REFERÊNCIAS	90

1. INTRODUÇÃO

O relatório organizado pela Diretoria de Estudos e Pesquisas de Acesso à Justiça (DEPAJ), fez um levantamento de casos (relacionados a 47 processos do Estado do Rio de Janeiro), e apresentou resultados indicando que 80% dos acusados eram negros e 86% tiveram a prisão preventiva decretada, por períodos de privação de liberdade que variaram de cinco dias a três anos, ainda que, ao final, tenham sido absolvidos. Segundo Bernades et al. (2023) cerca de 70% dos casos de condenação injustas no Brasil sejam devido a falso reconhecimento. Os achados por mais que sejam incipientes, podem apontar uma dinâmica semelhante aos casos de relatos de fatos por testemunhas, ambos têm os processos relacionados a memória, emoções, passagem do tempo, vivências passadas e entre outros. Esses cenários podem se agravar ainda mais quando falamos do testemunho infantil, tantas vezes subestimado e questionado, é atualmente um dos principais alvos dos estudos, principalmente no contexto de crimes de violência sexual e separações litigiosas (Souza; Cruz, 2023).

A sugestionabilidade interrogativa (SI) entendida como “o grau em que os indivíduos aceitam e incorporam informação durante um interrogatório formal, afetando a sua resposta posterior” (Gudjonsson, 2003) é uma das principais ameaças à qualidade do testemunho humano e, conseqüentemente, pode contribuir para erros de decisões por parte do tribunal (Gudjonsson, 2003). Nesta direção, Huff et al. (1996, citado em Kękuś *et al.*, 2020) sugeriram que a razão mais comum para a condenação de pessoas inocentes foi o depoimento de testemunhas oculares falsos e não confiáveis. No Brasil, existem poucos dados que se referem ao tema, mas um deles é o Relatório Consolidado sobre Reconhecimento Fotográfico em Sede Policial, construído pela Diretoria de Estudos e Pesquisas de Acesso à Justiça da Defensoria Pública do Estado do Rio de Janeiro (DPE/RJ, 2021).

O sistema judiciário solicita dos profissionais da psicologia meios para embasar suas decisões em dados científicos, essa área da psicologia que tem intersecção com o direito é chamada da Psicologia Jurídica, que abrange desde perícias trabalhistas até psicologia investigativa em cenas de crimes contra a vida. Como subárea específica existe a psicologia do Testemunho que se dedica ao estudo da “relação existente entre o sujeito, a testemunha e o objeto (o acontecimento), que se define por ser um ato singular e irrepetível” (Ferreira, 2018).

Para o profissional que desenvolve seu trabalho nessa área, frequentemente chegam demandas de avaliações psicológicas diversas e uma delas é a de confiabilidade do testemunho. Demanda essa citada por Gisli H. Gudjonsson (1984) em seu artigo “A New Scale Of Interrogative Suggestibility” onde relata que a sua necessidade de desenvolver uma escala que se proponha a medir a sugestionabilidade interrogativa, surgiu diante da demanda de seu trabalho de perito em tribunais na Inglaterra, em que se deparou com uma demanda específica que ele chamou de “credibilidade do testemunho”. Este autor destaca que mesmo diante da escassez de instrumentos que pudessem suprir essa demanda resolveu se aprofundar na temática e com isso conseguiu suprir a necessidade do tribunal. Diante disso, começou os estudos para o desenvolvimento das Gudjonsson Suggestibility Scale - GSS 1 e 2.

As Escalas de Sugestionabilidade de Gudjonsson (Gudjonsson Suggestibility Scale – GSS 1 e 2) são instrumentos amplamente utilizados na avaliação da sugestionabilidade em contextos forenses, clínicos e acadêmicos (Ridley; Gudjonsson, 2013; Clare *et al.*, 1994; Özcan, Ziyalar, 2023; Wachi *et al.*, 2018). Essa medida possui duas versões, a GSS 1 e a GSS 2, a diferença entre as duas escalas é apenas a natureza da narrativa que as constitui, onde na GSS 1 é de conteúdo forense, o que a faz mais indicada para adultos; enquanto a GSS 2, por ter uma narrativa de um acontecimento

mais trivial, sendo mais indicada para crianças. As escalas medem dois aspectos da sugestionabilidade: *Rendimento* (Yields) que é a tendência de um indivíduo de ceder a perguntas enganosas, e *Mudança* (Shift), que é a medida em que as respostas anteriores são alteradas como resultado do feedback negativo. As escalas provaram ter confiabilidade e validade satisfatórias, e muitos correlatos de sugestionabilidade interrogativa foram identificados (Gudjonsson, 1989, 1992).

Assim, observando a escassez de instrumentos nessa área, bem como a frequente utilização da GSS entre pesquisadores e psicólogos da área jurídico/ forense, este estudo teve como objetivo realizar uma revisão sistemática e metanálise das propriedades psicométricas da GSS, investigando, portanto, suas evidências de validade e fidedignidade como medida de sugestionabilidade. Para alcançar esse objetivo foi utilizado o método de Revisão Sistemática PRISMA (Higgins *et al.*, 2022), que permite, entre outros fins agregar evidências a fim de guiar práticas. Para isso foram revisados estudos publicados que utilizaram a escala em diferentes populações e contextos de pesquisa, no final será possível apresentar um vasto panorama de evidências psicométricas deste instrumento, o que possibilita observar a qualidade dos índices e, conseqüentemente, dos dados que o instrumento é capaz de coletar.

Diante do cenário introduzido anteriormente, entende-se que para melhor compreender as questões apresentadas e os objetivos pretendidos, se faz necessário apresentar uma breve revisão da literatura acerca da temática.

2 REVISÃO DE LITERATURA

2.1 PSICOLOGIA E TESTEMUNHO

A Psicologia Jurídica é a área da psicologia que corresponde a toda aplicação do saber psicológico às questões relacionadas ao Direito (Rovinski, 2004). Nesse contexto, os profissionais de psicologia são chamados a transmitir o conhecimento técnico ao juiz, produzindo prova para auxiliá-lo em seu livre convencimento e levar ao processo judicial a documentação técnica sobre o fato em questão.

A psicologia jurídica desempenha um papel crucial no sistema legal, pois pode fornecer uma compreensão aprofundada do comportamento humano, avaliar as competências psicológicas dos indivíduos envolvidos em processos legais, auxiliar no desenvolvimento de júris imparciais, fornece suporte a vítimas e testemunhas, além de contribuir para a redução da reincidência criminal mediante avaliações e intervenções psicológicas adequadas (Trindade, 2014). Assim, a interseção entre a psicologia e o sistema jurídico é uma das maneiras de promover a justiça, a equidade e a eficácia nas decisões legais, aumentando a possibilidade de garantir a proteção dos direitos individuais e a promoção do bem-estar social.

Brito (1993) pontua que as primeiras atuações do Psicólogo Jurídico, no final do século XIX, se relacionavam à Psicologia do Testemunho. Com as solicitações dos psicólogos aos tribunais, objetivava-se verificar, através de estudos experimentais dos processos psicológicos, a fidedignidade do relato, tentando uma compreensão dos processos internos que propiciassem ou dificultassem a veracidade dos relatos de vítimas, agressores e testemunhos.

A Psicologia do Testemunho se dedica, portanto, ao estudo dos processos cognitivos e emocionais envolvidos na recordação e relato de eventos testemunhados. Sua importância reside na contribuição para o sistema de justiça criminal, fornecendo subsídios para avaliação de vários aspectos do testemunho, como credibilidade e precisão dos depoimentos de testemunhas oculares e vítimas. Em suma, a psicologia do

testemunho desempenha um papel crucial no sistema de justiça criminal ao fornecer insights sobre os processos mentais subjacentes aos depoimentos de testemunhas, compreender as limitações e influências da memória humana é essencial para a avaliação adequada dos relatos baseada em evidências científicas (Ferreira, 2018). Nessa área teórica muitos desafios são encontrados, aborda temas como a precisão da memória, sugestionabilidade, influência de questões formuladas, confiabilidade de testemunhos e fatores que podem afetar a acurácia de relatos em contextos legais.

A partir do momento em que sugestionabilidade interrogativa refere-se à “o grau em que os indivíduos aceitam e incorporam informação durante um interrogatório formal, afetando a sua resposta posterior”, entende-se que o modo como as perguntas são formuladas e apresentadas, o estilo do entrevistador (distante, empático, agressivo), a ambiência e outros fatores podem exercer influência sobre a precisão e a confiabilidade dos relatos. Em um estudo Bain e Baxter (2000) avaliaram o efeito de dois estilos de entrevistadores nas medidas de sugestionabilidade interrogativa obtidas pela primeira das Escalas de Sugestibilidade de Gudjonsson (GSS1). Foi levantada a hipótese de que uma atitude geralmente abrupta adotada pelo entrevistador produziria maior distância psicológica e, portanto, pontuações mais altas no GSS1, do que uma atitude amigável. Como resultado, verificaram que os participantes testados na condição ‘abrupta’ obtiveram pontuações mais altas em Mudança e Sugestibilidade Total do que aqueles na condição ‘amigável’.

Assim, ao considerar a SI durante entrevistas de coleta de testemunho, pode ajudar a avaliar a precisão e a confiabilidade dos depoimentos, pois pessoas altamente sugestionáveis podem ser mais suscetíveis a informações externas ou sugestões durante interrogatórios, o que pode levar a distorções ou falsas memórias em seus relatos (Özcan; Ziyalar, 2023; Singh; Gudjonsson, 1988; Richardson; Smith,

1993; Clare *et al.*, 1994). Portanto, medir a sugestionabilidade pode fornecer insights sobre a probabilidade de um testemunho ser influenciado por fatores externos, auxiliando na avaliação da credibilidade das informações fornecidas. Neste sentido, sugere-se que os entrevistadores devem adotar práticas que minimizem sugestões implícitas ou explícitas (Favero *et al.*, 2023; Farrugia; Gabbert, 2019).

Para que um relato seja considerado preciso e fidedigno ele precisa se referir a uma descrição detalhada e exata de um evento ou experiência, sem distorções, omissões ou adições de informações. Esse tipo de relato é caracterizado pela consistência interna, pela ausência de contradições e pela aderência aos fatos observados de forma objetiva e imparcial. Essas características são fundamentais para a tomada de decisões legais justas e baseadas em evidências sólidas. Embora seja desejável que um relato seja o mais preciso e fidedigno possível, é importante reconhecer que a percepção de estímulos, o armazenamento e a evocação da memória humana são suscetíveis a falhas e distorções. Portanto, é improvável que um relato seja totalmente preciso e fidedigno, mas é possível minimizar essas imprecisões por meio de conhecimento sobre os processos relacionados à memória, de práticas adequadas de entrevista, investigação contextuais e análise crítica dos relatos fornecidos (Loftus, 2005; Mandler, 1980; Henriques; Pompeu, 2020).

Visto isso, compreender que a memória é uma das fontes de informação mais importantes durante um interrogatório no processo de investigação, e que ela é constituída por uma combinação de processos de codificação, armazenamento e recuperação, demonstra como o relato de fatos é um sistema complexo de ser compreendido. Além disso, outros processos podem influenciar, como os processos psicológicos (atenção, concentração e outros), emoções vivenciadas, traumas, a interpretação

pessoal, contextos culturais, tempo decorrido desde o evento até a coleta do testemunho, o grau de estresse envolvido, a influência de informações externas e a sugestibilidade da testemunha. Tudo isso pode contribuir para imprecisões ou erros em um relato, como acontece na criação de falsas memórias (Britto de Melo et. al., 2024; Loftus, 2005; Souza; Cruz, 2023).

Datam na primeira metade do século XX os primeiros estudos que faziam referência ao conceito de falsas memórias, tendo como seu autor mais relevante Alfred Binet (1900) que conduziu em experimento inédito com crianças, avaliando os efeitos que as sugestões apresentadas por entrevistadores tinham nesse público. O estudo consistia basicamente em apresentar um conjunto de seis objetos (moeda, etiqueta, botão, retrato, gravura e selo) durante 10 segundos. A memória foi avaliada através de recordação livre (perguntas neutras), perguntas diretivas, questões tendenciosas e enganosas Bjorklund (2000, citado em Taborsa, 2018).

Binet percebeu que raramente as crianças corrigiam as respostas errôneas apresentadas diante de questões tendenciosas ou enganosas, o que fez com que concluísse que eles faziam isso não por falha na memória, mas sim, por se sentirem pressionadas a responder. Isso fez com que o autor postulasse que as informações sugeridas eram agregadas à memória da criança. Binet também foi o primeiro a categorizar como o processo de sugestão pode afetar a memória e conseqüentemente o testemunho: a autosugestão, proveniente de processos internos e a sugestão externa, decorrente da influência por parte de terceiros.

Atualmente, o conceito de falsas memórias, é entendido por Stein e Pergher (2001) como a possibilidade de recordar eventos que de fato não aconteceram, assim, informações equivocadas são registradas na memória e, posteriormente, recordadas

como se tivessem sido verdadeiramente vivenciadas, razão pela qual as falsas memórias não podem ser consideradas “mentiras” daquele que relata, visto que a incorporação é tão forte que a lembrança é considerada verídica por aquele sujeito. Essas memórias, de forma generalista, podem ser divididas em espontâneas (distorções endógenas) ou sugeridas (vindas do ambiente externo), a primeira é fruto dos processos de distorções mnemônicas endógenas, sendo uma tentativa do sujeito de completar os fragmentos mnêmicos naturais dos processos de armazenamento, retenção e recuperação das informações na mente humana. Já a segunda, se caracteriza como uma informação externa que é incorporada à memória como se de fato tivesse acontecido, incorporação esta que é possibilitada pela chamada sugestionabilidade, nomeado a “tendência de se incorporar informações alteradas às suas recordações pessoais, emanadas de fontes externas, intencional ou acidentalmente” (Stein & Pergher, 2001). Atualmente, as pesquisas na área de neurociências estão avançando com relação ao conhecimento dos processos psicológicos e neurológicos mobilizados no processo de falsas memórias, evidenciando e concretizando ainda mais esse fenômeno (Shao *et al.*, 2023; Okado, 2005).

Assim, as falsas memórias sugeridas por Stein (2010) se relacionam com os processos de sugestão e sugestionabilidade, trazendo à tona a necessidade de aprofundamento de pesquisas nessa temática. Conhecer esses processos é essencial para a atuação do profissional psicólogo na área jurídica e forense, pois quando se trata de coleta e avaliação do testemunho é importante conhecer sobre as influências dos processos psicológicos de armazenamento e recuperação de informações e fatos (memória), deve-se, também, ser considerada a passagem do tempo entre o acontecimento e o relato, quais tipos de entrevistas aquele sujeito já foi submetido; sistema

de crenças e/ou expectativas; quantas vezes as memórias do evento foram acessadas, entre outros (Bernstein; Loftus, 2009; Henriques; Pompeu, 2020; Hoscheidt, 2013; Keķus, 2020). Esses fatores, em conjunto com a avaliação psicológica, oferecem dados que podem apontar circunstâncias importantes para esclarecimento de fatos importantes no processo legal (Souza; Cruz, 2023; Stein; Pergher; Feix, 2009). Na perspectiva jurídica, pode-se perceber um movimento de questionamento e melhoramento das técnicas de coleta do testemunho (Tomé; Fayet Júnior, 2017; Brito; Pereira, 2012), e os estudos acadêmicos na área da psicologia jurídica acompanham esse movimento, apesar de ainda acontecer de forma incipiente quando falamos do cenário nacional.

A discussão sobre o testemunho fica ainda mais delicada quando falamos de crimes sexuais contra crianças e adolescentes, em que geralmente a palavra da vítima é a única prova presente no processo para incriminar o réu, julgar esse tipo de processo exige muito de todos os profissionais atuantes. Assim, a produção de provas com maior fidedignidade e confiabilidade torna-se crucial, para alcançar esse objetivo é necessário o aperfeiçoamento das metodologias de escuta e avaliação dos testemunhos (Souza; Cruz, 2023; Stein; Pergher; Feix, 2009; Gava et al.,2013). Nesse contexto, o estudo dos processos de sugestão (como estímulo externo, como ideação ou crença comunicável) e sugestionabilidade, entendida como “dentro de uma interação social fechada, as pessoas passam a aceitar mensagens comunicadas durante questionamentos formais, como resultado da qual, sua resposta comportamental subsequente é afetada” (Gudjonsson e Clark, 1986, p. 84), pode contribuir para o melhor entendimento dos processos que atuam durante a tomada de testemunho em contextos forenses e policiais.

Ser incapaz de resistir à sugestão, implica um risco de testemunhos errôneos e falsos, com consequências potencialmente de longo alcance num processo legal. Por exemplo, Chan et al. (2017) mostraram que ceder às sugestões numa entrevista forense, mesmo quando advertidos contra fazê-lo, alterou a memória da testemunha ocular numa entrevista de acompanhamento, tornando o depoimento menos preciso. Diante do exposto, faz-se oportuno o aprofundamento nos conceitos que envolvem o construto sugestionabilidade, alvo desse estudo.

2.2 SUGESTÃO E SUGESTIONABILIDADE

A sugestão e a sugestionabilidade estão presentes em vários estudos que envolvem memória e testemunhos. Apesar de ainda ser necessário o desenvolvimento de mais pesquisas, esses conceitos podem ser percebidos ativamente no cotidiano das pessoas, desde “conselhos” de familiares e amigos até o bombardeio de sugestões que atingem a vida dos indivíduos por meio das publicidades ostensivas, logaritmos das redes sociais e outros meios de controle usados atualmente (Krasinski; Tonelli, 2018).

A sugestão foi definida como uma forma de ideação ou crença comunicável, que uma vez aceita tem a capacidade de exercer mudanças profundas no humor, pensamentos, percepções e comportamentos (Parris, 2016). A sugestionabilidade, pode ser entendida como a tendência/disposição ou atitude favorável de aderir à sugestão, de modo a acatar e dar forma às sugestões recebidas. Schacter (2003 *apud* Saraiva; Albuquerque, 2015) vai ao encontro dessa definição ao sugerir que a sugestionabilidade é a tendência individual para integrar informações distorcidas nas nossas memórias pessoais. Para uma sugestão ser considerada bem-sucedida é necessário que a pessoa sugestionada não somente acate a sugestão, como também, apropria-

se dela, faça dela sua ação genuína. Sendo assim, a sugestionabilidade é uma variável psicológica responsável por tratar esse estímulo e fazer com que ele acabe por se reverter em resposta (sugestão é um estímulo externo e sugestionabilidade é um processo interno) (Gudjonsson, 2003).

Há muito tempo é proposto que existem dois tipos de influência sugestionável e dois tipos de sugestionabilidade correspondentes a elas: direta e indireta (Eysenck; Furneaux, 1945). A sugestão direta envolve influência aberta e não oculta, enquanto a sugestão indireta diz respeito à influência que está oculta e o participante não sabe que a sugestionabilidade está sendo medida, no entanto, as evidências científicas da existência dos dois fatores têm sido escassas (Polczyk, 2016).

As primeiras referências aos termos “sugestão” e “sugestionabilidade” surgiram com o estudo do fenômeno hipnose, que teve destaque inicialmente na época da Segunda Guerra mundial, como uma das técnicas de intervenção usada no corpo militar. O desmembramento do construto sugestionabilidade do fenômeno hipnose, só aconteceu mais tarde no século XX com Binet (1990) que desenvolveu estudos sobre a sugestionabilidade, percebendo que existiam outras variações, e que a sugestionabilidade existia em outros contextos, não somente no da hipnose, como era mais conhecida anteriormente, entre elas a sugestionabilidade moral (Dias, et.al., 1999), sensorial (Gheorghiu, et.al., 2003) e interrogativa (Gudjonsson, 2003).

Cattell (1895 *apud* Gudjonsson, 2003) também investigou o domínio do testemunho, através de perguntas sugestivas colocadas a estudantes universitários e o grau de confiança que estes tinham nas suas respostas. Mas foi Binet quem mais focou suas pesquisas na sugestionabilidade, principalmente, com o público infantil, como já citado anteriormente, abordando as características individuais que poderiam estar diretamente relacionadas com a presença da sugestionabilidade, concluindo que

alguns sujeitos eram mais propícios que outros em tarefas específicas e, ainda, citou pela primeira vez o termo sugestionabilidade investigativa (SI), foco desse estudo.

Na psicologia contemporânea, um ressurgimento das questões relacionadas com a sugestão e a sugestionabilidade é aparente, principalmente, mas não exclusivamente, nas áreas da psicologia forense (Loftus; Palmer, 1974; Gudjonsson, 1997), na pesquisa sobre hipnose (Halligan; Oakley, 2014; Oakley; Halligan, 2013), na pesquisa sobre efeitos placebo (Colloca; Benedetti, 2009) e ainda ligada a transtornos neurológicos (Wieder *et al.*, 2020). A seguir o estudo da SI será aprofundado em um subtópico exclusivo, visto a necessidade de entendimento do construto para atingir os objetivos dessa revisão.

2.2.1 Sugestionabilidade interrogativa

Durante o estudo da sugestionabilidade interrogativa foram desenvolvidas algumas teorias na tentativa de melhor explicar esse fenômeno, a partir delas, alguns instrumentos de medidas foram criados na expectativa de mensurar de forma confiável esse fenômeno e auxiliar os profissionais na tomada de decisão que atuam diretamente com essas temáticas. Até o momento, existem duas teorias principais da sugestionabilidade interrogativa: a das diferenças individuais e experimentais, respectivamente Schooler e Loftus (1986, 1993, *apud* Gudjonsson, 2003).

A teoria das diferenças individuais baseia-se na análise do modelo psicossocial de sugestionabilidade interrogativa (Gudjonsson; Clark, 1986). Seu foco principal é examinar as diferenças individuais em resposta a perguntas indutoras e perguntas interrogativas sob pressão e os fatores cognitivos, sociais e emocionais associados à sugestionabilidade, assim, medir o efeito da pressão interrogativa através de feedback negativo é a principal característica do modelo proposto.

Em contraste, a abordagem experimental, desenvolvida a partir do trabalho de Elizabeth Loftus (Loftus; Palmer, 1974), concentra-se principalmente nas diferentes condições (por exemplo, atraso de tempo, a semelhança entre o evento e as informações enganosas, a presença de um aviso, a credibilidade da fonte e a plausibilidade de informações enganosas) e mecanismo (detecção de discrepâncias) que influenciam as respostas às questões principais.

Considerando essas perspectivas teóricas, Gísli Hannes Gudjonsson, psicólogo irlandês, desenvolveu sua teoria e descreveu a sugestionabilidade interrogativa (SI) como sendo “o grau em que os indivíduos aceitam e incorporam informação durante um interrogatório formal, afetando a sua resposta posterior” (Gudjonsson, 2003), ou seja, sugestionabilidade interrogativa, nada mais é do que o conceito de sugestionabilidade aplicada ao contexto de interrogatórios, comum na área jurídica e suas várias nuances. Esse autor em seus estudos ainda relacionou a sugestionabilidade ao estágio de desenvolvimento humano, à maturação de traços de personalidade, ao repertório verbal e à desejabilidade social (Gudjonsson, 2003, 1992,1998, 2020; Gudjonsson; Clark, 1986).

É, portanto, considerada um tipo particular de sugestionabilidade, distinguindo-se das demais por apresentar quatro elementos fundamentais diferenciadores: (1) implica um interrogatório numa interação social fechada; (2) contém um componente forte de incerteza; (3) envolve normalmente situações de elevado estresse com consequências importantes para as partes envolvidas; e, por fim, (4) remete para eventos, experiências e conhecimento passado, ao invés de aspectos sensoriais ou motores do momento presente (Gudjonsson, 1984).

A premissa básica deste modelo é que a SI depende das estratégias de enfrentamento que as pessoas podem gerar e implementar quando confrontadas com dois

aspectos importantes no processo de interrogatório – incerteza e expectativas (Gudjonsson, 2003). O modelo Gudjonsson e Clark (1986), estabelece ainda a existência de dois tipos distintos de sugestionabilidade interrogativa: primeiro relacionado com as ideias apresentadas por Binet (1990) que diz respeito ao impacto que as questões sugestivas ou direcionadas têm no testemunho do indivíduo; e outro, que se refere ao grau em que os entrevistadores são capazes de alterar respostas indesejadas (verdadeiras ou não) por meio de feedback negativo.

A partir desse modelo de SI, Gudjonsson criou em 1983, a “Escala de Sugestionabilidade de Gudjonsson, em duas versões (GSS-1 e GSS-2), cuja pretensão é medir de forma objetiva e confiável a sugestionabilidade interrogativa. Diferindo entre elas somente o conteúdo da narrativa e do questionamento, sendo a primeira composta por uma narrativa com conteúdo forense e a segunda não, sendo mais indicada para estudos com crianças a partir dos 7 anos. A escala tem aplicabilidade em contextos forenses, pesquisas e atuações clínicas e no decorrer dos anos vem sendo utilizada em pesquisas para medir sugestionabilidade em geral e não somente interrogativa. Esse instrumento vem tendo seus aspectos psicométricos testados nas últimas três décadas e já foi adaptado para vários países como Itália, Holanda e Portugal (Polczyk, 2005; Pires, et.al., 2014; Merckelbach, et.al., 1998).

Outras escalas também surgiram propondo medidas para SI, como o teste de sugestionabilidade de Würzburg (WST, Bottenberg; Wehner, 1972 *apud* Rowsell; Colloff, 2023), carente de estudos de validade; e o Teste de Sugestionabilidade de Declarações de Bonn (BTSS, Endres, 1997), que foi criado para medir diferenças individuais na sugestionabilidade de crianças de 4 a 10 anos de idade. Esta escala foi desenvolvida em resposta às críticas de que a maioria das escalas de sugestionabilidade focavam no público adulto, sendo consideradas complexas para crianças pequenas.

Adota a abordagem da sugestionabilidade interrogativa, aplicando os mesmos princípios de Rendimento e Mudança à população testemunha infantil (Rowell; Colloff, 2023).

O desenvolvimento desses instrumentos específicos (BTSS, Endres, 1997; WST, Bottenberg; Wehner, 1972 *apud* Rowell; Colloff, 2023; Gudjonsson, 1984) contribuíram para o desenvolvimento dos estudos que conseguiram comprovar que as questões sugestivas conseguem originar respostas distorcidas porque são elaboradas de forma a induzir a resposta desejada, independentemente da sua exatidão. Esses achados desencadearam uma onda de questionamentos sobre metodologias de interrogatórios e coleta de testemunho, indicando que essas técnicas poderiam induzir respostas ou modificar relatos (Cecconello; Milne; Stein, 2022). Outro fator importante encontrado foi o grau com que os entrevistadores poderiam influenciar na resposta dos sujeitos, como é demonstrado no instrumento por meio do Feedback negativo apresentado durante o procedimento de aplicação (Gudjonsson, 2003).

A este respeito, Linton e Sheehan (1994) realizaram um estudo usando uma versão modificada da GSS para verificar achados de pesquisas anteriores que dizem que mesmo quando nenhum feedback negativo foi apresentado, os sujeitos alteravam algumas de suas respostas quando interrogados novamente. Uma razão provável para esta descoberta é que repetidos questionamentos ou feedback neutro podem atuar como uma forma de feedback negativo implícito, ou seja, os sujeitos assumem que cometeram erros e, por isso, estão sendo interrogados novamente. Outras pesquisas precisam ser realizadas para estabelecer se as respostas a perguntas repetidas são independentes das respostas ao feedback negativo. Outra possível explicação para os efeitos de repetidos questionamentos é que alguns sujeitos podem não

se lembrar de como responderam a todas as perguntas anteriores, portanto, fornecem respostas diferentes.

A sugestionabilidade ainda pode ser agravada em situações em que o relato é obtido por meio de pressão ou coerção, como no contexto de interrogatório policial, entrevistas coercitivas são práticas frequentes e podem induzir falsas confissões. Tais confissões falsas têm desempenhado um papel proeminente em condenações injustas, como ilustrado, por exemplo, pelo programa chamado “Innocence Project” (<http://www.innocenceproject.org>) presente em vários países em que 20-25% dos casos atendidos por eles são de falsas confissões induzidas pela polícia.

O estudo de Kassin e Wrightsman (1985) diferencia entre três tipos de falsas confissões, são elas: primeiro tipo são as falsas confissões voluntárias, que surgem sem pressão externa durante um interrogatório policial, ou seja, a motivação para confessar voluntariamente são, por exemplo, para proteger outra pessoa ou para ganhar fama. O segundo tipo são as falsas confissões coagidas, que ocorrem devido à pressão policial e técnicas de interrogatório coercitivo. Nesta categoria, os confessores ainda sabem que são inocentes, mas muitas vezes só querem sair da situação desconfortável do interrogatório. O terceiro tipo consiste em confissões internalizadas coagidas que ocorrem quando suspeitos inocentes começam a acreditar erroneamente em sua culpa devido à pressão policial e à natureza sugestiva do interrogatório. Esta última estaria intimamente ligada ao nível de sugestionabilidade do interrogado (Otgaar *et. al*, 2020)

Essa situação se agrava ainda mais quando os envolvidos nos processos judiciais são pessoas em vulnerabilidade psicológica como pessoas com transtornos mentais e/ou pessoas com deficiência intelectual (Farrugia; Gabbert, 2019). Portanto,

é fundamental que os profissionais envolvidos na coleta de depoimentos tenham treinamento adequado para minimizar esses efeitos indesejáveis (Favero *et.al.*,2023). Assim é evidente a necessidade de aprofundamento nos mecanismos de funcionamento da sugestionabilidade enquanto construto psicológico e outras variáveis relacionadas.

Com esse objetivo muitos estudos investigaram variáveis relacionadas a SI com o objetivo de chegar a fatores fortemente relacionados ou quem sabe preditores do fenômeno. Entre as principais variáveis estão: inteligência (Gudjonsson; Rutter; Clare, 1995; Singh; Gudjonsson, 1992; Clare *et.al*, 1994; Young; Powell; Dudgeo., 2003), idade (Saraiva; Albuquerque, 2015; Cruz; Pinho, 2014; Gudjonsson, *et.al.*, 2016; Rowsell; Colloff, 2024; Miles, Clare; Bull, 2002), memória (Cunha, *et.al.*, 2007; Raver, Lindholm; Gustafsson, 2023; Saraiva; Albuquerque, 2015) e personalidade (Haraldsson, 1985; Gorassini, 2006). Pode-se citar ainda variáveis menos estudadas como desejabilidade social (Krasinski; Tonelli, 2018), coping (Souza Filho, 2018), ansiedade (Gudjonsson, 1988; Gudjonsson *et al.* 1995; Santtila; Ekholm; Niemi,1999; Polczyk, 2005) e conformidade (Otgaar, *et. al.* 2020; Hansen; Smeets; Jellicic, 2010).

Entre essas variáveis, as diferenças relacionadas à idade inicialmente começaram a ser mais pesquisadas, devido ao questionamento diante dos testemunhos infantis, porém, até o momento não foram encontradas evidências que sustentem essas contestações, se seguidas as boas práticas de escuta infantil. Em sua revisão Klemfuss e Olaguez (2018) identificaram o comprometimento intelectual como uma das diferenças individuais mais confiáveis para prevê a sugestionabilidade em crianças. O comprometimento intelectual pode ser considerado um fator de risco quando essas crianças são comparadas a seus pares cronológicos, ou seja, da mesma idade. Já

quando a avaliação foi realizada comparando as crianças com comprometimento individual e seus pares com idade corrigida, as diferenças de escores não foram significativas.

Outro achado com evidências mais confiáveis são os que ligam a maior vulnerabilidade a sugestão dos idosos ao declive de memória que podem surgir com a idade (Biondi, et.al. 2020). Junto as questões relacionadas à memória, este estudo verificou que os participantes idosos pontuaram mais baixo em medidas de memória e inteligência e mostraram um desempenho ruim em teste de SI.

Sobre a relação entre sugestionabilidade e níveis de inteligência, Bianco e Cuci (2015) tiveram como um dos objetivos estabelecer a validade discriminante da GSS 1 e GSS 2, para isso as escalas foram administradas a duas amostras (50 e 52 participantes) junto com medidas cognitivas, de personalidade e afetivas (como enfrentamento e ansiedade). Os instrumentos aplicados foram: três subtestes (Vocabulário, Compreensão e Analogias) da Escala de Inteligência de Adultos de Wechsler – Revisada (WAIS-R) e a subescala de Inteligência Fluida das Matrizes Progressivas Padrão (SPM). Nesse mesmo estudo ainda foram correlacionadas memória por meio do Teste de Rey e a personalidade empregando a forma abreviada do Questionário de Personalidade de Eysenck – Revisado (EPQ-R).

Os resultados demonstraram correlações moderadas e significativas entre os indicadores de recordação de memória, Vocabulário e Analogias (subescalas do WAIS-R) mostrando que as recordações imediatas e tardias refletem as habilidades verbais de um indivíduo. Além disso, correlações moderadas e significativas surgiram entre índices de memória de curto e longo prazo medidos pelo teste de Rey e os indicadores de memória do GSS. De fato, as escalas de recordação imediata e tardia podem ser consideradas como medidas de desempenho de memória (Gudjonsson, 1997).

Além disso, os dados não mostraram correlações significativas com habilidades verbais, memória e inteligência. Esse dado, segundo as autoras reflete o fato de que os participantes apresentam habilidades cognitivas médias, o que segundo Gudjonson (1988), a associação entre medidas de inteligência e sugestibilidade foi mais claramente observada em indivíduos com QI abaixo da média, não sendo significativa em sujeitos com habilidades na média ou alta.

Outro fator interessante de ser pontuado é a relação entre sexo e a sugestibilidade que nos primórdios teorizava-se sobre a existência de diferenças, mas com o desenvolvimento das pesquisas foi concluído que elas não são significativas. Um exemplo disso é o estudo de Wachi et al. (2019) que comparou a pontuação de pessoas do sexo masculino e feminino em um $n=442$, os escores de sugestibilidade não indicaram diferenças de gênero, porém, encontraram o desempenho elevado nas subescalas de recordação das participantes femininas, isso pode sugerir que elas são mais propensas a fornecer respostas corretas às perguntas não sugestivas do que os participantes do sexo masculino. Outro estudo que aborda sexo como variável é o de Hein et al. (2021), o qual a partir de uma análise de variância fatorial também não encontraram diferenças entre os participantes ($n=560$) de diferentes sexos, mas em pontuações de recordação livre, participantes do sexo feminino pontuaram mais alto do que os do sexo masculino.

Sobre a ansiedade, variável que faz parte da teoria base da GSS formulada por Gudjonson, este autor indica que a ansiedade é uma das variáveis que pode influenciar na sugestibilidade, por exemplo, em sua publicação "The Psychology of Interrogations and Confessions" ele conclui "Shift é a medida de como as pessoas respondem à pressão interrogativa, o que a vincula particularmente à ansiedade e aos processos de enfrentamento" (2003. p.413, tradução nossa), assim, o shift enquanto fator

seria influenciado pelo processo de ansiedade. Em outro artigo, Gudjonsson (1988) verificou que a ansiedade de estado se correlacionou significativamente com a sugestionabilidade, particularmente após o feedback negativo ter sido administrado. Porém, nem todos os pesquisadores conseguem confirmar essa correlação, a exemplo, Polczyk (2005) que teve como resultado mais intrigante a falta de correlação entre ansiedade e sugestionabilidade interrogativa, porque tal relação foi observada na literatura e faz sentido do ponto de vista teórico. Já a hipótese sobre a relação entre inteligência geral, memória e sugestionabilidade interrogativa foi confirmada. Por ser um dos instrumentos mais utilizados enquanto medida objetiva de sugestionabilidade faz-se necessário aprofundar sobre o contexto em que foi criada e seus procedimentos de aplicação e pontuação.

2.3 ESCALA DE SUGESTIONABILIDADE DE GUDJONSSON

A Escala de Sugestionabilidade de Gudjonsson (Gudjonsson Suggestibility Scales GSS), foi desenvolvida em 1983 por um psicólogo irlandês chamado Gísli Hannes Gudjonsson com o objetivo de apresentar medidas objetivas à vulnerabilidade de pessoas a dar relatos errôneos quando interrogados em conteúdos forenses. Existem duas versões dessa escala, a GSS 1 e a GSS 2 diferindo entre elas somente o conteúdo da narrativa e do questionamento, sendo a primeira um conteúdo forense e a segunda não, sendo mais indicada para estudos com crianças a partir dos 7 anos. Elas são particularmente aplicáveis a contextos de entrevista policial, mas podem ser aplicadas de forma útil a qualquer situação de entrevista, incluindo pesquisas e prática clínica. Podem ser utilizadas com populações forenses, vítimas e testemunhas de crimes, pessoas com transtornos mentais e crianças e adolescentes.

Esse instrumento é composto por uma narrativa, apresentada oralmente ou tocado de um gravador, seguida de uma tarefa de recordação imediata, uma de

recordação atrasada e uma de questionamento. Inicialmente é apresentado ao sujeito como um teste de memória expondo uma narrativa contendo 40 fatos. Em seguida, é solicitado ao sujeito que reproduza em formato de relato livre, o máximo de informações possível (caracterizando a tarefa de recordação imediata) esse relato dá origem a um escore pontuado de 0 a 40. Após 50 minutos é realizado o “delayed recall” (recordação atrasada), e mais uma vez é solicitada a recuperação do parágrafo apresentado. Após esse momento o questionário, contendo 20 (vinte) questões no estilo dicotômico (sim/não), das quais 15 (quinze) são enganosas. Dando continuidade ao processo, o feedback negativo é apresentado, em que o indivíduo é informado de que cometeu erros (mesmo que nenhum erro tenha sido cometido) e que é necessário responder às perguntas novamente (o mesmo questionário aplicado anteriormente). Essa fase dá origem a escores pontuados de 0 a 15 (yield/rendimento e shift/mudança) (Gudjonsson, 1992; Merckelbach. *et al.*, 1998; Gudjonsson, 2020).

A GSS 1 e 2 fornecem medidas de recordação da memória verbal, tanto imediata quanto atrasada, e de sugestibilidade interrogativa, que se mostrou dividida em dois componentes. Os dois primeiros o autor denominou de “Yield” (Rendimento/Produção), que diz respeito ao número de respostas erradas que o indivíduo apresenta durante o primeiro questionamento, e o “Shift” (Mudança), que diz respeito ao número de respostas modificadas após o feedback negativo. Rendimento e Mudança são normalmente somados e são referidos como “Sugestibilidade Total”. A GSS 1 e a GSS 2 também apresentam dados clínicos por meio de medidas objetivas que o autor chamou de “Confabulation” (confabulação), caracterizando-se como a soma de dois componentes: 1) “Fabrications” (fabricações) que acontecem quando

uma informação é adicionada à história e 2) “Distortions” (distorções), quando o conteúdo da história é modificado de alguma maneira, ambos refletem a presença de erros de memória (Merckelbach, *et al.*, 1998; Gudjonsson, 2020).

A escala tem a vantagem de medir o impacto de ambas as questões sugestivas (Yield /Rendimento) e pressão interpessoal (Shift/Mudança). Os dois tipos de suggestionabilidade parecem ser bastante independentes na medida em que são fracamente correlacionados e carregam fatores separados. Os dois tipos de suggestionabilidade, Rendimento e Mudança, podem levar a diferentes inferências e práticas. Assim, conhecer os tipos de perguntas sugestivas às quais os sujeitos são particularmente suscetíveis e até que ponto pode ser induzidos ao erro, podem fornecer informações práticas úteis sobre a potencial de confiabilidade do depoimento das testemunhas. Por sua vez, a aplicação de feedback crítico (feedback negativo) que são as pontuações resultantes de rendimento (rendimento 2) e deslocamento, por outro lado, representam pressão relativamente maior à suggestionabilidade e pode, portanto, estar mais ligada à ansiedade e aos processos de enfrentamento (Gudjonsson, 2003).

A maior parte das pesquisas com o GSS 1 e GSS 2 têm se preocupado com dois tipos de informações que podem ser facilmente derivadas das escalas. Estes correspondem a Yield 1 e Shift. Porém, rendimento 2, memória (imediata e atrasada) e a confabulação (distorções e fabricações) também podem ser medidas para fins clínicos e de pesquisa. Para melhor compreensão, o quadro abaixo, apresenta as formas de medida possíveis durante o uso das GSS 1 e 2:

Tabela 1 - Esquema com os tipos de medidas possíveis com o uso das Escala de Sugestionabilidade de Gudjonsson (GSS 1 e 2)

1. Recordação imediata (<i>Immediate recall</i>)	Mede a recordação verbal imediata da narrativa do GSS 1 e 2 apresentando uma indicação do nível de atenção, concentração e capacidade de memória. O número máximo de “ideias” que os sujeitos podem recordar é 40.
2. Recuperação atrasada (<i>Delayed recall</i>)	A recordação tardia da narrativa no GSS 1 e 2 é geralmente obtida cerca de 50 minutos após a recordação imediata. Normalmente se deteriora em cerca de um ou dois pontos durante um período de 50 minutos.
3. Rendimento 1 (<i>Yield 1</i>)	Refere-se ao número de sugestões que o sujeito se rende na GSS 1 e GSS 2 antes do feedback negativo.
4. Mudança (<i>Shift</i>)	Refere-se ao número de vezes em que houve uma mudança distinta nas respostas do sujeito após feedback negativo
5. Rendimento 2 (<i>Yield 2</i>)	Refere-se ao número de questões principais que o sujeito cede após o feedback negativo ter sido administrado. Portanto, o rendimento 2 representa o número de sugestões aceitas após a interrogativa pressão. O rendimento 2 indica o tipo de mudança que ocorreu como resultado do feedback negativo.
6. Sugestibilidade total (<i>Total Suggestibility</i>)	Esta é a soma de Yield 1 e Shift. Isto dá uma indicação do nível geral de sugestionabilidade do sujeito.
7. Confabulação (<i>Confabulation</i>)	A confabulação refere-se a problemas no processamento de memória onde as pessoas substituem lacunas em sua memória com experiências imaginárias que eles acreditam ser verdadeiro. Dividida em: distorções, que representam uma grande mudança nos deta-

lhes de um sistema existente “ideia”; e fabricações, onde um elemento novo ou inovador é adicionado à narrativa. Distorções e fabricações são normalmente adicionadas para perfazer o número total de confabulações.

Fonte: Gudjonsson, (2003).

Gudjonsson (1984) apresentou as propriedades psicométricas da GSS, a partir da realização de Análise Fatorial das pontuações derivadas das 20 questões do Rendimento 1. Posteriormente, dois fatores foram testados usando o procedimento Vari-max. Segundo o pesquisador, os itens Yield (Rendimento 1) e Shift (Mudança) são claramente carregados em fatores separados, sugerindo uma estrutura bifatorial para a GSS. Os coeficientes Alfa para as pontuações dos 15 itens Yield (rendimento) e 15 itens Shift (mudança) foram 0,77 e 0,67, respectivamente.

Posteriormente, Gudjonsson (1992) apresentou dados psicométricos da GSS 2, indicando ser uma forma paralela da GSS1. As respostas às 20 perguntas (“Yield/Rendimento 1”) e as pontuações “Shift/Mudança” a estas perguntas após feedback negativo foram analisadas por meio de análise fatorial, utilizando o procedimento Varimax. Os coeficientes alfas foram 0,87, 0,90 e 0,79, para “Yield/Rendimento I”, “Yield/Rendimento 2” e “Shift/Mudança”, respectivamente. Estes resultados indicam uma melhora na consistência interna identificada na GSS 1, onde os coeficientes alfas foram modestos (Gudjonsson, 1984). Embora os itens “Yield/Rendimento” e “Shift/Mudança” sejam carregados em fatores separados, os dois fatores estão significativamente correlacionados. A correlação com “Shift” é maior com “Rendimento 2” ($r = 0,57$, $P < 0,001$) do que com “Rendimento 1” ($r = 0,36$, $P < 0,001$),

Estudos psicométricos destas escalas com a população portuguesa (Pires, 2011; Pires *et al.*, 2014) apresentaram índice de fidedignidade (Alfa de Cronbach),

semelhante àquela encontrada por Gudjonsson (1984), especificamente no que se refere à Rendimento 1 (0,74) e Rendimento 2 (0,76). A avaliação da estabilidade temporal com recurso à aplicação da GSS-1 e GSS-2 à mesma amostra, indicou valores satisfatórios para todas as medidas (Yield 1 = 0,39; Yield 2 = 0,46; Sugestionabilidade Total = 0,32), à exceção da Mudança (= 0,11) (Pires *et al.*, 2014).

A escala tem sido utilizada em pesquisas com o objetivo de conhecer mais profundamente as nuances da sugestionabilidade e seus efeitos práticos. Como na pesquisa de Ridley e Gudjonsson (2013), que utilizaram uma amostra com 180 participantes, 90 dos quais tinham histórico de suspeita de abuso sexual e 90 “controles” com quociente de inteligência (QI) compatível. O grupo de vítimas obteve pontuação significativamente mais alta do que o grupo controle em Shift (Mudança) ($d = 0,72$), Yield 2 (Rendimento 2) ($d = 1,16$) e Sugestionabilidade Total ($d = 0,53$), mas inferior na recordação tardia ($d = 0,95$) no acompanhamento de 1 semana. Segundo os autores, os resultados podem sugerir a existência de uma história de abuso sexual significativamente relacionada à redução da capacidade de lidar com a pressão interrogativa em termos de feedback negativo, mas não com Rendimento 1 ou sugestionabilidade retardada. Isso sugere a importância da avaliação socioemocional, e não somente de fatores cognitivos.

Os autores ainda conceituam à sugestionabilidade imediata, como sendo o ato do entrevistado concordar ou aceitar a informação enganosa no momento do interrogatório (Gudjonsson, 2003, 2018), mas pode não se tornar incorporado em sua lembrança subsequente ao longo do tempo. Em contraste, a sugestionabilidade tardia refere-se à medida em que a informação falsa é incorporada à memória da testemunha (Ridley & Gudjonsson, 2013). Assim, abuso infantil por si só não leva a um au-

mento da sugestionabilidade, mas pode ser é a presença de sintomas psicopatológicos associados ao trauma que aumenta o risco de inserção de erros na memória (Godman, Quas, & Ogle, 2009). Os sintomas do trauma podem prejudicar a capacidade das pessoas de julgar a fonte de informação e formar ligação de memória satisfatória (ou seja, ser capaz de segregar o conteúdo de um evento observado a partir de informações enganosas subsequentes). Concluíram os autores que existe uma correlação entre a intensidade da sintomatologia traumática e a sugestionabilidade, mas que é necessário a realização de novas pesquisas para conhecer mais profundamente essa relação (Vagni, et.al. 2020)

Diante do exposto, o presente estudo se propõe a realizar uma revisão sistemática com metanálise das propriedades psicométricas da Escala de Sugestionabilidade de Gudjonsson em suas duas versões (GSS 1 e GSS 2), para tanto, se faz necessário apresentar uma breve descrição atual das propriedades psicométricas que serão aqui investigadas.

2.4 PROPRIEDADES PSICOMÉTRICAS

Ao realizar pesquisas científicas e atividades práticas em avaliação psicológica é essencial contar com instrumentos de medição confiáveis e válidos e a avaliação das propriedades psicométricas são critérios fundamentais para garantir a credibilidade dos resultados obtidos (CFP, 2022). A qualidade das propriedades psicométricas de um instrumento é fundamental para garantir sua utilidade e credibilidade. Instrumentos com propriedades psicométricas robustas são mais confiáveis na obtenção de resultados precisos e válidos, permitindo que pesquisadores e profissionais tomem decisões com bases científicas. contribuindo para a construção de um corpo de conhecimento sólido e confiável. (Hutz, 2015).

No contexto brasileiro, a qualidade dos instrumentos psicológicos que serão utilizados na prática profissional é verificada pela Comissão Consultiva em Avaliação Psicológica, vinculada ao Conselho Federal de Psicologia (CFP), por meio do Sistema de Avaliação de Testes Psicológicos (SATEPSI). Neste documento as diretrizes referentes à validade, precisão e normatização dos instrumentos psicométricos são analisados, podendo ser aprovados ou não para uso na prática profissional. Aqueles que utilizarem em suas avaliações instrumentos não reconhecidos por esse sistema podem ser enquadrados em grave falta ética (CFP, 2022).

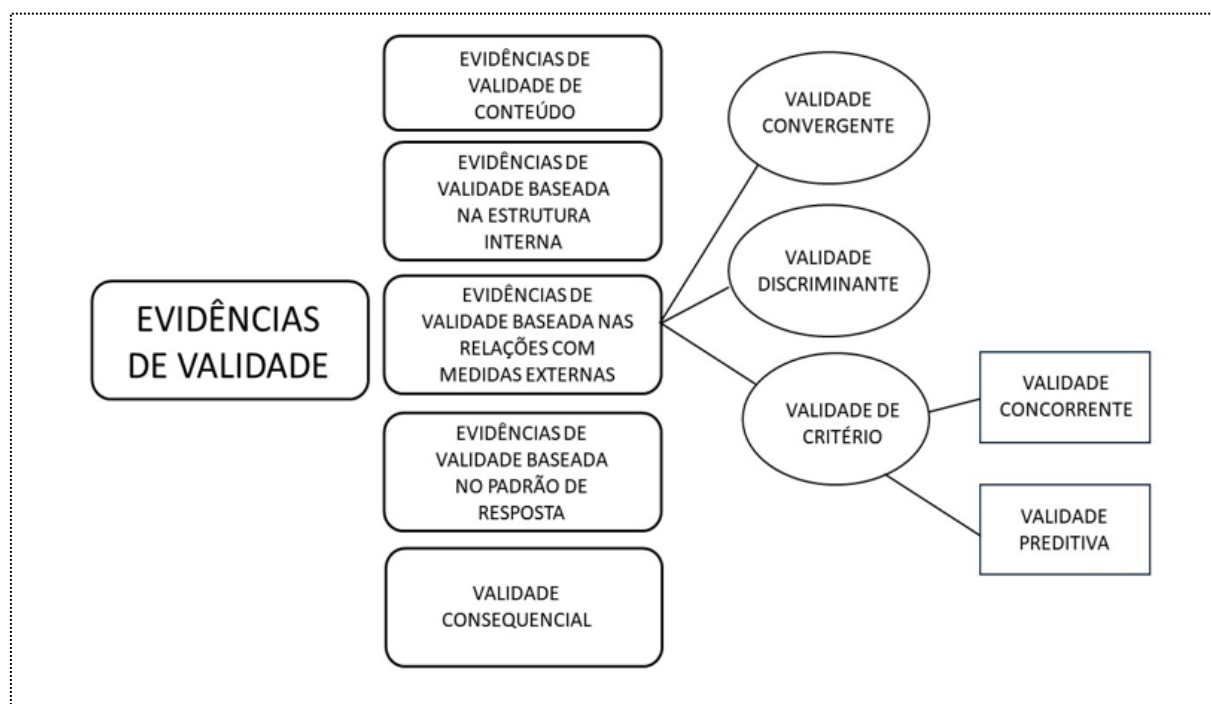
2.4.1 Validade

A validade é uma propriedade psicométrica crucial para indicar a qualidade de um instrumento, sua definição clássica indicava que a validade “se refere ao grau que o teste mede aquilo que se propõe a medir” Buckingham (1921, citado por Pacico, et.al., 2015), ou seja, o teste tem validade quando os itens realmente representam com veracidade a expressão do traço latente (Pasquali, 2013). No processo histórico de desenvolvimento desse conceito surgiram dois modelos principais, são eles: o modelo tripartite e o modelo atual. O primeiro diferencia três tipos: validade de conteúdo, que se refere à representatividade das questões ou itens do instrumento em relação ao construto em estudo; a validade de critério, que verifica se as medidas do instrumento estão correlacionadas com outras variáveis externas relevantes e a validade de construto, por sua vez, avalia em que medida a estrutura está sendo bem representada empiricamente.

Já o modelo atual é descrito nos Padrões de Testagem Psicológica e Educacional por três agências de referência na área *American Educational Research Association* (AERA), *American Psychological Association* (APA) e *National Council on Measurement in Education* (NCME) (2014). Neste modelo, a validade é conceituada como

“o grau em que todas as evidências acumuladas corroboram a interpretação pretendida dos escores do teste”, neste sentido, a validade pode ser entendida como uma propriedade dos escores gerados pelo teste em um determinado contexto e não do instrumento em si. Além disso, as evidências de validade são classificadas como: 1) Evidências de validade de conteúdo; 2) Evidências de validade baseada na estrutura interna; 3) Evidências de validade baseada nas relações com medidas externas; 4) Evidências de validade baseada no padrão de resposta; e 5) Validade consequencial. (AERA, APA & NCME, 2014). Ressalta-se que todas as evidências de validade e confiabilidade acumuladas corroboraram para a interpretação pretendida dos escores do instrumento (Damásio & Borsa, 2017).

Figura 01 – Tipos de evidências de validade.



Fonte: AERA, APA & NCME, (2014)

2.4.2 Evidências de validade baseadas no conteúdo

Aqui o objetivo é avaliar o quão bem o instrumento avalia o construto, com foco em aspectos gramaticais, semânticos e linguísticos dos itens atentando para que esses estejam construídos de maneira acessível ao público-alvo; verifica também se o conjunto de itens de fato faz parte do construto de interesse e se ele está sendo adequadamente mensurado. Essa evidência é especialmente importante quando se trata de um processo de construção e/ou adaptação de um instrumento, erros no processo de verificação da validade de conteúdo pode comprometer todas as outras medidas do instrumento. A avaliação pode ser feita por juízes experts no construto que está sendo estudado ou ainda em psicometria e pela população alvo (ninguém melhor para avaliar se de fato os itens estão sendo entendidos e interpretados de maneira devida), por meio de fichas avaliativas criadas para cada construto (AERA, APA & NCME, 2014; Pasquali, 2012, Damásio & Borsa, 2017).

2.4.3. Evidências de validade baseada na estrutura interna

Esta avaliação visa observar se a estrutura empírica encontrada reflete corretamente a estrutura teórica do construto alvo. Quando um instrumento é desenvolvido a intenção é que ele consiga abordar todos os fatores que estruturam o construto que se deseja medir, por exemplo, quando o instrumento se propõe a medir a sugestibilidade, o pesquisador precisa escolher, entre as correntes teóricas, qual ele deseja se embasar, no caso da escala em questão a teoria é a desenvolvida pelo próprio autor que divide a sugestibilidade interrogativa em dois fatores 1) a extensão que os sujeitos cedem a perguntas sugestivas; (2) até que ponto os sujeitos podem ser obrigados a mudar suas respostas uma vez que a pressão interpessoal foi aplicada.

Visto isso, o instrumento precisará contemplar os dois fatores citados teoricamente, no caso da GSS Rendimento e Mudança.

A principal técnica padrão para avaliar a validade interna é a Análise Fatorial, entendida como um conjunto de técnicas de análise de dados que pretende reduzir os itens a um algumas variáveis latentes, explicando assim, a covariância entre os itens. As variáveis latentes não podem ser observadas (característica de grande parte dos fenômenos psicológicos) elas são inferidas a partir das relações com variáveis observáveis descritas pelos itens. Assim, a análise fatorial permitirá agrupar os itens que correspondem a inteligência linguística, por exemplo. Existem dois tipos de análise fatorial: a exploratória (AFE), em que a estrutura fatorial é encontrada por meio da correlação entre os itens; e a fatorial confirmatória (AFC) em que o pesquisador fornece uma estrutura fatorial pré-determinada com o objetivo de testá-la (Damásio & Borsa, 2017; Pacico, et.al., 2015). Achada a estrutura fatorial, por meio da AFE é possível avaliar a sua qualidade por meio de índices de adequação de ajustes e das medidas de confiabilidade.

2.4.4. Evidências de validade baseada nas relações com medidas externas

Aqui é avaliada se os escores encontrados pelo instrumento se associam de maneira descrita na teoria com medidas externas. Eles se dividem em três grupos: validade convergente (são utilizados instrumentos que mensuram o construto semelhantes que se associam de determinada maneira citada na literatura), um exemplo disse é a ansiedade e estresse que se associam positivamente; validade discriminante é utilizado instrumentos que mensuram construtos opostos que não tem relações teóricas e conseqüentemente não tem relações empíricas, como por exemplo, atenção dirigida e qualidade de vida; validade de critério que se subdivide em concorrente (mesma variável avaliada por dois instrumentos distintos) e preditiva (entende que o

grau de desempenho em um teste pode prever o resultado de em outro instrumento ou situação), como exemplo, pessoas que tem altos escores em inteligência provavelmente terão um melhor desempenho acadêmico, mais utilizado em estudos longitudinais, pois subentende-se a passagem do tempo (Pacico *et al.*, 2015; Pasquali, 2019).

2.4.5 Evidências de validade baseada no padrão de resposta

Avalia aspectos teóricos e empíricos sobre a forma como os participantes respondem os itens e os processos cognitivos envolvidos durante o procedimento de resposta. O próprio manual dos Standards for Educational and Psychological Testing (AERA, APA & NCME, 2014) não especificam classificações e técnicas específicas para essa avaliação, no entanto alguns autores citam técnicas como: grupo focal; rastreamento ocular e contagem do tempo de resposta.

2.4.6. Validade consequencial

Aqui os pesquisadores buscam avaliar as consequências dos testes e dos procedimentos de aplicação na população, já que principalmente na utilização em contextos clínicos e de avaliação psicológica em contextos variados a aplicação dos testes tem consequências práticas como diagnósticos de transtornos, licenças de motoristas, portes de armas e outros. Essa validade é muito pouco pesquisada, mas esse fato não tira sua importância inclusive para pensar questões éticas relacionadas a essa temática (AERA, APA & NCME, 2014).

3. Fidedignidade ou precisão

A avaliação mais utilizada é a medida com confiabilidade também chamada de fidedignidade ou precisão, tem como objetivo mensurar os erros de mensuração, identifica qualquer flutuação nos escores que podem ser resultantes de processo de medição (Zanon & Hauck Filho, 2015). Para analisar os dados de confiabilidade as formas mais utilizadas são a consistência interna e o teste-reteste. A primeira consiste

em conhecer em que grau o padrão de resposta aos itens é preciso baseado nas medidas de correlação entre os itens, o método mais utilizado para esse fim é o Alfa de Cronbach, altamente difundido entre os pesquisadores e softwares de medidas, apesar de existirem técnicas mais modernas e eficientes que levam em consideração as cargas fatoriais (o que o alfa de Cronbach negligencia), como as medidas de confiabilidade compostas, o Ômega de McDonald, entre outros (Damásio & Borba, 2017).

O segundo diz respeito a reaplicação dos testes depois de um determinado intervalo de tempo (para estudos longitudinais) ou coletas duplicadas para testar a precisão (visto em pesquisas transversais). Segundo Zanon e Hauck Filho (2015) se constitui como método mais intuitivo para avaliação da fidedignidade no decorrer de um período de tempo. Após a reaplicação é realizada a correlação entre os escores nos diferentes momentos, essa correlação é uma análise estatística entre duas ou mais variáveis que produz um coeficiente. O período de tempo parece desempenhar um papel importante sobre os escores, já que se o teste for reaplicado em um período de tempo mais curto tende a apresentar maiores coeficientes de correlações, se o período entre as aplicações formais for longo, os coeficientes tendem a diminuir.

Neste sentido, observa-se uma preocupação da Psicologia e áreas afins na pesquisa e na avaliação da sugestionabilidade como fator componente dos testemunhos, necessitando de instrumentos precisos que apresentem propriedades psicométricas aceitáveis (Alexandre & Coluci, 2011). O que justifica a realização desse estudo, o qual tem por objetivo principal realizar uma Revisão Sistemática e Meta-análise da Escada de Sugestionabilidade de Gudjonsson (Gudjonsson Suggestibility Scales GSS), a fim, de verificar suas propriedades psicométricas.

Essa pesquisa justifica-se pelo fato de que o uso de instrumento com a qualidade baseada em evidências científicas e com propriedades psicométricas robustas

é essencial para o desenvolvimento de boas práticas no campo da psicologia. Junta-se a isso, o fato de que o estudo da sugestionabilidade interrogativa reveste-se de importância já que a escala é utilizada em outros países como parte integrante nas avaliações, se faz necessário entender a robustez das propriedades psicométrica da GSS descritas até o momento e o quanto as escalas podem contribuir para o entendimento de aspectos relacionados a vulnerabilidade dos sujeitos a sugestionabilidade investigativa no contexto de investigações forenses e testemunhos adultos e infantis. Assim, este estudo teve como objetivo realizar uma revisão sistemática e metanálise das propriedades psicométricas da GSS, investigando, portanto, suas evidências de validade e fidedignidade como medida de sugestionabilidade

3 MÉTODO

O presente projeto se caracteriza com uma pesquisa teórica, secundária de Revisão Sistemática que pretende verificar as propriedades psicométricas da Escala de Sugestionabilidade de Gudjonsson (*Gudjonsson Suggestibility Scales – GSS 1 e 2*) que objetiva medir a sugestionabilidade interrogativa. A Revisão Sistemática é uma forma de pesquisa que utiliza como fonte de dados a literatura sobre determinado tema, disponibilizando um resumo das evidências, mediante a aplicação de métodos explícitos e sistematizados de busca, apreciação crítica e síntese da informação selecionada (Sampaio & Mancini, 2006).

Para alcançar o referido objetivo, foi adotada a estratégia de revisão sistemática e Metanálise da literatura seguindo o método PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), que apresenta um procedimento metodológico composto por várias etapas: 1) Produzir um protocolo de investigação e efetuar o seu registro; 2) Definir os critérios de inclusão e exclusão; 3) Desenvolver uma estratégia de busca; 4) Selecionar os estudos; 5) Avaliar o risco de viés dos estudos que

serão incluídos; 6) Extrair os dados; 7) Sintetizar os dados e avaliar a certeza da evidência; 8) Apresentar os resultados. O protocolo desta revisão sistemática foi registrado no *International Prospective Register of Systematic Reviews* (PROSPERO) sob o número CRD42023464581 (Chien *et al.*, 2012).

Critérios de elegibilidade

Para ser incluído nessa revisão sistemática o estudo precisou: apresentar as propriedades psicométricas das escalas GSS 1 e 2 (evidências baseadas no conteúdo, evidências baseadas no processo de resposta ao item, evidências baseadas na estrutura interna, evidências baseadas nas relações com variáveis externas e fidedignidade ou confiabilidade da medida); serem estudos originais de pesquisa envolvendo seres humanos e estarem indexados nas bases de dados eletrônicas utilizadas. Foram excluídos estudos secundários, literatura cinzenta, estudos não publicados nas plataformas indexadas, estudos duplicados, estudos que apresentaram o uso da GSS sem informações sobre suas propriedades psicométricas.

Estratégia de busca

Os descritores utilizados foram: (*Gudjonsson Suggestibility Scale*) OR (*Gudjonsson Suggestibility Scales*) para buscar estudos nas seguintes bases de dados: LILACS, PubMed, Embase, Scopus, CINAHL, Academic Search Premier, MEDLINE, APA Journals, PePSIC, Scielo, PsycINFO, sendo a consulta realizada nos dias 15 e 16 de setembro de 2023. As buscas nas bases de dados eletrônicas foram realizadas sem restrições de idioma e ano de publicação.

Tabela 2 - Esquema estratégia de busca

Bases de dados	Estratégias de busca
LILACS, PubMed, Embase, Scopus, CINAHL, Academic Search Premier, MEDLINE, APA Journals, PePSIC, Scielo, PsycINFO	(Gudjonsson Suggestibility Scale) OR (Gudjonsson Suggestibility Scales)

Seleção dos estudos

Terminado o processo de busca dos estudos nas base de dados, foram excluídos os duplicados e os remanescentes foram baixados para o aplicativo da *web Rayyan QCRI* (<http://rayyan.qcri.Org/>) (Hupe, 2019; Ouzzani *et al.*, 2016), para que fossem selecionados com base no título e resumo por dois avaliadores (R.L.S.S. e M.P.G.). Os títulos e resumos foram analisados de acordo com critérios de elegibilidade, para identificação de estudos relevantes, classificando-os como: “sim” (incluído) ou “não” (excluído). Por fim, os estudos selecionados para leitura na íntegra foram submetidos à busca cruzada, processo pelo qual é analisada na lista de referência dos artigos para identificação de estudos relevantes que não teriam surgido na busca nas bases de dados eletrônicas (Higgins *et al.*, 2022).

Extração dos dados

No processo de extração dos dados os artigos selecionados foram examinados por dois avaliadores (R.L.S.S. e F.N.S.), de forma independente, devidamente treinados e familiarizados com a planilha de extração de dados, para garantir maior confiabilidade entre os avaliadores e a adequação dos itens extraídos. A planilha de extração de dados foi construída tendo como base os critérios de elegibilidade, utilizando o software de planilha eletrônica Excel (*software Microsoft Excel; Microsoft*

Corporation, WA, EUA) e foi realizada a calibração com um estudo da área. Foram extraídos os seguintes dados dos estudos: autores, tipo de estudo, número total da amostra, características do estudo, idioma de validação, tipo de aplicação do instrumento, resultados das propriedades psicométricas (evidências baseadas no conteúdo, evidências baseadas no processo de resposta ao item, evidências baseadas na estrutura interna, evidências baseadas nas relações com variáveis externas e fidedignidade ou confiabilidade da medida), tipo de financiamento e conflitos de interesse.

Os estudos selecionados foram lidos e examinados integralmente pelos avaliadores (R.L.S.S. e F.N.S.) de forma independente, usando um formulário descrito anteriormente (Higgins *et al.*, 2019; Mokkink *et al.*, 2018). A posteriori, os formulários foram comparados e todas as discordâncias entre os avaliadores foram discutidas em busca de um consenso, não sendo possível alcançá-lo, um terceiro avaliador foi consultado (M.P.G.).

Avaliação do risco de viés

A avaliação de viés é uma das etapas que diferenciam a Revisão sistemática de outras revisões, visto que é averiguada se houve algum processo de apreciação crítica dos estudos que foram incluídos. Aqui entende-se viés como sendo erro sistemático, ou seja, uma tendência a se desviar da verdade de maneira proposital ou não. Assim, os autores das revisões devem avaliar se o viés pode subestimar ou superestimar os dados encontrados (Higgins *et al.*, 2019; Mokkink *et al.*, 2018; Brink & Louw, 2012).

Para realizar a avaliação da qualidade metodológica dos estudos incluídos nesta revisão foi utilizada a ferramenta *Critical appraisal tool for validity and reliability*

studies of objective clinical tools (Brink & Louw, 2012) Este instrumento consiste em 13 itens em formato de pergunta que podem ser respondidas com “sim ou não”, são elas: (1) Se foram usados sujeitos humanos, os autores deram uma descrição detalhada da amostra de sujeitos usados para realizar o teste?; (2) Os autores esclareceram a qualificação ou competência do(s) avaliador(es) que aplicaram o teste?; (3) O padrão de referência foi explicado?; (4) Se a confiabilidade entre os avaliadores foi testada, os avaliadores estavam cegos para os achados de outros avaliadores?; (5) Se a confiabilidade intra-avaliador foi testada, os avaliadores estavam cegos para suas próprias descobertas anteriores do teste sob avaliação?; (6) A ordem do teste foi variada?; (7) Se sujeitos humanos foram usados, o período de tempo entre o padrão de referência e o teste de índice foi curto o suficiente para ter certeza razoável de que a condição alvo não mudou entre os dois testes?; (8) A estabilidade (ou estabilidade teórica) da variável que está sendo medida foi levada em consideração ao determinar a adequação do intervalo de tempo entre medidas repetidas?; (9) O padrão de referência foi independente?; (10) A execução do teste foi descrita com detalhes suficientes para permitir a replicação do teste?; (11) A execução do padrão de referência foi descrita com detalhes suficientes para permitir sua replicação?; (12) As retiradas do estudo foram explicadas?; (13) Os métodos estatísticos foram apropriados para o propósito do estudo?

Para utilização dessa ferramenta os avaliadores foram submetidos a um treinamento e familiarização para avaliação do risco de viés, objetivando o treinamento foi realizada a calibração com um estudo da área. A análise foi realizada por dois avaliadores (R.L.S.S. e F.N.S.), independentes que avaliaram a qualidade metodológica dos estudos selecionados seguindo os critérios descritos pelo instrumento citado anteriormente. Todas as inconsistências entre os dois avaliadores foram discutidas a fim de

alcançar um consenso. Na ausência de consenso, um terceiro avaliador (M. P. G.) foi consultado para determinar a pontuação, ou não, do item.

Análise de dados

Inicialmente, foram analisadas as propriedades psicométricas dos estudos selecionados, em seguida foi realizada uma meta-análise dos resultados quantitativos dos estudos incluídos na revisão. A metanálise pode ser considerada como uma média ponderada das estimativas observadas nos estudos. Neste estudo, foram calculadas meta-análises de coeficientes de correlação usando a transformação Z de Fisher. As meta-análises foram calculadas usando modelos de efeitos aleatórios (EA), calculados pelo método da máxima verossimilhança restrita, assumindo que os estudos selecionados foram amostrados a partir de um conjunto maior de estudos (Higgins *et al.*, 2003; Prinsen *et al.*, 2018).

Destaca-se que o modelo de efeitos aleatórios foi escolhido em vez do modelo de efeito fixo devido a fatores experimentais, como a metodologia do estudo (por exemplo, condições das amostras), que podem influenciar o relato dos comportamentos interpessoais das amostras e a confiabilidade dos resultados dos estudos. Para além desses fatores, o modelo de efeitos aleatórios permite uma maior generalização externa, em comparação com o modelo de efeito fixo.

A heterogeneidade entre os estudos foi avaliada por meio da estatística do teste Cochran Q (adotando o nível de significância $p < 0,1$) e a inconsistência foi avaliada utilizando índice I^2 de Higgins (Higgins *et al.*, 2003). Como critério de avaliação foram adotados os seguintes referenciais: valores $\leq 40\%$ indicam baixa heterogeneidade; 30% a 60% indicam heterogeneidade moderada; $> 50\%$ a 90% indicam heterogeneidade substancial e $> 75\%$ a 100% indicam heterogeneidade considerável (Higgins *et.*

al, 2022). Quando $I^2 > 50\%$ e *tau squared* (τ^2) > 1 , foram acompanhados pela significância estatística ($p < 0,05$), foi considerada a existência de heterogeneidade significativa.

Para a análise da consistência interna e para a interpretação qualitativa, foram utilizados os valores do alfa de *Cronbach*, avaliados segundo as seguintes categorias: excelente ($\geq 0,85$); bom (0,80 – 0,84); moderado (0,75 – 0,79) e justo (0,70 – 0,74). Os valores de alfa de Cronbach foram determinados considerando os fatores e o tamanho amostral (Ponterotto; Ruckdeschel, 2007; Santos *et al.*, 2020). Para os resultados da evidência baseada na estrutura interna, foram considerados aceitáveis os seguintes critérios: *Comparative Fit Index* (CFI) ou *Tucker Lewis Index* (TLI) $> 0,95$; *Root-Mean-Square Error of Approximation* (RMSEA) $< 0,06$; ou *Standardized Root Mean Square Residual* (SRMR) $< 0,082$, *Average Variance Extracted* (AVE) $> 0,5$ (Prinsen, *et.al.*, 2018; Terwee, *et.al.*, 2007; Hair *et.al.*, 2021). Para os resultados da evidência baseada nas relações com variáveis externas (validade de critério, do tipo convergente ou discriminante; e do tipo concorrente ou preditiva), foram considerados aceitos os resultados $> 0,70$ (Prinsen, *et.al.*, 2018). Para os resultados de confiabilidade – fidedignidade, avaliados pelo *Intraclass Correlato Coefficients* (ICC) e pela confiabilidade composta, foram considerados aceitáveis resultados $> 0,70$ (Prinsen, *et.al.*, 2018). Todas as análises e as figuras foram desenvolvidas no software estatístico programa R versão 4.1.3 utilizando o pacote meta.

4. RESULTADOS

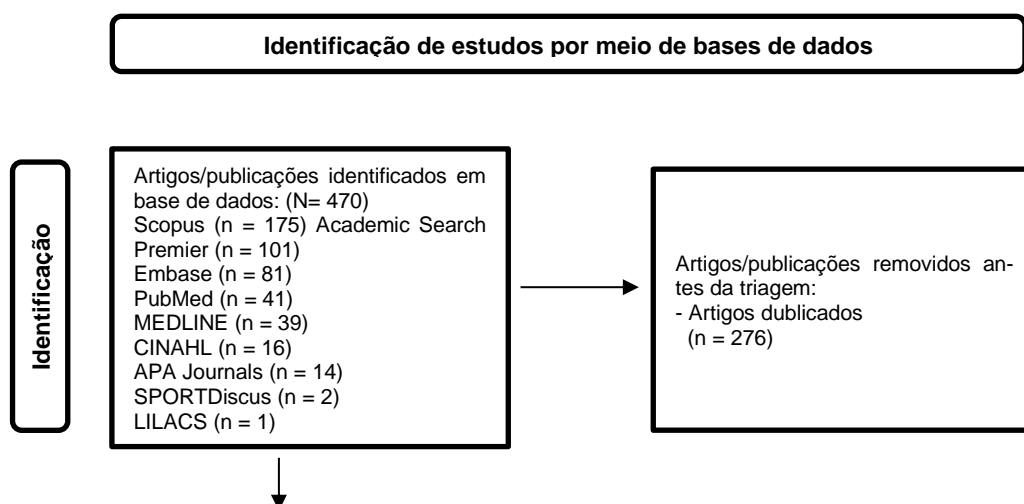
Durante o processo de busca foram encontradas 470 publicações em 09 bases de dados eletrônicas selecionadas [Scopus (n = 175); Academic Search Premier (n =

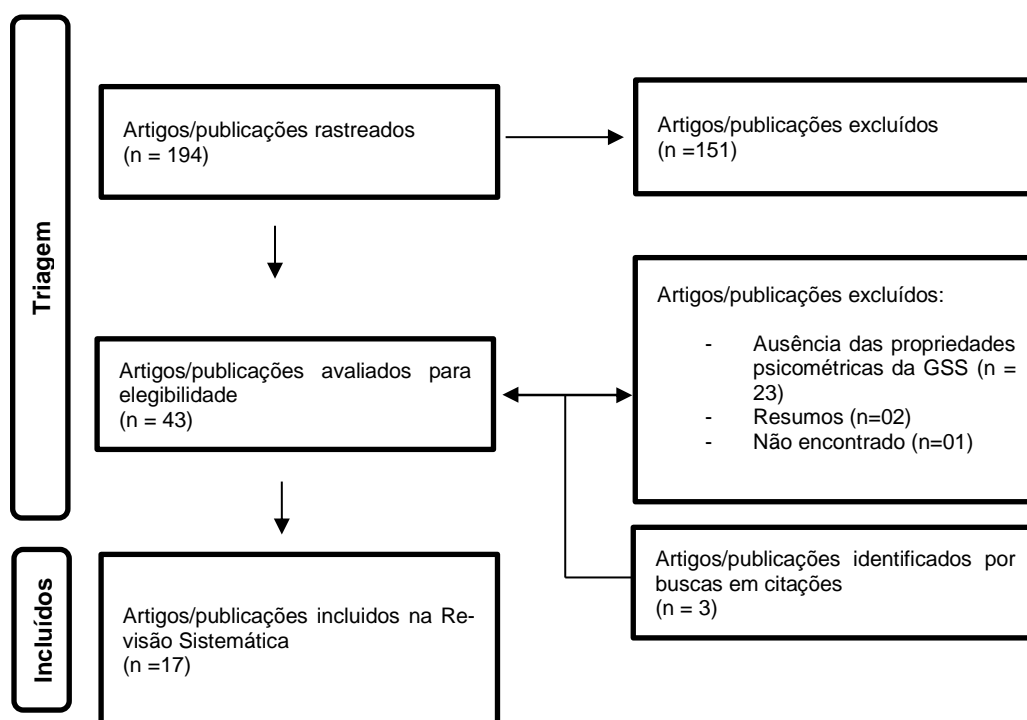
101); Embase (n = 81); PubMed (n = 41); MEDLINE (n = 39); CINAHL (n = 16); APA Journals (n = 14); SPORTDiscus (n = 2); LILACS (n = 1)], das quais 276 artigos duplicados foram excluídos.

Após a triagem por título e resumo, foram excluídos 151 artigos, restando 43 artigos para a leitura do texto completo. Nesta etapa de seleção por título/resumo a concordância foi de 74.2%, com 50 conflitos entre os avaliadores, os quais foram avaliados por um terceiro avaliador. Após a análise do terceiro avaliador restaram 43 artigos que passaram para a fase de leitura na íntegra.

Após a leitura na íntegra, 17 artigos foram eleitos para a extração quantitativa e qualitativa, nesta etapa de leitura na íntegra, a concordância foi de 69,77% (13 conflitos entre os avaliadores). Ao fim dessa fase foi realizada a busca cruzada para verificar nas referências dos estudos selecionados a existências de estudos que pudessem se encaixar nos critérios de elegibilidade para a revisão e que ainda não tivessem sido localizados pelas buscas em bases de dados. Aqui foram encontrados 03 artigos que se encaixavam nos critérios de elegibilidade, como não tinham acesso liberado, foi solicitado ao autor, porém não tivemos retorno até a conclusão dessa revisão.

Figura 2 -Flowchart da seleção dos estudos.





4.1. Características dos artigos

Para a síntese narrativa desta Revisão Sistemática, foram incluídos 17 (dezesete) artigos, com período de publicação, variando entre 1984 e 2023. O idioma de publicação de 16 (dezesesseis) estudos foi o inglês e 01 (um) deles foi publicado no idioma espanhol. As escalas GSS 1 e 2 foram administradas em 10 idiomas: inglês (Australia, Reino Unido, Canadá e Estados unidos), português de Portugal, polonês, espanhol, turco, italiano, alemão holandês, sueco e japonês.

A amostra total dos estudos foi de 4.752 sujeitos de pesquisa, com uma variação de tamanho da amostra de 31 a 1.261 participantes por estudo sendo 2.236 mulheres, 2.331 homens, 182 pertenciam a estudos que não especificaram sexo e 03 (três) sujeitos preferiram não especificar. A maioria dos participantes selecionados foram de ambos os sexos, porém, 02 (dois) estudos restringiram sua amostra apenas ao sexo masculino. Com relação à idade dos participantes, variou entre 06 (seis) e 78

anos; sendo 15 estudos com adultos e 02 (dois) com adolescentes e crianças (Tabela 3).

Dentre os estudos analisados 14 eram do tipo pesquisa observacional transversal, 02 (dois) longitudinais com o tempo de intervalo entre as aplicações da GSS variando de 1 dias a 4 meses e 02 (dois) dias com desenho experimental. Quanto ao tipo de escala pesquisada, 05 (cinco) estudos levantaram dados de ambas as escalas, 06 (seis) da GSS 1 e 06 (seis) da GSS 2.

Dentre as propriedades psicométricas, foram encontradas nos artigos selecionados: evidências de confiabilidade e fidedignidade, sendo esta a mais frequente, aparecendo 17 vezes; evidências baseadas na estrutura interna, evidências baseadas nas relações com variáveis externas e evidências baseadas no conteúdo (Tabela 3). Os demais tipos de propriedades psicométricas previstas (AERA, APA & NCME, 2014) não foram identificados nos estudos analisados.

Tabela 3. Síntese e características dos resultados dos estudos do GSSS.

Autores	Ano	Estudos	N° amostral total	Características dos estudos	Idioma de Validação	Aplicação do GSSS	Tipos de propriedades psicométricas	Conflito de Interesse (CI) Fonte de Financiamento (FF)
H. Merckelbach P. Muris I. Wessel P. J. Van Koppen	1998	Estudo 1 Observacional Longitudinal	N = 40	Estudantes universitários (37 mulheres e 3 homens); Média de idade: 20,6 anos (DP = 2,2, faixa etária de 18 a 26 anos)	Holandês GSS1 e GSS2	Presencial	EC	CI – ND FF - ND
H. Merckelbac P. Muris I. Wessel P. J. Van Koppen	1998	Estudo 2 Experimental	N=53	Estudantes universitários (41 mulheres e 12 homens. Média de idade: 20 anos (faixa etária de 18 a 27 anos)	Holandês GSS1 e GSS2	Presencial	EBE	CI – ND FF - ND
G. H. Gudjonsson	1987	Observacional Longitudinal	N= 90	Três grupos de sujeitos foram testados: (Grupo 1) 14 homens e 14 mulheres com idade média de 27 anos (DP = 7,7). (Grupo 2) 32 pacientes 'forenses' e 'clientes' encaminhados ao autor por advogados de defesa ou promotoria. 27 homens e 05 mulheres no grupo, com idade média de 31 anos (DP	Inglaterra GSS 1 e GSS2	Presencial	EC	CI – ND FF - ND

				= 11,9). (Grupo 3) 30 casos 'forenses' semelhantes aos do grupo 2, com idade média de 31 anos (DP = 10,3). 28 homens e 02 mulheres.				
I. C. H. Clare G. H. Gudjonsson S. C. Rutter P. Cross	1994	Observacional- Transversal	n=101	Idade dos participantes: entre 17 e 69 anos (média: 33,2 anos; DP: 10,5) Amostra com diferentes níveis de inteligência, com QI total entre 59 e 130 (média: 79,7; DP: 15,8), medido pela Escala de Inteligência Wechsler para Adultos - Revisada(WAIS-R).	Inglês (Reino Unido) GSS2	Presencial	EC	CI – ND FF - ND
Fee-Elisabeth Hein V. Scheuble A. Beauducel A. Leue	2021	Observacional- Transversal	N= 560	Idade média: 24,20 anos (DP = 4.60); 287 mulheres e 273 homens. Estudantes universitários das Universidades de Kiel e Bonn, Alemanha.	Alemão GSS1	On-line	EC	CI=ND FF= Financiada pela German Research Foundation (BE 2443/11-1, LE 2240/6-1)

G. E. Gignaca M. B. Powell	2009	Observacional Transversal	N=220	5,3 a 13,2 anos (M = 10,1, DP = 2,16; 43,2% do sexo feminino)	Australia (inglês) GSS2	NR	EC EBI	CI – ND FF - ND
T. Wachi K. Watanabe K. Yokota Y. Otsuka K. Hiram	2019	Observacional Longitudinal	N= 442	53 entrevistados foram removidos devido a respostas inadequadas ou tempos de resposta excessivamente longos. Assim, a amostra final foi de 442 (218 (49,3%) homens e 224 (50,7%) mulheres). As idades média e mediana foram 40,12 (DP = 10,66) e 40 anos, respectivamente (intervalo: 20–59 anos).	Japonês GSS2	On-line e Presencial	EBC EC	CI – ND FF - ND
R. Polczyk	2016	Observacional- Transversal	N=118	(69 mulheres e 49 homens), com idade média de 22,20 anos (DP = 2,00, intervalo de 18 a 29).	Polonesa GSS2	Presencial	EC	CI – ND FF - ND
E. I. Özcan N. Ziyalar	2023	Observacional- Transversal	N=175	Idades dos participantes variaram de 19 a 36. 104 eram mulheres e 68 eram homens, e 3 pessoas não especificaram seu gênero. A maioria dos participantes	Turco GSS 2	Presencial	EC	CI- Nenhum conflito de interesses foi declarado pelos Autores. FF- Os autores declararam que este es-

				era de estudantes de graduação. A idade média dos participantes era de 23,3				tudo não recebeu apoio financeiro.
R. Polczyk	2005	Observacional-Transversal	N=355	A amostra 1 foi administrada com o GSS 1; consistia em 148 sujeitos, 89 mulheres e 59 homens, com uma idade média de 22,8, DP=4,0 (intervalo 18–56). Destes, de 76 a 80 sujeitos também completaram outras medidas de diferença individual. A amostra 2 compreendeu 207 sujeitos (120 mulheres e 87 homens; média idade 22,4, DP=3,5) que completaram a Escala de Sugestibilidade de Gudjonsson 2	Polônês GSS1 e GSS2	Presencial	EC	CI – ND FF - ND
G. H. Gudjonsson	1992	Observacional-Transversal	N= 129	100 eram pacientes forenses do sexo masculino, com idade média de 28 (DP =10), e 29 eram homens adultos	Inglês, Reino Unido.	Presencial	EC	CI – ND FF - ND

				normais, com idade média de 26 (DP = 5).	GSS1 e GSS2			
G. H. Gudjonsson K. K. Singh	1984	Observacional- Transversal	N= 31	Meninos delinquentes, idade média de 14,3 anos (DP = 1.3, intervalo de 11-16).	Inglês Reino Unido GSS1	Presencial	EC	CI – ND FF - ND
G. H. Gudjonsson	1984	Observacional- Transversal	N=195	58 mulheres "normais", idade média de 31 anos (DP = 12,7); 56 homens "normais", idade média de 24 anos (DP = 5,8); 40 pacientes forenses com idade média de 30 anos (DP = 9,1); 41 crianças (principalmente meninos delinquentes), idade média de 14 anos (DP = 1,4).	Inglês (Reino Unido) GSS1	Presencial	EC	CI – ND FF - ND
G. Richardson	1993	Observacional Transversal	N=57	Jovens (48 homens, 9 mulheres) em uma unidade especializada, todos exibindo	Inglês (Inglaterra)	Presencial	EC	CI – ND FF - ND

P. Smith				comportamento perturbado e desordenado. faixa etária foi de 10,9 a 17,6 anos (idade média = 15,5)	GSS1			
K. Singh G. H. Gudjonsson	1987	Observacional Transversal	N=285	209 (73%) eram homens e 76 (27%) eram mulheres. A faixa etária foi de 11 a 65 anos.	Inglês (Inglaterra) GSS1	Presencial	EC	CI – ND FF - ND
T. Wachi K. Watanabe K. Yokota Y.Otsuka K. Hirama	2019	Experimental	N= 218	A amostra consistiu homens = 108, 49,5%; mulheres = 110, 50,5% que se envolveram tanto na recordação imediata no primeiro dia quanto na recordação tardia dois dias depois. A idade média foi de 39,87 (DP = 10,71) com uma variação de 20 a 59	Japonesa GSS2	Online	EC	CI – ND FF - ND
A. Bianco	2015	ESTUDO 01	N=1.261	GRUPO1= 624 (GSS1) adultos foram recrutadas	Italiano	Presencial	EC EBI	CI – ND FF - ND

A. Curci		Observacional- Transversal		por meio do método de amostragem de bola de neve. (52,6% mulheres), com idade média de 37,54 (DP = 13,91; intervalo = 16–78). GRUPO2= 637;(GSS2) (58,4% mulheres) com idade média de 36,67 (DP=14,29; intervalo=16–84);	GSS1 e GSS 2			
A. Bianco A. Curci	2015	ESTUDO 02 Observacional- Transversal	N=102	GRUPO 01= bateria de teste = GSS1. 50 participantes (52% mulheres), com uma idade média de 37,66 (DP=13,74). GRUPO 02= 52 (GSS2) participantes (51,9% mulheres) com uma média de idade de 36,27 (DP=14,14);	Italiano GSS1 e GSS 2	Presencial	EC EBE	CI – ND FF - ND

L. Angelucci	2005	Observacional- Transversal	N=320	Estudantes universitários. 143 mulheres e 177 ho- mens, de 16 a 28 anos (M=19,48)	Espanhol (Venezu- ela) GSS1	Presencial	EBI EC	CI – ND FF - ND
--------------	------	-------------------------------	-------	--	--	------------	---------------	------------------------

Legenda: EBC = evidências baseadas no conteúdo, EBI = evidências de validade baseadas na estrutura interna, EBE = evidências de validade baseadas nas relações com variáveis externas, EC = evidências de confiabilidade/fidedignidade, NR = não relatado e ND - Não declarado

4.2. Análise do Risco de Viés

Para essa etapa da análise foi utilizada a ferramenta *Critical appraisal tool for validity and reliability studies of objective clinical tools* (Brink; Louw, 2012) que tem como objetivo ser usada em revisões sistemáticas de estudos que testam a validade e confiabilidade de instrumentos, fazendo uma avaliação crítica da qualidade metodológica desses estudos (Brink; Louw, 2011).

Ressalta-se que durante a avaliação dos estudos, decidiu-se utilizar a nomenclatura “Nr” (não relatada) para alguns procedimentos encontrados, mas não explicados pela literatura utilizada como referência, por exemplo, em alguns estudos que foram investigados não tinha como objetivo avaliar aquele tipo específico de validade baseada na relação com medidas externas, tratavam especificamente da aplicação de teste padrão ouro para medida do construto avaliado, sendo eles os critérios: 03, 07, 09 e 11. Por sua vez, a nomenclatura “Na” (não se aplica) já é utilizada pela própria ferramenta e é de acordo com o padrão de validade desenvolvidos e testados por Brink e Louw (2011).

Ao avaliar os critérios de validade e confiabilidade nos estudos revisados, observa-se que todos os estudos supriram os critérios: 01 (descrição detalhada da amostra de sujeitos usados para realizar o teste), 08 (estabilidade teórica da variável sendo medida foi levada em consideração ao determinar a adequação do intervalo de tempo entre medidas repetidas), 12 (as retiradas de sujeitos do estudo foram explicadas), 13 (os métodos estatísticos foram apropriados para o propósito do estudo) e 10 (a execução do teste foi descrita em detalhes suficientes para permitir a replicação do teste) exceto o estudo de Singh e Gudjonsson, (1987). Já o critério 02 (os autores esclareceram a qualificação ou competência do(s) avaliador(es) que realizaram o teste) foi pontuado positivamente somente por 06 (seis) estudos: Richardson e Smith (1993),

Polczyk, (2016), Polczyk, (2005), Hein F. *et al.*, (2021), Gudjonsso, (1987) e Clare *et al.*, (1994).

Quanto aos critérios que avaliam a confiabilidade, somente dois estudos apresentaram cegamento Interavaliadores (critério 04), são eles Richardson e Smith, (1993) e Clare *et al.*, (1994). Já o cegamento intravaliador (critério 05) não foi alcançado por nenhum dos estudos avaliados, assim como o critério 06 que trata da ordem variada de aplicação do teste em que o único estudo aprovado, foi cumprido apenas no estudo de Polczyk, (2005) (Figura 3). Em conjunto, podemos observar que os itens 4, 5 e 6 tiveram comprometimento na metodologia o que pode introduzir risco de viés na revisão sistemática (Figura 3).

Alta frequência de “não aplicável” e “não relatados” pode sugerir que muitos não atendem itens críticos de validade e confiabilidade, indicando que não seguiram aos padrões metodológicos rigorosos ou ainda que alguns aspectos metodológicos não foram considerados importantes ou relevantes para certos estudos, o que pode também limitar a aplicabilidade dos resultados.

Figura 3 - Avaliação da qualidade metodológica dos estudos incluídos

	Statistical methods	Withdrawals from the study	Standard Replicability	Index test replicability	Reference standard independent	Adequacy in the time interval	Application time	Randomization in application order	Raters blinded (intrater)	Raters blinded (interrater)	Reference standard	Competence of the rater(s)	Sample description
Wachi T. et al., 2019b	+	+	!	+	!	+	!	-	-	-	!	-	+
Wachi T. et al., 2019a	+	+	!	+	!	+	!	-	?	-	!	-	+
Singh K. K., Gudjonsson G. H., 1987	+	+	!	-	!	+	!	-	-	-	!	-	+
Richardson G., Smith P., 1993	+	+	!	+	!	+	!	-	?	+	!	+	+
Polczyk R., 2016	+	+	!	+	!	+	!	-	-	-	!	+	+
Polczyk R., 2005	+	+	!	+	!	+	!	+	-	-	!	+	+
Özcan E.I., Ziyalar N., 2023	+	+	!	+	!	+	!	-	-	-	!	-	+
Merckelba, et al., 1988	+	+	!	+	!	+	!	-	-	-	!	-	+
Hein F. et al., 2021	+	+	!	+	!	+	!	!	!	!	!	+	+
Gudjonsson G. H., 1992	+	+	!	+	!	+	!	-	-	-	!	-	+
Gudjonsson G. H., 1984	+	+	!	+	!	+	!	-	-	-	!	-	+
Gudjonsson G. H., Singh K. K., 1984	+	+	!	+	!	+	!	-	?	-	!	-	+
Gudjonsson G. H., 1987	+	+	!	+	!	+	!	-	-	-	!	+	+
Gignac G. E., Powell M., 2009	+	+	!	+	!	+	!	-	-	-	!	-	+
Clare, I. C. H. et al., 1994	+	+	!	+	!	+	!	-	?	+	!	+	+
Bianco A., Curci A., 2015	+	+	!	+	!	+	!	-	-	-	!	-	+
Angelucci L., 2005	+	+	!	+	!	+	!	-	-	-	!	-	+

Legenda: ● = Sim ● = Não ● = Não relatado ● = Não aplicado

4.3. Síntese das Evidências de Confiabilidade por meio da Meta-análise

As evidências de confiabilidade para a escala GSS foram avaliadas com base em três fatores: Yields, Shift e Total. Para o fator Yields, a meta-análise incluiu 10 estudos com um total de 1823 observações. O modelo de efeitos aleatórios revelou um alfa de Cronbach (α) = 0,75 (IC 95% [0,67; 0,81], $p < 0,001$). A heterogeneidade foi significativa, com $I^2 = 94,3\%$, indicando alta variabilidade entre os estudos (Figura 03). Para o fator Shift, 11 estudos foram incluídos, com 2108 observações. O modelo de efeitos aleatórios mostrou um $\alpha = 0,67$ (IC 95% [0,64; 0,70], $p < 0,001$). A heterogeneidade foi menor em comparação ao fator Yields, com $I^2 = 36,5\%$, indicando moderada variabilidade entre os estudos (Figura 04). Para o fator Total, 7 estudos com 1129 observações foram analisados. O modelo de efeito aleatórios mostrou um $\alpha = 0,82$ (IC 95% [0,63; 0,91], $p < 0,001$). A heterogeneidade foi muito alta, com $I^2 = 97,5\%$, indicando variabilidade substancial entre os estudos (Figura 05).

Em resumo, as evidências de confiabilidade mensuradas pelo alfa de *Cronbach*, entre os dois fatores da GSS, variaram de 0,67 a 0,75, no modelo bifatorial; e da escala total de 0,58 a 0,82, representando um nível de consistência interna classificado de “justo” a “moderado”, respectivamente (Ponterotto; Ruckdeschel, 2007; Santos *et al.*, 2020), mas abaixo do recomendado no fator Shift e Total.

Figura 04- Comparação das evidências de consistência interna do fator Yield entre os estudos da escala GSS, do modelo bifatorial.

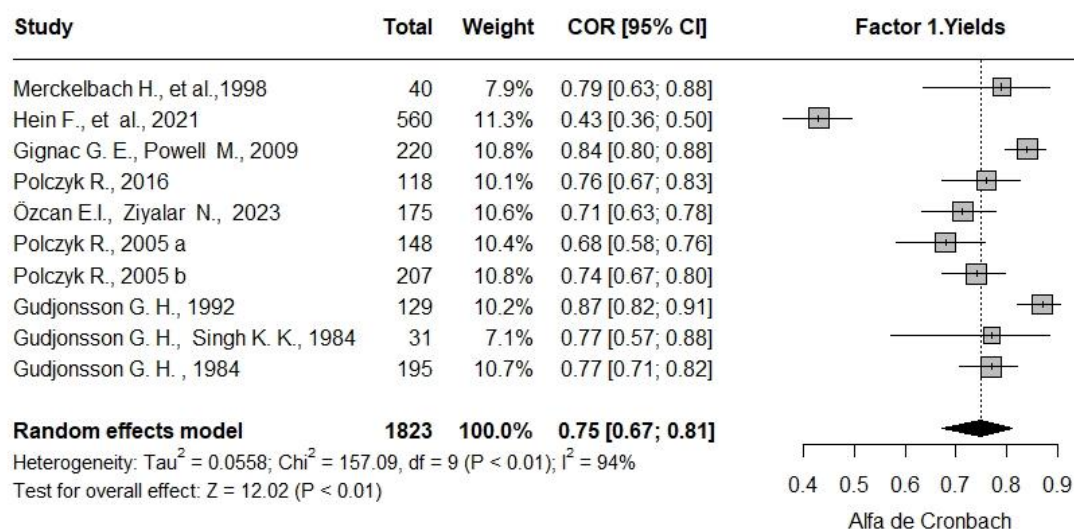


Figura 05 - Comparação das evidências de consistência interna do fator Shift entre os estudos da escala GSS, do modelo bifatorial.

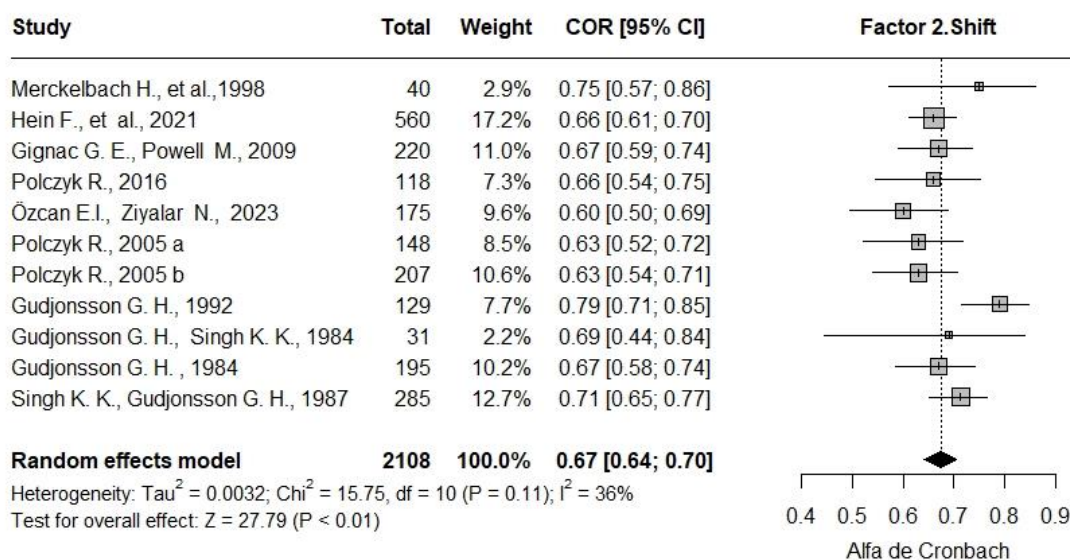
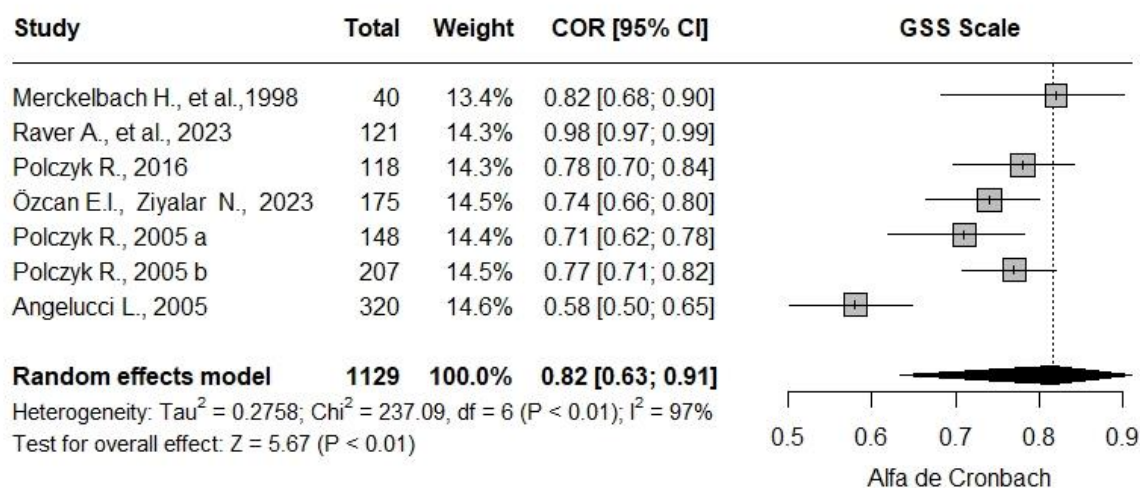


Figura 06 - Comparação das evidências de consistência interna do escore total entre os estudos da escala GSS, do modelo bifatorial.



4.3.1. Análise de subgrupos

Na análise de subgrupos para o fator Yield, comparando os tipos de escala GSS 1 e GSS 2, os resultados mostraram que a GSS 2 teve um $\alpha = 0,79$ (IC 95% [0,76; 0,81], $I^2 = 77,9\%$) e a GSS 1 um $\alpha = 0,57$ (IC 95% [0,52; 0,61], $I^2 = 94,5\%$). A diferença entre os grupos foi estatisticamente significativa ($p < 0,001$). Na análise de subgrupos para o fator Shift, a escala GSS 2 teve um $\alpha = 0,67$ (IC 95% [0,63; 0,71], $I^2 = 61,3\%$), enquanto a GSS 1 apresentou um $\alpha = 0,67$ (IC 95% [0,64; 0,70], $I^2 = 0,0\%$). Não houve diferença significativa entre os grupos ($p = 0,9861$). Na análise de subgrupos para o fator Total, a escala GSS 2 teve um $\alpha = 0,77$ (IC 95% [0,73; 0,80], $I^2 = 0,0\%$), enquanto a GSS1 apresentou um $\alpha = 0,78$ (IC 95% [0,75; 0,81], $I^2 = 99,1\%$). Não houve diferença significativa entre os grupos ($p = 0,5199$).

Figura 07 - Comparação das evidências de consistência interna do fator Yield, dividido pelos subgrupos GSS1 e GSS 2.

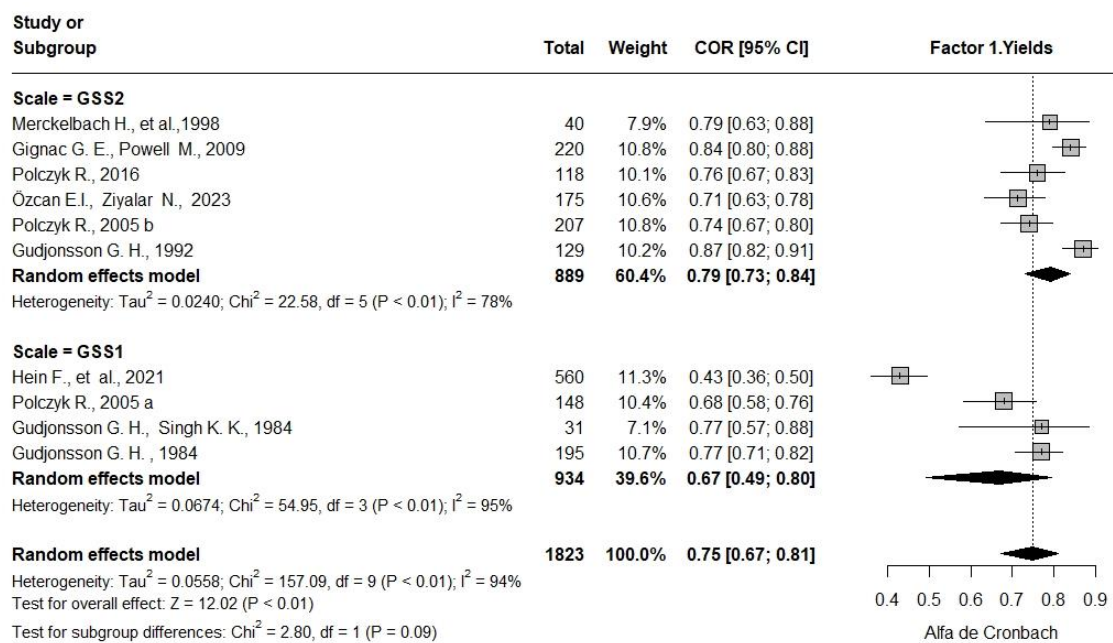


Figura 08 - Comparação das evidências de consistência interna do fator Shift, dividido pelos subgrupos GSS1 e GSS 2.

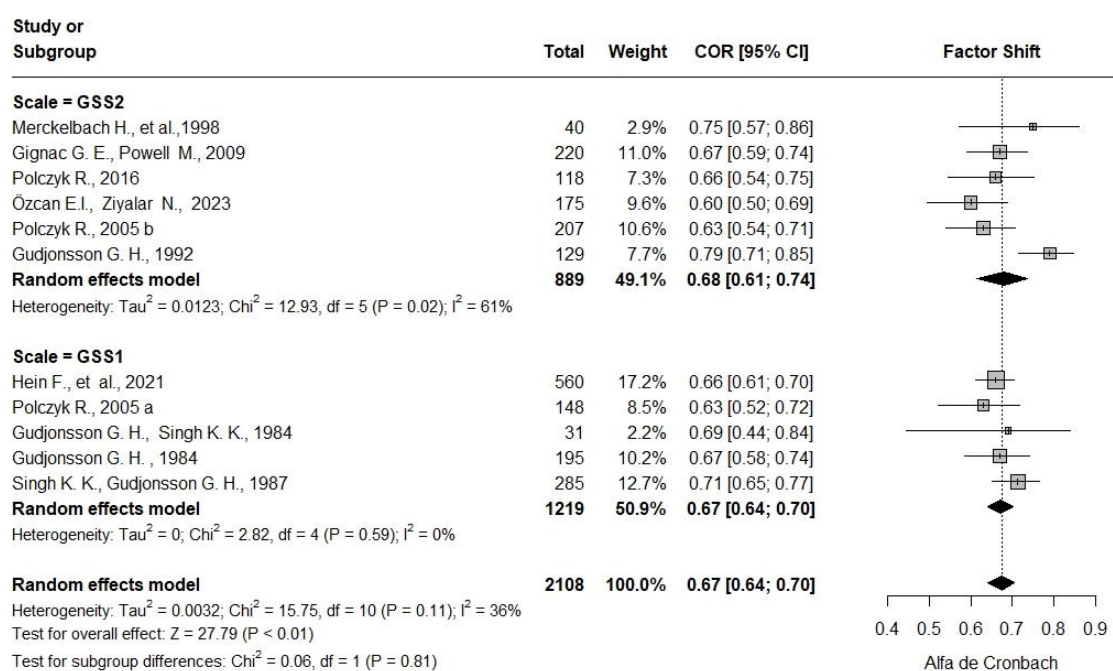
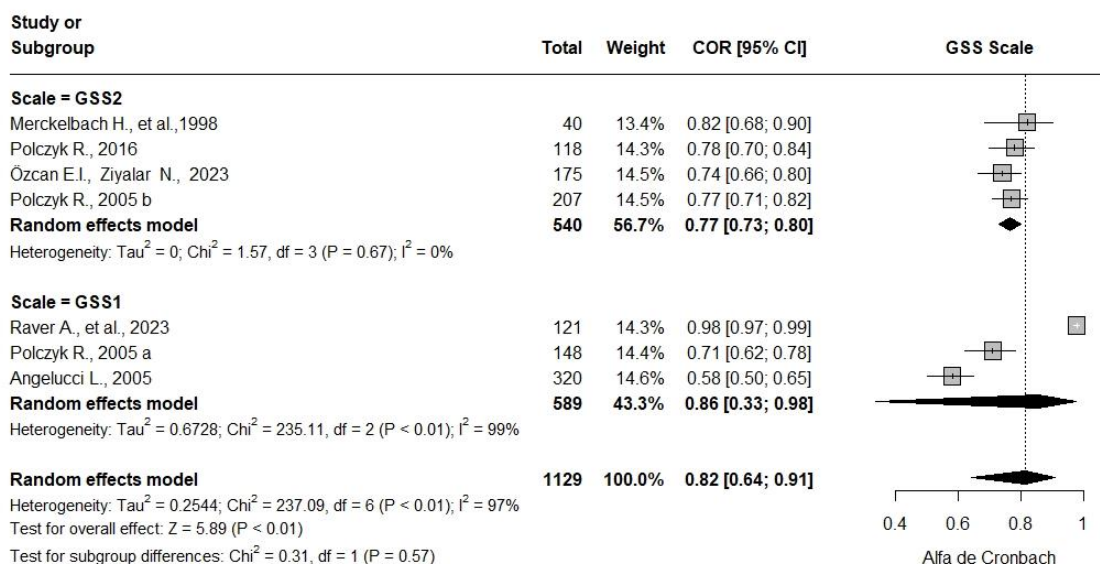


Figura 09 - Comparação das evidências de consistência interna do fator total, dividido pelos subgrupos GSS1 e GSS 2.



Em resumo, os resultados indicam que no geral a escala GSS apresenta evidências de fidedignidade com base no alfa de Cronbach com valores classificados como moderados (0,75 – 0,79) e justos (0,70 – 0,74) (Ponterotto, 2007; Santos, 2020), com correlações fortes no modelo de efeitos aleatórios. Porém, foram verificados no fator Yield um estudo (Hein *et al.*, 2021) com alfa de Cronbach abaixo do aceitável ($\alpha = 0,43$) com uma alta heterogeneidade nesse fator ($I^2 = 94,5\%$). Já no fator Shift, o alfa dos estudos metanalisados foi um pouco abaixo do recomendado pela literatura [$(\alpha = 0,67$; IC 95% [0,65; 0,69], $p < 0,001$)], com uma baixa heterogeneidade entre os estudos ($I^2 = 36,5\%$). Assim, a heterogeneidade variou entre os fatores, sendo mais alta para os fatores Yields e Total. A análise de subgrupos revelou diferenças significativas entre os tipos de escala GSS 1 e GSS 2, especialmente para o fator Yield.

3. DISCUSSÃO

Esta é a primeira revisão sistemática que avaliou as propriedades psicométricas da GSS, essa escala tem como objetivo medir de maneira objetiva a vulnerabilidade dos sujeitos à sugestionabilidade interrogativa (Gudjonsson, 1984). Esta revisão forneceu evidências que podem nortear sua aplicabilidade, qualidade metodológica e evidências das propriedades psicométricas. A seguir discute-se os resultados das evidências encontradas, iniciando com os dados qualitativos e em seguida as evidências da metanálise.

Seguindo as recomendações da área de Psicometria, foram analisados nessa revisão sistemática, os estudos que apresentavam propriedades psicométricas da GSS 1 e 2. Quanto as *Evidências de Validade baseadas no Conteúdo (EBC)*, esse tipo de validade foi identificada no estudo de Ozcan (2023), este autor relata que a tradução e a retrotradução do formulário da GSS foram feitas por 04 (quatro) psicólogos que conhecem o idioma inglês. As retrotraduções foram comparadas com o formulário original, a tradução mais parecida com os formulários originais foi escolhida para o estudo. Percebe-se que o processo de adaptação desse estudo, da maneira em que foi descrito, ficou resumido a questões gramaticais e linguísticas resumidas a tradução.

Este mesmo tipo de evidência foi verificado no estudo de Angelucci (2006) que teve como objetivo realizar uma análise psicométrica de uma versão em espanhol da GSS, a fim de determinar a confiabilidade e a validade do instrumento. O estudo relata que o processo de adaptação teve início a partir da versão original de Gudjonsson (1984) da GSS 1, que foi traduzida pela própria autora. A tradução foi avaliada e corrigida por 03 (três) juízes: 02 (dois) professores de inglês e 01 (um) de psicologia. Também foram realizadas mudanças adaptando as questões e a narrativa ao contexto

venezuelano, como exemplo a autora citou: “sul da califórnia” em vez de “sul de Croydon” “5 dolares” no lugar de “5 libras” e “Embaixada dos Estados Unidos” no lugar de “Embaixada Britânica”. Ao final desse processo ela obteve a versão final da GSS em espanhol que foi utilizada no estudo relatado pelo artigo.

Outra adaptação encontrada foi a de Polczyk (2005), segundo o autor as escalas GSS1 e GSS2 foram cuidadosamente traduzidos do inglês para o polonês para garantir a equivalência entre os idiomas. Alguns detalhes da narrativa, como nomes geográficos, foram alterados para garantir a compatibilidade com as realidades polonesas. Nenhuma retrotradução foi realizada, pois o autor afirmou que pequenas alterações no conteúdo da história ou nas perguntas não afetariam a ideia geral das escalas.

Atualmente, as melhores práticas no processo de adaptação recomendam a existência de pelo menos essas três fases: tradução direta, retrotradução e comitês de especialistas (Beaton *et al.*, 2000; Epstein *et al.*, 2015). Damásio (2015) baseado em Beaton *et al.*, (2000) recomenda os seguintes passos para realização do processo de adaptação de instrumentos: 1) tradução do instrumento; 2) síntese das versões traduzidas; 3) avaliação das sínteses por juízes experts; 4) avaliação do instrumento pelo público alvo 5) tradução reversa e 6) estudo piloto, sendo esse último passo opcional). Segundo esses autores, as evidências de validade de conteúdo podem ser consideradas um dos processos mais importantes na construção e adaptação de instrumentos, que quando não realizada de forma adequada pode prejudicar o entendimento do instrumento pela população alvo, comprometendo uma avaliação adequada do construto de interesse.

Sendo assim, os processos de adaptação para turco, espanhol e polonês descritos (Ozcan, 2023; Angelucci, 2006; Polczyk, 2005) seguem parcialmente os passos

para uma validação de conteúdo considerada adequada pela literatura, podendo comprometer em certa medida a avaliação da sugestionabilidade interrogativa nesses países. Os demais estudos de adaptação e validação das escalas GSS para outros idiomas, foram a adaptação para a língua portuguesa de Portugal, porém por ter sido publicada apenas em uma tese de doutorado (Pires, 2011), não cumpriu os critérios de inclusão adotados nessa revisão sistemática. Por sua vez, as adaptações para o alemão (Reutemann, 2006; Schemm, 2006), japonês (Watanabe et al., 2013) e italiano (Curci; Bianco, 2014) não foram analisadas nessa revisão por não se ter conseguido tais publicações na íntegra.

No que diz respeito as *Evidências de Validade baseadas na Estrutura Interna* (EBI), foram encontrados estudos que realizaram Análises Fatoriais Confirmatórias (AFC) totalizando 07 (sete) publicações (Özcan et al., 2023; Gignac; Powell (2009); Bianco; Curci, 2015; Angelluci, 2005; Hein et al., 2021; Gudjonsson, 1984,1992).

As primeiras indicações da estrutura fatorial das escalas GSS 1 e 2 foram apresentadas pelos criadores dessas medidas (Gudjonsson, 1984,1992). No primeiro estudo da década de 80 (Gudjonsson, 1984), esse autor realizou uma Análises de componentes principais (com rotação varimax), contando com uma amostra de 195 sujeitos ingleses. Dois fatores da sugestionabilidade surgiram: (1) Yield1 (rendimento); e (2) Shift (mudança). Essa estrutura fatorial identificada, foi compatível teoricamente com a proposta de Gudjonsson para explicação da sugestionabilidade interrogativa.

Já no estudo da década de 90 (Gudjonsson, 1992) realizou a aplicação da GSS 2 em 129 sujeitos todos do sexo masculino. foi realizada uma análise fatorial dos dados, também com rotação Varimax, sendo, portanto, o procedimento estatístico idêntico ao usado anteriormente com o GSS 1 (Gudjonsson, 1984). O resultado obtido

mostra que as pontuações Yield 1 (Rendimento 1) e Shift (Deslocamento) carregam em fatores separados o que apoia a visão teórica de Gudjonsson e Clark (1986).

Apesar da inegável contribuição da teoria sobre SI e da criação das escalas GSS 1 e 2 (Gudjonsson, 1984,1992), estudos posteriores testaram a estrutura fatorial das medidas em outros contextos e com diferentes amostras. Em alguns desses estudos foram verificadas algumas inconsistências na estrutura fatorial quando comparada com os estudos originais. Por exemplo, Bianco e Curci (2015) com o objetivo de investigar as propriedades psicométricas da adaptação italiana do GSS 1 e GSS 2, foi identificada uma estrutura fatorial com três fatores (Yield 1, Yield 2 e Shift) e não bifatorial (Yield 1 e Shift), como observado inicialmente (Gudjonsson, 1984,1992). Por meio de Análises Fatoriais Confirmatórias (AFC) com todos os 60 itens do GSS1 e GSS2 usando o software LISREL, para cada formulário GSS, foram testadas estruturas de um, dois e três fatores e comparados os índices de ajuste correspondentes. Os índices de ajuste utilizados foram Chi-square/df, RMSEA, SRMR, AIC, considerados apropriados para avaliar ajuste de modelos (Prinsen, et.al., 2018; Terwee, et.al., 2007; Hair et.al, 2021).

As autoras descreveram nessa publicação dois estudos, o primeiro recrutou duas amostras de pessoas adultas e italianas, realizado por meio do método de amostragem de bola de neve (N =1.261) divididos em dois grupos. A amostra 01 foi formada por 624 sujeitos que responderam a GSS1 e a amostra 02 por 637 sujeitos responderam o GSS 2. O procedimento padrão de administração foi seguido, como resultados, verificou-se que a solução de três fatores forneceu um modelo de ajuste melhor ajustado aos dados ($\chi^2/gf=4,40$, RMSEA=0,07, AIC=7763.75, SRMR=0,08), embora as diferenças entre os índices de ajuste para os três modelos não tenham sido significativas ($X^2/gf = 6,16$, n.s.). As diferenças entre os índices de ajuste dos três modelos

foram mínimas ou nulas ($(X^2/gl = 6,03, n.s.)$), o que pode demonstrar que as três estruturas alternativas foram igualmente plausíveis para explicar os dados (Bianco; Curci, 2015).

Por sua vez, Gignac e Powell (2009) focam suas análises nos componentes fatorais do GSS2 e questionam o modelo bifatorial. Para estes autores, a solução de componente que foi rotacionada por Gudjonsson (1992) via varimax, é um procedimento de rotação que força os componentes a serem ortogonais (independentes) sendo um método questionado atualmente. No caso da escala GSS, seguindo o procedimento padrão, Yield e Shift, devem ser somados para formar a pontuação de Suggestibilidade Total, o que implicaria em correlação positiva entre Yield e Shift. Assim, para Gignac e Powell (2009) a subescala Shift deve ser considerada como uma pontuação de diferença no índice, porque é a diferença entre as respectivas respostas de itens do Yield 1 ao Yield 2 que formaria o Shift.

Os autores propuseram, então, o cálculo de duas pontuações Shift alternativas: (1) Shift-positiva e (2) Shift-negativa. A primeira foi operacionalizada como a soma de respostas alteradas de Yield 1 para Yield 2, correspondendo a não endossar uma pergunta principal (antes do feedback) para posteriormente endossar uma pergunta principal (após o feedback). O segundo foi pensado como sendo a soma de respostas alteradas de Yield 1 para Yield 2, correspondendo a concordar com a pergunta principal e então não concordar com a pergunta principal, após receber feedback negativo.

Para testar esse modelo Gignac e Powell (2009) realizaram uma Análise Fatorial confirmatória com uma amostra de 220 crianças ($M=10,1, DP = 2.16$) australianas. Como resultado, as pontuações do Yield 1 demonstraram medir de forma confiável um construto (a extensão em que os sujeitos cedem a perguntas sugestivas) relacio-

nado a diferenças individuais na rendição a perguntas tendenciosas, porém, as pontuações da subescala Shift (a extensão em que os sujeitos podem ser levados a mudar suas respostas após o feedback negativo) foram associadas a um nível inadequado de confiabilidade de consistência interna (ω de McDonald's = 0,60). Segundo os autores, esse fato pode explicar o porquê muitos pesquisadores falharam em encontrar correlações teoricamente significativas entre Shift e critérios externos.

No entanto, o modelo proposto não teve sustentabilidade quanto aos dados psicométricos o que, segundo os autores, isso não anula o fato de que Shift incorpora dois processos de resposta diferentes positivos e negativos e essa pode ser a razão para a baixa confiabilidade de consistência interna. Visto isso, o artigo indica o uso cauteloso das GSS e não indica o uso da subescala Shift até que mais estudos de validade sejam realizados.

Hein et al. (2021) teve como objetivo investigar as propriedades psicométricas e a estrutura fatorial de uma versão online alemã da Escala de Sugestionabilidade de Gudjonsson 1 (GSS 1). Para tanto, esses autores realizaram uma AFC com dados coletados de uma amostra composta por 560 estudantes alemães. O modelo bifatorial teve um ajuste aceitável para Yield 1: [$\chi^2 = 174,43$, $df = 151$, $p = 0,09$]; RMSEA = 0,02; CFI = 0,89]; e para Yield 2: [$\chi^2 = 184,02$, $df = 149$, $p < 0,05$]; RMSEA = 0,02; CFI = 0,94], uma vez que a aceitação de desinformação sugerida pelas perguntas principais foram a pontuação do Yield 2. O modelo hipotético de fator único para Shift mostrou um ajuste de modelo muito bom [$\chi^2=181,74$, $df=170$, $p= 0,26$; RMSEA=0,01, CFI=0,98]. Assim, os dados do GSS 1 online alemão confirmaram a base teórica do GSS 1, pois os dois fatores Yield e Shift foram confirmados em um CFA. Neste estudo também foi apresentado resultados de análise de invariância fatorial, concluindo a invariância dos modelos com relação a sexo e instituição. A invariância dos scores da

GSS com relação aos sexos é uma conclusão já consensual na literatura (Wachi et al. 2019; Hein et al.,2021), tanto que a própria escala não apresenta nenhuma forma de avaliação ou pontuação diferente entre homens e mulheres.

Entretanto, é importante ressaltar que versões online das GSS podem apresentar como limitações o fato de que o relato é apresentado de maneira escrita e os participantes podem demorar o tempo que for preciso para lê-lo, já na versão presencial a história é contada oralmente, de modo que os participantes podem trapacear escrevendo pontos da história para recordar posteriormente. Porém, segundo Wachi et al., 2018 a versão online da GSS poderia servir como um método alternativo à versão padrão face a face, visto que, como os entrevistadores às vezes têm dificuldade em afirmar o feedback negativo com precisão e firmeza, a versão online poderia reduzir os efeitos do comportamento do entrevistador e padronizar o procedimento. Ela também poderia ser administrada quando os entrevistadores devidamente treinados não estivessem disponíveis.

Outro artigo que trouxe EBI por meio da AFC foi o de Özcan et al. (2023) este estudo objetivou adaptar a escala de sugestibilidade de Gudjonsson-2 (GSS-2) para o contexto turco, contando com a participação de 175 participantes, sendo esses estudantes de graduação). A análise fatorial confirmatória apresentou os seguintes índices de ajuste: ($\chi^2/gf=0,05$, RMSEA=0,001, GFI=0,99, SRMR=0,016). Segundo os autores a AFC falhou, o que resultou em 1 fator considerado unidimensional, no entanto, o fator Shift não mostrou propriedades unidimensionais, sendo que alguns itens não se encaixavam na estrutura do fator. Não houve relação significativa entre as pontuações do GSS-2 e as outras escalas determinadas para validade de construto.

Por fim, Angelucci (2006) teve como objetivo realizar uma análise psicométrica de uma versão em espanhol da GSS, a fim de determinar a confiabilidade e a validade

do instrumento. A GSS 1 foi administrada em uma amostra de 320 estudantes universitários venezuelanos, selecionada por disponibilidade. O procedimento seguido nesse estudo diferiu dos demais no fato que fora coletada a pontuação somente de Yield (rendimento/produção) sem a administração do feedback negativo. Segundo a pesquisadora essa possibilidade é prevista por Gudjonsson (1984) que indica a possibilidade de trabalhar as medidas das subescalas de forma independente tanto para pontuação como para análise psicométrica. Para análise de dados foi realizada uma análise de componentes principais com rotação Varimax. Os resultados revelaram a existência de dois componentes subjacentes a escala. O primeiro com o autovalor de 2,60, explicando 13% da variância total da escala e o segundo com 1,75, explicando 8,7% da variância. Diante dos dados coletados a autora avalia que os fatores obtidos apresentaram uma baixa consistência interna ($\alpha = 0,62$ e $\alpha = 0,47$, respectivamente) e explicaram de forma baixa a variância da escala, sendo, portanto, importante a realização de novos estudos para avaliar a validade da adaptação das escalas GSS no contexto espanhol.

Observado o procedimento seguido pela outra no processo de adaptação (não apresentou padrão ouro estabelecido pela literatura) e a realização de análises fatoriais com métodos não mais recomendados, a versão espanhola é carente de dados que confirmem a validade baseada na estrutura interna. E como já foi dito, a autora também aponta que não utilizou o feedback negativo em seu procedimento de aplicação da escala, sendo coletado somente a pontuação de Yield. Como o próprio Gudjonsson (1984) afirma, as subescalas que formam a GSS podem ser trabalhadas separadamente, porém não se sabe até que ponto isso pode ter influenciado nos dados psicométricos apresentados.

Partindo para verificação dos estudos que buscaram apresentar *Evidências de Validade baseadas nas relações com Variáveis Externas* (EBE) das escalas GSS 1 e 2, foi verificado nessa revisão sistemática, apenas dois estudos que se propuseram a testar esse tipo de evidência, porém durante as buscas foram encontrados estudos de correlação entre sugestionabilidade e memória (Merckelba; Wessel; Koppen, 1998; Cunha, et.al. 2007; Raver, et.al. 2023; Saraiva e Albuquerque, 2015); inteligência (Clare et al.,1994; Polczyk, 2005; Milne et al.,2002; Biondi et al.,2020; Singh; Gudjons-son, 1992); e personalidade (Pires et al.,2014; Hansen et al.,2010; Baxter, et al.2013; Haraldsson, 1985; Polczyk, 2005; Gorassini, 2006).

Um dos estudos descritos no artigo de Merckelbach *et al.*, (1998) foi examinada a validade preditiva das pontuações de rendimento (Yield) do GSS 1. Segundo Pacico et al. (2015) esse tipo de validade indica a força da relação entre os escores do teste e o resultado de um outro instrumento ou situação, obtidos após um intervalo de tempo. Contando com a participação de 53 estudantes de graduação, estes completaram a GSS 1 e então foram expostos a uma série de slides, dos quais o slide do meio, era crítico, pois retratava uma vítima de acidente. Após um intervalo de tempo (não foi especificado quanto), os participantes receberam uma série de perguntas de memória sobre este slide crítico, entre as quais havia 10 perguntas principais. A pontuação média do fator Yield para esta amostra foi de 4,0 (DP = 1,2), enquanto a pontuação média nas perguntas sugestivas sobre o slide crítico foi de 1,4 (DP = 1,4). A correlação de Pearson entre as pontuações em Yield dos participantes e a extensão em que eles cederam às perguntas principais foi pequena, mas significativa: $r = 0,22$, $p < 0,05$.

Por sua vez, no estudo 2 de Bianco e Curci (2015) o GSS 1 e o GSS 2 foram administrados juntos com medidas cognitivas, de personalidade e afetivas (como enfrentamento e ansiedade) para estabelecer sua validade discriminante (divergente). Para isso, foram realizadas aplicação de duas baterias de testes com subgrupos aleatoriamente selecionados da amostra normativa. O primeiro subgrupo foi formado por 50 participantes que preencheram a bateria com os seguintes testes: versão italiana GSS 1, três subescalas que medem inteligência verbal (Vocabulário, Compreensão e Analogias) da Escala de Inteligência de Adultos de Wechsler – Revisada (WAIS-R); As Matrizes Progressivas Padrão (MPS) para investigar inteligência fluida; teste de Rey para avaliar fatores relacionados a memória; a forma abreviada do Questionário de Personalidade de Eysenck – Revisado (EPQ-R) para avaliar dimensões da personalidade; o Inventário de Ansiedade Traço-Estado (STAI-Y) e a Orientação de Enfrentamento para Problemas Experimentados versão Italiana (COPE-NIV) foi empregada para avaliar o comportamento e os pensamentos dos participantes em resposta à situação interrogativa. O segundo subgrupo foi formado por 52 participantes que preencheram a mesma bateria, substituindo apenas a GSS 1 pela GSS 2.

Os resultados obtidos sugerem que o GSS 1 e o GSS 2 discriminam entre a sugestibilidade interrogativa e outras medidas cognitivas. A falta de uma forte associação entre as pontuações do GSS e as pontuações de personalidade sugere também que a sugestibilidade interrogativa, quando medida em uma amostra italiana, pode ser um aspecto distinto da personalidade. Não foram encontradas correlações significativas entre a sugestibilidade e a tendência a experimentar ansiedade contrapondo a ideia inicial de Gudjonsson (1983) e em consonância com outros estudos (Haraldsson, 1985; Pires *et al.*, 2013; Polczyk, 2005) assim como também não foram encontradas correlações com orientação para enfrentamento.

É importante destacar que para ser selecionado como evidencia EBE o estudo necessitava ter esse objetivo claramente descrito, esse critério excluiu a maior parte dos estudos de correlação encontrados no processo de busca. Estima-se que essa escassez seja devido à dificuldade em estabelecer correlações significativas ao ponto de o autor poder afirmar uma possível validade de critério, discriminante ou convergente. Portanto, com o pouco grau de evidências da relação com outros instrumentos e construtos, ainda não é possível, por exemplo, estabelecer redes nomológicas da sugestibilidade interrogativa.

Considerando as *Evidências de confiabilidade* (EC) do tipo consistência interna, foi verificado que das análises de consistência interna existentes, as mais utilizadas nos estudos analisados foram o coeficiente de alpha de Cronbach (α), em sua maioria, aparecendo somente um estudo que realizou o coeficiente de Kuder–Richardson (K-R 20) (Bianco; Curci, 2015) e um estudo que conduziu uma análise de ômega de *McDonald's* (Gignac; Powell, 2009). Quanto ao Alpha, ele foi desenvolvido por Lee Cronbach em 1951 para fornecer uma medida da consistência interna de um instrumento e é expresso como um número entre 0 e 1 e a literatura considerada como aceitável para se indicar consistência interna de um fator ou escala, valores de alfa a partir de 0,70 (Ponterotto; Ruckdeschel, 2007; Santos *et al.*, 2020).

Gudjonsson (1984) realizou um dos primeiros cálculos de alpha publicados para a GSS, ele apresentou o Coeficiente Alpha dos 15 itens de Yield e dos 15 itens Shift como sendo os valores 0,77 e 0,67, respectivamente para a amostra observada (n=195 sujeitos). Observando um pior desempenho do fator Shift o estudo de Singh e Gudjonsson (1987) realizou um procedimento testando outras formas de calcular a pontuação de shift, testando várias hipóteses, mas a que teve melhor resultado, elevando o alpha de shift para 0,71, mas com um o acréscimo dos 5 itens afirmativos,

anteriormente desconsiderados, sendo indicado pelos autores realizar o cálculo baseado em todos os 20 itens da escala e não somente nos 15 como feito originalmente. Esse estudo fez com que fossem realizadas modificações no procedimento padrão de aplicação das escalas. Entretanto, é consenso na literatura da área que o aumento dos itens pode influenciar no valor do alfa (Damásio, 2017), não sendo, portanto, um procedimento que garante a precisão do instrumento.

Bianco e Curci (2015) apresentaram valores de alfa Cronbach das subescalas Recordação imediata (Immediate recall) =0,95 e de Recordação atrasadas (Delayed recall) =0,93 que estão bem acima do ponto de corte 0,70. Para Tavakol e Dennick, (2011) valores de alfa muito alto (acima de 0,90), pode sugerir que alguns itens são redundantes, pois estão testando a mesma pergunta, mas em uma aparência diferente. Esse foi o único estudo que junto ao alpha foi calculado o coeficiente de Kuder-Richardson das subescalas rendimento 1 (yield 1) = 0,66, rendimento 2 (yield2) = 0,71, mudança (shift) = 0,58 e sugestibilidade total = 0,68, valores esses considerados questionáveis e abaixo da média considerando o restante dos estudos.

O estudo de Gignac e Powell (2009) foi o único presente na revisão sistemática que apresentou em seus resultados os valores de Ômega de McDonald's que, da mesma maneira que o Alpha de Cronbach, é um estimador de confiabilidade com índices bastante sensíveis de consistência interna, muito mais que o Alpha (Wagner, 2021). As pontuações da subescala Yield 1 obtiveram um ômega de McDonald's (ω) =0,83, sugerindo um nível respeitável de confiabilidade de consistência interna. Em contraste, no fator Shift o ômega de McDonald's foi abaixo do recomendado (ω = 0,60), sugerindo um nível inadequado de confiabilidade de consistência interna. A baixa confiabilidade associada às pontuações do Shift pode explicar por que vários pesquisadores falharam em encontrar correlações teoricamente significativas entre o

Shift e os critérios externos. Ao final o autor contraindica a utilização da subescala shift para avaliação da sugestionabilidade e conseqüentemente a utilização da GSS deve ser realizada com cautela.

Em geral, esses resultados indicam que na maioria dos estudos, o fator Shift apresentou indicadores de consistência interna, inferiores aos de Yield, essa diferenciação também foi vista nas análises fatoriais discutidas no tópico anterior, assim como nos estudos de correlação que não foram incluídos na Revisão Sistemática, mas foram utilizados no processo de construção da revisão de literatura. A metanálise realizada também apontaram um desempenho menor de Shift.

No presente estudo, apenas o Alfa de *Cronbach* foi metanalisado devido a sua maior ocorrência nos estudos identificados, por ser o índice mais descrito na literatura; e, ainda, o mais utilizado para verificação da consistência interna de instrumentos psicológicos; porém, sugere-se que em estudos futuros de avaliação da GSS, os demais indicadores de consistência interna (ômega de *McDonald's* e a confiabilidade composta) sejam verificados.

Todo os valores alpha de Cronbach foram metanalisados, agrupados pela metanálise, de maneira geral foram considerados moderados (0,75 – 0,79) e justos (0,70 – 0,74) seguindo os parâmetros já citados (excelente ($\geq 0,85$); bom (0,80 – 0,84); moderado (0,75 – 0,79) e justo (0,70 – 0,74), Ponterotto; Ruckdeschel, 2007; Santos *et al.*, 2020), o que aponta para a confiabilidade da GSS. Como já foi relatado os resultados indicam que a escala GSS apresenta propriedades psicométricas aceitáveis, o que sugere que a escala GSS é confiável na medição de sugestionabilidade interrogativa. O fato de a heterogeneidade ter sido maior para Yield e Total sugere que os alfas encontrados foram diferentes entre os estudos para esse fator.

A alta heterogeneidade (I^2 alto) entre os valores de alpha indica que os resultados desses fatores variaram significativamente entre os diferentes estudos, possivelmente devido a diferenças metodológicas, populacionais ou contextuais. Comparar os resultados entre diferentes subgrupos pode revelar padrões que ajudam a entender a heterogeneidade, aqui comparamos as duas versões da GSS. As diferenças significativas entre os tipos de escala GSS2 e GSS1, especialmente para o fator Yield deu a GSS2 melhores desempenho especialmente para o fator Yield, podendo-se sugerir para pesquisadores e profissionais qual versão da escala utilizar em diferentes contextos ou populações. Visto isso, sugere-se que mais estudos sejam realizados para entender as fontes de variação.

No que se refere ao tipo de fidedignidade *Teste-reteste*, Zanon Filho (2015) afirma que é um método de avaliação da consistência dos escores ao longo de um intervalo de tempo. Ele se baseia em uma correlação dos escores dos mesmos sujeitos avaliados em momentos distintos. Os escores de correlação que foram considerados adequados são os superiores a 0,80. Aparentemente, não existe um consenso do padrão de intervalo de tempo que precisa ser dado para minimizar as interferências de outros fatores, como por exemplo a memória e a aquisição/desenvolvimento de novas habilidade do sujeito e outros fatores que possam produzir distorções consideráveis nos escores finais coletados pela escala.

O estudo que se propôs a avaliar a confiabilidade teste-reteste foi o de Merkelbach et al. (1998) que apresentou como objetivo investigar as propriedades psicométricas (consistência interna e estabilidade), além de sua relação com medidas de metacognição. O procedimento seguido foi: os participantes ouviram a curta história do GSS 1, então se envolveram em uma breve tarefa que consistia em escalas de autorrelato (entre as quais o DES e o CFQ), e finalmente responderam às 20 questões

do GSS. Em seguida, os participantes receberam o feedback negativo e responderam as 20 questões novamente. Para ter uma impressão da estabilidade do GSS, a narrativa do GSS foi lida para os alunos pela segunda vez, quatro semanas depois e então eles responderam a escala de rendimento novamente. A correlação teste-reteste foi de 0,55 ($p < 0,01$), considerado inadequado.

Outro estudo que apresentou esse tipo de confiabilidade foi o de Gisli H. Gudjonsson, (1987), o qual buscou encontrar indícios de confiabilidade teste-reteste da sugestionabilidade interrogativa. O autor relata que o procedimento de teste-reteste foi feito pelos mesmos entrevistadores e que em sua maioria foi o próprio autor, porém não entra em detalhes sobre porcentagens. A amostra foi composta por 90 participantes, divididos em 03 (três) grupos. Os sujeitos foram testados em duas ocasiões separadas (a primeira vez com a GSS 1 e a segunda com a GSS 2) sob circunstâncias semelhantes pelo mesmo experimentador, após intervalos de tempo que variaram entre 01 (uma) semana a 4 (quatro) meses (dependendo da disponibilidade do sujeito). Como resultados foi apresentada uma tabela com valores de correlações entre a GSS 1 e 2 ($r=0,92$, $p < 0,001$). Assim, para o autor a sugestionabilidade interrogativa parece ser uma característica comportamental estável com relação ao tempo no contexto testado.

Outro tipo de método para aferir fidedignidade de instrumentos psicométricos é o *Método das metades* (split-half), o qual consiste em verificar a correlação entre as duas metades do teste, quanto maior o valor da correlação maior a fidedignidade. O grande problema desse método é determinar como o teste deve ser dividido (Zanon; Hauck-Filho, 2015) podendo chegar a resultados diferentes dependendo da distribuição realizada dos itens entre as metades. Urbina (2007) sugere que se deve levar em consideração dois pontos principais: a) diferença sistemática entre os itens ao longo

do teste, por exemplo itens mais difíceis e os mais fáceis ou ainda subescalas que medem conteúdos diferentes; b) quando o desempenho do teste depende exclusivamente do tempo. Assim, para Urbina essas duas condições podem influenciar diretamente nos dados de fidedignidade do método duas metas.

Hein et al. (2021) também realizou o método split-half, realizando a divisão do conjunto total de itens em uma primeira metade (itens 1 a 10) e em uma segunda metade (itens 11–20). O autor não especificou se utilizou alguma técnica de divisão dos itens entre as metades. Os resultados nos seguintes coeficientes split-half corrigidos de Spearman-Brown foram: Yield 1 = 0,45, Yield 2 = 0,71 e Shift = 0,68. Zanon e Hauck-Filho (2015) destacam que o nível mínimo de confiabilidade recomendado pelo método das metades é $\geq 0,70$. Os resultados mostraram que Yield 1 foi ligeiramente inferior a Yield 2 e Shift, sendo assim, a subescala Yield 2 (0,71) foi a que teve melhor escore. O autor também calculou estimativa de Confiabilidade (Squared Factor Score Determinacy), usada para se aproximar mais do modelo de análise fatorial confirmatória (CFA). Equivalente à confiabilidade do fator de regressão, também conhecida como Hancock's H, que apresentou os seguintes valores: Yield 1: H = 0,59, Yield 2: H = 0,71, Shift: H = 0,71. Comparado ao nível mínimo de confiabilidade recomendado de H = 0,70 a 0,80, Yield 1 é ligeiramente inferior a Yield 2 e Shift.

Finalmente, quanto à confiabilidade interavaliadores, também considerado um método para estimar fidedignidade de instrumentos psicométricos, esta foi verificada em alguns estudos desta revisão sistemática, como o de Wachi et al. (2019) os quais calcularam a confiabilidade interavaliadores para recordação imediata e o resultado foi: $r = 0,99$. As pontuações do teste kappa para Yield1, Yield2 e Shift foram 0,97, 0,99

e 0,97, respectivamente. Para recordação atrasada no segundo dia, o primeiro codificador codificou todas as amostras ($n = 218$) e o segundo codificou 30% ($n = 66$). A confiabilidade interavaliadores deste recall foi $r = 0,99$.

Richardson e Smith (1993) com o objetivo de investigar a confiabilidade entre avaliadores da Escala de Sugestionabilidade de Gudjonsson (GSS1) para uma amostra de 57 sujeitos ingleses, verificaram correlações de confiabilidade entre avaliadores altamente significativas para todas as pontuações de sugestionabilidade. Os valores encontrados foram: Rendimento 1 (Yield 1) $r = 0,98$, $P = 0,001$; Rendimento 2 (Yield 2) $r = 0,99$, $P = 0,001$; Deslocamento (Shift) $r = 0,95$, $P = 0,001$; sugestionabilidade total $r = 0,99$, $P = 0,001$. O número total de perguntas/respostas individuais pontuadas por cada avaliador foi de 2280, a porcentagem de discordâncias de pontuação entre os dois avaliadores foi de 1,4%.

Por sua vez, Clare et al. (1994) realizaram um estudo com o objetivo de estender o trabalho realizado por Richardson e Smith (1993) na forma original do GSS, fornecendo dados de confiabilidade entre avaliadores para todas as medidas que podem ser derivadas do GSS 2. Contando uma amostra de 101 sujeitos ingleses, coletaram as recordações imediatas e tardias dos sujeitos, assim como, às respostas as perguntas antes e depois do feedback negativo, as quais foram escritas na íntegra e transcritas pelo investigador. Cópias das respostas foram dadas a três avaliadores treinados que seguiram as instruções padrão de pontuação da GSS. A confiabilidade entre avaliadores, estimada por meio de coeficientes de correlação intraclassa, calculados por meio da análise de variância, teve como resultados: Immediate recall – 0,97; Delayed recall – 0,95; Yield 1 – 0,99; Yield 2 – 0,99; Shift – 0,99; Total Suggestibility – 0,99 com $p < ,001$. Assim, a concordância significativa entre os três avaliadores indicou um alto nível de confiabilidade entre avaliadores.

A discussão sobre a influência do entrevistador nos resultados dos escores de sugestionabilidade é alvo de estudos (Baxter, 2003, 2006; Bain et al.,2010; Dukala, K. e Polczyk,2014). Segundo Boon (2010) até dois terços da variância na sugestionabilidade interrogativa podem estar potencialmente sob o controle do entrevistador. Esses dados apontam uma limitação no uso da escala, visto que, o profissional da ponta que por ventura sinta necessidade de avaliar a sugestionabilidade, deverá ser muito bem treinado. No estudo de Clare et al. (1994) em que os entrevistadores foram treinados pelo próprio Gudjonsson, criador da escala, eles criticaram a falta de informação e lacunas que ficam para aqueles aplicadores que dependem somente das instruções contidas no manual da escala ou ainda em artigos de validação.

Outro ponto importante, já discutido nesse trabalho e agora corroborado é a necessidade de treinamento específico para profissionais que tenham contato com vítimas e testemunhas, pois se em ambiente controlado, como de aplicação de uma escala, é possível constatar níveis de sugestão emitidos pelos entrevistadores em ambiente natural, em ambientes não controlados como aqueles em que os sujeitos são escutados por uma rede (saúde, segurança, assistência social) muitas vezes sem conhecimento específico a respeito de todos os processos que envolvem o testemunho, podem ser mais susceptíveis ao aumento da sugestionabilidade interrogativa.

5. CONCLUSÃO

As escalas GSS são utilizadas para medir sugestionabilidade interrogativa e segundo os autores essas medidas tem aplicabilidade em contextos clínico, pesquisa e jurídico (Gudjonsson, 2013). A partir dos resultados observados nessa revisão sistemática, percebe-se que a GSS é um instrumento que apresenta propriedades psi-

cométricas, aceitáveis indicando um potencial para utilização em pesquisa para aprofundar o conhecimento sobre a sugestionabilidade e outros construtos relacionados. Para fins clínicos a literatura sugere a utilização das subescalas de memória (imediate e tardia) e a confabulação (distorções e fabricações) que podem ajudar a fazer inferências sobre criação de falsas memórias (Ridley; Gudjonsson, 2013). No entanto, essas subescalas aparecem pouco nos estudos examinados, portanto não apresentam propriedades psicométricas descritas, principalmente a de confabulação, que embasa sua utilização para este fim (Clare *et al.*, 1994).

Segundo Ridley e Gudjonsson (2013) as escalas são particularmente aplicáveis a questões legais, como interrogatórios de testemunhas de crimes por policiais e interrogatórios de suspeitos de crimes. Os resultados dos estudos mostram a importância das perguntas e feedbacks durante o interrogatório, deixando evidente que perguntas sugestionáveis e o feedback negativo dado pelo entrevistador causam mudanças no depoimento da testemunha (Özcan; Ziyalar, 2023; Wachi *et al.*, 2018). Porém, para sua utilização no contexto jurídico, estudos psicométricos mais robustos e, sobretudo, definições teóricas e a estrutura fatorial precisam ser aprofundados para melhor compreensão da SI.

Portanto, ressalta-se que a GSS pode ser uma ferramenta útil no rastreio da vulnerabilidade dos sujeitos a sugestão, mas, entende-se que esse como todo instrumento deve ser usado com cautela e ressalvas. Se fosse validada no contexto brasileiro, poderia ser utilizada como fonte complementar no contexto da avaliação psicológica, se assim fosse avaliada pelo SATEPSI (Sistema de Avaliação de Testes Psicológicos).

Os estudos apresentam também, implicações em relação às práticas jurídicas, uma vez que um entrevistado (por exemplo, uma testemunha ou suspeito) é induzido

ao erro por policiais ou outras fontes, o que pode tornar difícil a recuperação das informações originais. Esse efeito pode ser minimizado, quando os entrevistados são solicitados a recordar a história livremente e/ou focar em lembrar a história original ignorando a pressão social (Richardson G., Smith P., 1993; Özcan E.I., Ziyalar N., 2023; Clare, I. C. H. *et al.*, 1994). Assim, o aprofundamento das pesquisas com o objetivo de viabilizar a sua aplicabilidade nessa seara parece ter um potencial maior de aplicabilidade.

Chama-se atenção para os resultados alcançados pelos processos de aplicação online (Hein. *et al.*, 2021; Wachi. *et al.*, 2019b; Wachi. *et al.*, 2019a) Apesar de limitações como a falta de controle sobre os participantes que pode trapacear anotando respostas, como também, pode dificultar a participação dos sujeitos com dificuldade no manuseio de tecnologias, nossos achados têm implicações para a GSS e a prática policial. Primeiro, a versão online da GSS poderia servir como um método alternativo à versão padrão face a face. Como os entrevistadores às vezes têm dificuldade em afirmar o feedback negativo com precisão e firmeza, a versão online poderia reduzir os efeitos do comportamento do entrevistador e padronizar o procedimento.

A versão online também poderia ser administrada quando os entrevistadores apropriados não estão disponíveis, já que os entrevistados podem ter alta influência nos escores das escalas (Singh; Gudjonsson, 1988; Richardson; Smith., 1993; Clare, *et al.*, 1994; Bianco; Curci, 2015; Wachi *et al.*, 2019b). Esse fato evidencia uma das grandes limitações das escalas GSS que é influência do avaliador, assim sugere-se a realização de estudos de validade intra e inter avaliadores com maior rigor metodológico, visto que na avaliação de viés foi um dos critérios que tiveram mais pontuações negativas.

Por fim, esse estudo de revisão sistemática utilizou métodos rigorosos e sistematizados com o intuito de diminuir a ocorrência de vieses sendo considerado o padrão ouro em pesquisas para levantamento de evidências científicas (Higgins, 2011). Já a metanálise teve como objetivo otimizar os resultados da confiabilidade dos dados, pois realiza a análise quantitativa dos estudos incluídos na revisão, fornecendo informações adicionais ricas para o entendimento do instrumento investigado (Egger *et al.*, 1997). Mas como qualquer método, possui limitações que precisam ser consideradas ao interpretar seus resultados, neste estudo pode-se citar: a heterogeneidade encontrada nessa metanálise, que se refere a diferença de valores de alpha apresentados que variaram entre 0,43 a 0,67. que quando muito alta pode indicar estão medindo coisas ligeiramente diferentes ou que as condições sob as quais os estudos foram realizados são substancialmente diferentes, visto que os estudos foram realizados com populações diferentes em contexto socioculturais extremos.

Outro fator a ser considerado é a avaliação da qualidade metodológica geral dos estudos e o risco de viés na Revisão Sistemática, onde pode-se concluir que a qualidade geral dos estudos pode ser fragilizada, principalmente no que se trata do cegamento nos estudos de confiabilidade e a falta de pesquisas que avaliassem a validade de critério, o que prejudicou ainda mais a pontuação geral do estudo.

Conclui-se que a GSS, apresentou consistência interna aceitáveis o que possibilita seu uso em alguns contextos, com as várias ressalvas já apresentadas, por isso as escalas não podem ser consideradas padrão-ouro na identificação de vulnerabilidade à sugestão. Por ser uma escala de uso respaldado pelo tempo, recomenda-se que estudos de validade mais recentes sejam feitos e que abarquem as propriedades psicométricas mais aceitáveis atualmente, seguindo o recomendado pela literatura atual da área (AERA, APA, NCME, 2014).

REFERÊNCIAS

- AMERICAN EDUCATIONAL RESEARCH ASSOCIATION; AMERICAN PSYCHOLOGICAL ASSOCIATION; NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, 2014.
- BAIN, S. A.; BAXTER, J. S.; FELLOWES, V. Interacting influences on interrogative suggestibility. *Legal and Criminological Psychology*, v. 9, n. 2, p. 239-252, 2004. DOI: 10.1348/1355325041719419.
- BAXTER, J. S.; BOON, J. C. W.; MARLEY, C. Interrogative pressure and responses to minimally leading questions. *Personality and Individual Differences*, v. 40, n. 1, p. 87-98, 2006. DOI: 10.1016/j.paid.2005.06.017.
- BAXTER, J. S.; JACKSON, M.; BAIN, S. A. Interrogative suggestibility: interactions between interviewees' self-esteem and interviewer style. *Personality and Individual Differences*, v. 35, n. 6, p. 1285-1292, 2003. DOI: [https://doi.org/10.1016/S0191-8869\(02\)00349-5](https://doi.org/10.1016/S0191-8869(02)00349-5).
- BEATON, D. E.; BOMBARDIER, C.; GUILLEMIN, F.; FERRAZ, M. B. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, v. 25, n. 24, p. 3186–3191, 2000. DOI: 10.1097/00007632-200012150-00014.
- BERNSTEIN, D. M.; LOFTUS, E. F. How to Tell If a Particular Memory Is True or False. *A Journal of Association for Psychological Science*, v. 4, 2009.
- BINET, A. *La suggestibilité*. Paris: Bod Books on Demand. ed. 2019.
- BRASIL. Presidência da República. Subsecretaria Geral. Lei nº 13.431. Estabelece o sistema de garantia de direitos da criança e do adolescente vítima ou testemunha de violência e altera a Lei nº 8.069, de 13 de julho de 1990 (Estatuto da Criança e do Adolescente). *Diário Oficial da União*, Poder Executivo, Brasília, DF, 04 de abril de 2017. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/lei/l13431.htm. Acesso em: 2017.
- BRINK, Y.; LOUW, Q. A. Clinical instruments: reliability and validity critical appraisal. *Journal of Evaluation in Clinical Practice*, v. 18, n. 6, p. 1126-1132, 2012.
- BRITO, L. M. T. Reflexões em torno da Psicologia Jurídica. In: CRUZ, R. M.; MACIEL, S. K.; RAMIREZ, D. C. (org.). *O trabalho do psicólogo no campo jurídico*. São Paulo: Casa do Psicólogo, 2005.
- BRITO, L. M. T. DE; PEREIRA, J. B. Depoimento de crianças: um divisor de águas nos processos judiciais? *Psicologia-USF*, Itatiba, v. 17, n. 2, p. 285-293, 2012.

Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-82712012000200012&lng=en&nrm=iso.

BRITTO DE MELO, H.; PAGNUSSAT, J.; CECCONELLO, W. W.; FAVERO, G. C. A Abordagem Cognitiva Para Interrogatórios: Buscando Informações ao Invés de Confissões. *Revista Brasileira de Segurança Pública*, [S. l.], v. 18, n. 1, p. 14–29, 2024. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-82712012000200012&lng=en&nrm=iso.

CECCONELLO, W. W.; MILNE, R.; STEIN, L. M. Oitivas e interrogatórios baseados em evidências: considerações sobre entrevista investigativa aplicado na investigação criminal. *Revista Brasileira de Direito Processual Penal*, v. 8, n. 1, p. 489–510, jan. 2022.

CHIEN, P. F. W.; KHAN, K. S.; SIASSAKOS, D. Registration of systematic reviews: PROSPERO. *Wiley Online Library*.2012. DOI: 10.1111/j.1471-0528.2011.03242.x

CLARE, I. C. H.; GUDJONSSON, G. H.; RUTTER, S. C.; CROSS, P. The inter-rater reliability of the Gudjonsson Suggestibility Scale (Form 2). *British Journal of Clinical Psychology*, v. 33, p. 357-365, 1994. DOI: 10.1111/j.2044-8260.1994.tb01132.x.

COLLOCA, L.; BENEDETTI, F. Placebo analgesia induced by social observational learning. *Pain*, v. 144, n. 1-2, p. 28-34, 2009. DOI: 10.1016/j.pain.2009.01.033.

CONSELHO FEDERAL DE PSICOLOGIA. Resolução CFP nº 09/2018. Estabelece diretrizes para a realização de Avaliação Psicológica no exercício profissional da psicóloga e do psicólogo, regulamenta o Sistema de Avaliação de Testes Psicológicos - SATEPSI e revoga a Resolução CFP nº 09/2018. *Diário Oficial da União: seção 1*, Brasília, DF, n. 170, p. 45-48, 5 set. 2018. Disponível em: <https://www.cfp.org.br>.

CORREIA, A. R. P. *Uma nova escala de sugestionabilidade interrogativa análoga à GSS-1: estudo exploratório com uma amostra de jovens portugueses*. 2012. Dissertação (Mestrado Integrado em Psicologia) – Faculdade de Psicologia e Ciências da Educação, Universidade de Coimbra.

CRUZ, M. P.; PINHO, M. S. Sugestionabilidade em pessoas idosas: Um estudo com a Escala de Sugestionabilidade de Gudjonsson (GSS 1). *Análise Psicológica*, v. 32, n. 2, p. 199-213, 2014. DOI: 10.14417/ap.613.

CURCI, A.; BIANCO, A. *Gudjonsson Suggestibility Scales. Manuale d'uso*. Firenze: Giunti OS Organizzazioni Speciali, 2014.

DAMÁSIO, B. F.; BORSA, J. C. (org.). *Manual de Desenvolvimento de Instrumentos Psicológicos*. São Paulo: Vetor, 2017.

DAVIS, D.; LEO, R. A. Acute Suggestibility in Police Interrogation: Self-regulation Failure as a Primary Mechanism of Vulnerability. In: RIDLEY, A. M.; GABBERT, F.;

LA ROOY, D. J. (eds.). *Suggestibility in Legal Contexts: Psychological Research and Forensic Implications*. John Wiley & Sons, Ltd., 2013.

DIAS, M. G. B. B.; SALTZSTEIN, S. D.; MILLERY, M. Raciocínio moral em interação social: Um estudo sobre sugestibilidade. *Estudos de Psicologia*, v. 4, n. 2, p. 199-219, 1999.

DUKAŁA, K.; POLCZYK, R. Age and Interviewer Behavior as Predictors of Interrogative Suggestibility. *The Journals of Gerontology: Series B, Psychological Sciences and Social Sciences*, v. 69, n. 3, p. 348-355, 2014. DOI: <https://doi.org/10.1093/geronb/gbt037>.

ENDRES, J. The suggestibility of the child witness: The role of individual differences and their assessment. *The Journal of Credibility Assessment and Witness Psychology*, v. 1, n. 2, p. 44-67, 1997. Disponível em: <http://truth.charleshon-tsphtd.com/JCAAWP/9701/9701.html>.

EPSTEIN, J.; OSBORNE, R. H.; ELSWORTH, G. R.; BEATON, D. E.; GUILLEMIN, F. Cross-cultural adaptation of the Health Education Impact Questionnaire: experimental study showed expert committee, not back-translation, added value. *Journal of Clinical Epidemiology*, v. 68, n. 4, p. 360-369, 2015. DOI: 10.1016/j.jclinepi.2013.07.01.

EYSENCK, H. J.; FURNEAUX, W. D. Primary and secondary suggestibility: an experimental and statistical study. *Journal of Experimental Psychology*, v. 35, n. 6, p. 485-503, 1945. DOI: 10.1037/h0054976.

FAVERO, G. C.; KATZER, E. F.; MELO, F. P.; NETTO, K. A.; CECCONELLO, W. W. Entrevista investigativa: conceitos, características e aplicabilidade. In: LINO, D. (Org.). *Psicologia Investigativa: Teoria e Prática*. 2023. p. 71-90.

FARRUGIA, L.; GABBERT, F. Vulnerable suspects in police interviews: Exploring current practice in England and Wales. *J Invest Psychol Offender Profil*, 2019. p. 1-14. DOI: 10.1002/jip.

FERREIRA, B. G. Psicologia do Testemunho: Nos trilhos da mentira em busca da verdade. 2018. Dissertação (Mestrado em Psicologia) - Universidade Lusófona de Humanidades e Tecnologias, Escola de Psicologia e Ciências da Vida.

GAVA, L. L.; PELISOLI, C.; DELL'AGLIO, D. DALBOSCO. A perícia psicológica em casos de suspeita de abuso sexual infanto-juvenil. *Avaliação Psicológica*, Itatiba, v. 12, n. 2, p. 137-145, 2013. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712013000200005&lng=pt&nrm=iso.

GOODMAN, G. S.; QUAS, J. A.; OGLE, C. M. Child Maltreatment and Memory. *Annual Review of Psychology*, v. 61, p. 325-51, 2009.

GORASSINI, D. R.; HARRIS, J. A.; DIAMOND, A.; FLYNN-DASTOOR, E. Computer assessment of interrogative suggestibility. *Personality and Individual Differences*, v. 40, p. 569-577, 2006. DOI: 10.1016/j.paid.2005.08.003.

GUDJONSSON, G. H. A. A New Scale Of Interrogative Suggestibility. *Personality and Individual Differences*, v. 5, n. 3, p. 303-314, 1984.

GUDJONSSON, G. H. A. Psychology of Interrogations and Confessions: a handbook. John Wiley & Sons, Ltd., 2003. Cap.13 e 14, p. 332-412.

GUDJONSSON, G. H.; CLARK, N. K. Sugestionabilidade no interrogatório policial: um modelo psicológico social. *Comportamento Social*, v. 1, n. 2, p. 83–104, 1986.

GUDJONSSON, G. H. Interrogative suggestibility: Its relationship with assertiveness, social-evaluative anxiety, state anxiety and method of coping. *British Journal of Clinical Psychology*, v. 27, n. 2, p. 159-166, 1988. DOI: 10.1111/j.2044-8260.1988.tb00764.x.

GUDJONSSON, G. H. Interrogative suggestibility: factor analysis of the Gudjonsson Suggestibility Scale (GSS 2). *Personality and Individual Differences*, v. 13, n. 4, p. 479-481, 1992.

GUDJONSSON, G.; VAGNI, M.; MAIORANO, T.; PAJARDI, D. The relationship between trauma symptoms and immediate and delayed suggestibility in children who have been sexually abused. *Journal of Investigative Psychology and Offender Profiling*, 2020, p. 1–14. DOI: 10.1002/jip.1554.

GUDJONSSON, G. H.; LISTER, S. Interrogative Suggestibility and its Relationship with Self-Esteem and Control. *Journal of the Forensic Science Society*, v. 24, n. 2, p. 99–110, 1984. DOI: 10.1016/s0015-7368(84)72302-4.

GUDJONSSON, G. H.; RUTTER, S. C.; CLARE, I. C. H. The relationship between suggestibility and anxiety among suspects detained at police stations. *Psychological Medicine*, v. 25, n. 4, p. 875, 1995. DOI: 10.1017/s0033291700035133.

HAIR JR., J. F.; HULT, G. T. M.; RINGLE, C. M.; SARSTEDT, M. *A primer on partial least squares structural equation modeling (PLS-SEM)*. 1 ed. Thousand Oaks (CA): Sage Publications, 2021.

HARALDSSON, E. Interrogative suggestibility and its relationship with personality, perceptual defensiveness and extraordinary beliefs. *Personality and Individual Differences*, v. 6, n. 6, p. 765-167, 1985.

HALLIGAN, P. W.; OAKLEY, D. A. Hypnosis and beyond: Exploring the broader domain of suggestion. *Psychology of Consciousness: Theory, Research, and Practice*, v. 1, n. 2, p. 105–122, 2014. DOI: 10.1037/cns0000019.

HANSEN, I.; SMEETS, T.; JELICIC, M. Further data on interrogative suggestibility and compliance scores following instructed malingering. *Legal and Criminological Psychology*, v. 15, n. 2, p. 221-228, 2010. Disponível em: <https://doi.org/10.1348/135532509x447796>.

HENRIQUES, C. G. P.; POMPEU, J. C. Sugestionabilidade e depoimento infantil: protocolos de entrevista para minimização de falsas memórias. *Revista Diálogos Interdisciplinares*, v. 9, n. 2, 2020. ISSN 2317-3793.

HIGGINS, J. P. T.; THOMPSON, S. G.; DEEKS, J. J.; ALTMAN, G. D. Measuring inconsistency in meta-analyses. *BMJ*, v. 327, p. 132-35, 2003.

HIGGINS, J. P. T.; THOMAS, J.; CHANDLER, J.; CUMPSTON, M.; LI, T.; PAGE, M. J.; WELCH, V. A. (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2 ed. Chichester (UK): Cochrane, 2022. Disponível em: <www.training.cochrane.org/handbook>.

HOSCHEIDT, S. M.; DONGAONKAR, B.; PAYNE, J.; NADEL, L. Emotion, Stress, and Memory. *The Oxford Handbook of Cognitive Psychology*. Psychology, Cognitive Psychology Online, 2013. DOI: 10.1093/oxfordhb/9780195376746.013.0035.

JANOSON, M. A.; FRUMKIN, B. Review of the Gudjonsson Suggestibility Scales. Forensic Psychology, New York, NY, Clinical Psychology Associates, South Miami, 1997.

KASSIN, S. M.; WRIGHTSMAN, L. The psychology of Evidence and Trial Procedure. Beverly Hills: Sage Publications, 1985.

KĘKUŚ, M.; DZIUBAŃSKA, R.; KOMEŻA, I.; DUDEK, I.; CHYLIŃSKA, K.; SZPITALAK, M.; POLCZYK, R. Observation inflation and interrogative suggestibility: Different but related memory errors. *Polish Psychological Bulletin*, v. 51, n. 3, p. 219–225, 2020.

KLEMFUSS, J. Z.; OLAGUEZ, A. P. Individual Differences in Children's Suggestibility: An Updated Review. *Journal of Child Sexual Abuse*, 2018. DOI: 10.1080/10538712.2018.1508108.

KRASINSKI, K.; TONELLI, H. Neuropsicologia da sugestionabilidade e tomadas de decisão social. *Pluralidades em Saúde Mental*, Curitiba, v. 7, n. 1, p. 43-62, jan./jun. 2018.

LINTON, C. P.; SHEEHAN, P. W. A relação entre sugestionabilidade interrogativa e suscetibilidade à hipnose. *Australian Journal of Clinical and Experimental Hypnosis*, v. 22, p. 53-53, 1994.

LOFTUS, E. F.; PALMER, J. C. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning & Verbal Behavior*, v. 13, n. 5, p. 585–589, 1974. DOI: 10.1016/S0022-5371(74)80011-3.

LOFTUS, E. F. Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, v. 12, n. 4, p. 361–366, 2005. DOI: 10.1101/lm.94705.

MANDLER, G. Recognizing: The judgment of previous occurrence. *Psychological Review*, v. 87, n. 3, p. 252–271, 1980. DOI: 10.1037/0033-295X.87.3.252.

MATOS, D. A. S.; RODRIGUES, E. C. *Análise fatorial*. Brasília: Enap, 2019.

MERCKELBACH, H.; MURIS, P.; WESSEL, I.; VAN KOPPEN, P. J. The Gudjonsson Suggestibility Scale (GSS): Further data on its reliability, validity, and metacognition correlates. *Social Behavior and Personality*, v. 26, n. 2, p. 203-210, 1998. DOI: 10.2224/sbp.1998.26.2.203.

MILNE, R.; BULL, R. Back to basics: A componential analysis of the original cognitive interview mnemonics with three age groups. *Applied Cognitive Psychology*, v. 16, p. 743-753, 2002. DOI: 10.1002/acp.825.

MILNE, R.; CLARE, I. C. H.; BULL, R. Interrogative suggestibility among witnesses with mild intellectual disabilities: The use of an adaptation of the GSS. *Journal of Applied Research in Intellectual Disabilities*, v. 15, n. 1, p. 8-17, 2002. DOI: 10.1046/j.1360-2322.2001.00096.x.

OAKLEY, D. A.; HALLIGAN, P. W. Hypnotic suggestion: opportunities for cognitive neuroscience. *Nature Reviews Neuroscience*, v. 14, n. 8, p. 565–576, 2013. DOI: 10.1038/nrn3538.

OKADO, Y. Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory*, v. 12, n. 1, p. 3-11, 2005. DOI: 10.1101/lm.87605.

OTGAAR, H.; SCHELL-LEUGERS, J. M.; HOWE, M. L.; VILAR, A. D. L. F.; HOU-BEN, S. T. L.; MERCKELBACH, H. The link between suggestibility, compliance, and false confessions: A review using experimental and field studies. *Applied Cognitive Psychology*, v. 35, p. 445–455, 2021. DOI: 10.1002/acp.3788.

PACICO, J. C.; HUTZ, C. S.; SCHNEIDER, A. M. de A.; BANDEIRA, D. R. Validade. In: HUTZ, C. M.; BANDEIRA, D. R.; TRENTINI, C. M. [orgs.]. *Psicometria*. Porto Alegre: Artmed, 2015.

PARRIS, B. A. The Prefrontal Cortex and Suggestion: Hypnosis vs. Placebo Effects. *Frontiers in Psychology*, v. 7, p. 415, 2016. DOI: 10.3389/fpsyg.2016.00415.

PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. Petrópolis, RJ: Vozes, 2013.

PIRES, R.; SILVA, D. R.; FERREIRA, A. S. The Portuguese adaptation of the Gudjonsson Suggestibility Scale (GSS1) in a sample of inmates. *International Journal of Law and Psychiatry*, v. 37, p. 289–294, 2014. DOI: 10.1016/j.ijlp.2013.11.014.

PIRES, R. *Estilos de Personalidade e vulnerabilidade à sugestão no contexto de uma relação interpessoal [Personality styles and interrogative suggestibility]*. 2011. Tese (Doutorado) – Lisbon University, Lisbon, Portugal.

PRINSEN, C. A. C.; MOKKINK, L. B.; BOUTER, L. M.; ALONSO, J.; PATRICK, D. L.; de VET, H. C. W.; TERWEE, C. B. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, v. 27, n. 5, p. 1147-1157, 2018.

POLCZYK, R. Sugestibilidade interrogativa: estabilidade transcultural das propriedades psicométricas e correlacionais das escalas de sugestionabilidade de Gudjons-son. *Personalidade e Diferenças Individuais*, v. 38, n. 1, p. 177-186, 2005. DOI: 10.1016/j.paid.2004.03.018.

PONTEROTTO, J. G.; RUCKDESCHEL, D. E. An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual and Motor Skills*, v. 105, n. 3, p. 997-1014, 2007. DOI: 10.2466/pms.105.3.997-1014.

RAVER, A.; LINDHOLM, T.; GUSTAFSSON, P. U.; ALM, C. Memory accuracy, suggestibility and credibility in investigative interviews with native and non-native eyewitnesses. *Frontiers in Psychology*, v. 14, 2023. DOI: 10.3389/fpsyg.2023.1240822.

REUTEMANN, M. *Suggestibilität: situative und personale Prädiktoren*. Münster: Waxmann Verlag, 2006.

RIBEIRO, A. C. M. A. Sugestibilidade interrogativa em crianças dos 4 aos 9 anos: estudo normativo com o Bonn Test Statement of Suggestibility (BTSS) e análises de variáveis relacionadas com a sugestionabilidade. 2015. Tese (Doutorado em Psicologia). Universidade de Coimbra. Disponível em: <http://hdl.handle.net/10316/26855>.

ROVINSKI, S. L. R.; PELISOLI, C. Violência sexual contra crianças e adolescentes: testemunho e avaliação psicológica. 1ª ed. São Paulo: Vetor, 2019.

ROWSELL, K.; COLLOFF, M. F. Critique of a measure of interrogative suggestibility for children: The Bonn test of statement suggestibility. *Applied Cognitive Psychology*, v. 38, 2024. DOI: <https://doi.org/10.1002>.

SANTOS, A. S. D. Dos. Sugestibilidade interrogativa e o efeito da manipulação da fonte de feedback negativo na GSS1: um estudo com adolescentes. 2017. Dissertação (Mestrado Integrado em Psicologia) - Faculdade de Psicologia e de Ciências da Educação, Universidade de Coimbra, 2017.

SAMPAIO, R. F.; MANCINI, M. C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. *Revista Brasileira de Fisioterapia*, v. 11, n. 1, p. 83-89, 2007.

SANTTILA, P.; EKHOLM, M.; NIEMI, P. The effects of alcohol on interrogative suggestibility: The role of state-anxiety and mood states as mediating factors. *Legal and Criminological Psychology*, v. 4, n. 1, p. 1-13, 1999. DOI: 10.1348/135532599167707.

SANTOS, G. M.; STRATHDEE, S. A.; EL-BASSEL, N.; PATEL, P.; SUBRAMANIAN, D.; HORYNIAK, D.; COOK, R. R.; MCCULLAGH, C.; MAROTTA, P.; CHOKSI, F.; KANG, B.; ALLEN, I.; SHOPTAW, S. Psychometric properties of measures of substance use: a systematic review and meta-analysis of reliability, validity and diagnostic test accuracy. *BMC Medical Research Methodology*, v. 20, n. 1, p. 1-22, 2020.

SARAIVA, M.; ALBUQUERQUE, P. B. Influência da idade, desejabilidade social e memória na sugestionabilidade infantil. *Psicologia: Reflexão e Crítica*, v. 28, n. 2, p. 356-364, jan. 2015.

SCHEMM, K. V.; KRAUS, U.; GUBI-KELM, S.; KÖHNKEN, G. A German version of the forensic Gudjonsson Suggestibility Scale (GSS 1). In: GILES, S. (Ed.). *Psychological aspects of legal processes*. London: IA-IP Publishing, 2006.

SENE, A. S.; LOPES, E. J.; ROSSINI, J. C. F. Falsas memórias e tempo de reação: estudo com o procedimento de palavras associadas. *Psychologica*, p. 23-38, 2014.

SHAO, X.; LI, A.; CHEN, C. et al. Cross-stage neural pattern similarity in the hippocampus predicts false memory derived from post-event inaccurate information. *Nature Communications*, v. 14, 2299, 2023. DOI: <https://doi.org/10.1038/s41467-023-38046-y>.

STEIN, L. M.; PERGHER, G. K. Criando falsas memórias em adultos por meio de palavras associadas. *Psicologia: Reflexão e Crítica*, v. 14, n. 2, p. 353-366, 2001.

STEIN, L. M. Falsas memórias: fundamentos científicos e suas aplicações clínicas e jurídicas. Porto Alegre: Artmed, 2010.

STEIN, L. M.; PERGHER, G. K.; FEIX L. DA F. Desafios da oitiva de crianças no âmbito forense. Brasília-DF: Secretaria Especial dos Direitos Humanos da Presidência da República/Subsecretaria de Promoção dos Direitos da Criança e do Adolescente Programa Nacional de Enfrentamento da Violência Sexual contra Crianças e Adolescentes. Childhood Brasil (Instituto WCF-Brasil. Projeto Culturas e Práticas não Revitimizantes: Reflexão e Socialização de Metodologias Alternativas para Inquirir Crianças e Adolescentes em Processos Judiciais, 77p. 2009.

SOUZA, R.L.S.; CRUZ, S.T.M. 2023. Perícia Psicológica de vítimas e testemunhas de violência. In D. Lino (Org), *Psicologia Investigativa: Teoria e Prática* (pp.55-70).

SOUZA FILHO, M. L. de. Sugestionabilidade: suas características e correlações com outras variáveis psicológicas. Dissertação (Mestrado) Psicologia Cognitiva – Universidade Federal de Pernambuco. Recife, 2007.

TABORDA, J. M. M. A produção de falsas memórias no Paradigma DRM: o efeito da alteração de pistas para o reconhecimento de palavras. Dissertação (Mestre em Psicologia Clínica e da Saúde) - Universidade da Beira Interior Ciências Sociais e Humanas, 2018.

TAVAKOL, M.; DENNICK, R. Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. 2011. DOI: 10.5116/ijme.4dfb.8dfd

TERWEE, C. B.; BOT, S. D. M.; DE BOER, M. R.; VAN DER WINDT, D. A.; KNOL, D. L.; DEKKER, J.; BOUTER, L. M.; DE VET, H. C. W. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1):34-42. 2007.

TRINDADE, J. *Manual de Psicologia Jurídica para operadores do direito*. Livraria do Advogado, 2014.

TOMÉ, M.; FAYET JÚNIOR, N. O valor processual do depoimento (testemunhal ou vitimário) infantil. In: DIAS, F. da V.; BUDÓ, M. De N.; SANTOS, G. F. dos (Orgs.). *Criminologia, violência e controle social*. Porto Alegre, RS: Editora Fi, 2017.

URBINA, S. *Fundamentos da testagem psicológica*. Porto Alegre: Artmed, 2007.

VAGNI, M.; MAIORANO, T.; PAJARDI, D. Effects of post-traumatic stress disorder on interrogative suggestibility in minor witnesses of sexual abuse. *Current Psychology*, 41, 7681–7694, 2022. DOI: <https://doi.org/10.1007/s12144-020-01253-7>

WAGNER, V.; PEIXOTO, E. M.; OLIVEIRA, L. P. de. Propriedades Psicométricas e Adaptação Cultural da Basic Need Satisfaction in General Scale para uma População Brasileira de Usuários de Cadeira de Rodas. *Revista Brasileira de Educação Especial*, Marília, v. 27, e0214, 2021.. Acessos em 04 jan. 2024. Epub 15-Jun-2021. DOI: <https://doi.org/10.1590/1980-54702021v27e0214>.

WATANABE, K.; WACHI, T.; YOKOTA, K.; KURASHI, H.; YUSUKE, O.; ONO, S. The examination of the Gudjonsson Suggestibility Scale 2. *Japanese Journal of Criminal Psychology*, v. 51, p. 56-57, 2013.

WIEDER, L.; BROWN, R.; THOMPSON, T.; TERHUNE, D. B. Suggestibility in functional neurological disorder: a meta-analysis. *J Neurol Neurosurg Psychiatry*. Epub ahead of print. doi:10.1136/jnnp-2020 323706.

YOUNG, K.; POWELL, M. B.; DUDGEON, P. Individual differences in children's suggestibility: a comparison between intellectually disabled and mainstream samples. *Personality and Individual Differences*, v. 35, n. 1, p. 31–49, 2003. doi:10.1016/s0191-8869(02)00138-1.

ZANON, C.; HAUCK FILHO, N. Fidedignidade. In: Hutz, C. M.; Bandeira, D. R.; Trentini, C. M. (org.). *Psicométria*. Porto Alegre: Artmed, 2015.