

**Ariela Raissa Lima Costa**



**O CONTROLE DE DESEJABILIDADE SOCIAL VIA  
DIFERENTES FORMATOS DE RESPOSTA: AVALIAÇÃO DA  
TRÍADE SOMBRIA**

**Apoio:**



**CAMPINAS**

**2020**

**Ariela Raissa Lima Costa**

**O CONTROLE DE DESEJABILIDADE SOCIAL VIA  
DIFERENTES FORMATOS DE RESPOSTA: AVALIAÇÃO DA  
TRÍADE SOMBRIA**

Projeto de pesquisa apresentado ao Programa de Pós-Graduação *Stricto Sensu* em Psicologia da Universidade São Francisco, Área de Concentração - Avaliação Psicológica, para obtenção do título de Doutor.

ORIENTADOR: PROF. DR. NELSON HAUCK FILHO

CAMPINAS

2020

## PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU EM PSICOLOGIA

Ariela Raissa Lima Costa defendeu a tese **“O CONTROLE DE DESEJABILIDADE SOCIAL VIA DIFERENTES FORMATOS DE RESPOSTA: AVALIAÇÃO DA TRÍADE SOMBRIA”** aprovada pelo Programa de Pós-Graduação Stricto Sensu em Psicologia da Universidade São Francisco em 10 de julho de 2020 pela Banca Examinadora constituída por:

Prof. Dr. Nelson Hauck Filho  
Orientador e Presidente

Prof. Dr. Felipe Valentini  
Examinador

Prof. Dr. Lucas de Francisco Carvalho  
Examinador

Prof. Dr. Josemberg Moura de Andrade  
Examinador

Prof. Dr. Wagner de Lara Machado  
Examinador

## PROTOCOLO DE ASSINATURA(S)

O documento acima foi proposto para assinatura digital na plataforma Portal de Assinaturas Certisign. Para verificar as assinaturas clique no link: <https://www.portaldeassinaturas.com.br/Verificar/596D-DD32-3B4F-6B7F> ou vá até o site <https://www.portaldeassinaturas.com.br:443> e utilize o código abaixo para verificar se este documento é válido.

**Código para verificação: 596D-DD32-3B4F-6B7F**



### Hash do Documento

17467F22D06D06B9ED930C8A42560DA35C38D1066B36AEC4478296D97556D77B

O(s) nome(s) indicado(s) para assinatura, bem como seu(s) status em 23/07/2020 é(são) :

- Josemberg Moura de Andrade (Signatário) - 024.182.104-56 em 23/07/2020 15:07 UTC-03:00

**Tipo:** Assinatura Eletrônica

**Identificação:** Por email: josemberg.andrade@gmail.com; SMS: +5561991160751

### Evidências

**Client Timestamp** Thu Jul 23 2020 15:07:19 GMT-0300 (Horário Padrão de Brasília)

**Geolocation** Latitude: -15.725935400000001 Longitude: -47.8526193 Accuracy: 4077

**IP** 189.6.27.9

**Assinatura:**

### Hash Evidências:

E5EA9F05DE6EB1B2BFE707AF0EA79DEAABE86B504686996EDADB85AFD0DD8815

- Wagner de Lara Machado (Signatário) - 005.120.140-27 em 22/07/2020 13:40 UTC-03:00

**Tipo:** Assinatura Eletrônica

**Identificação:** Por email: wag.lm.psico@gmail.com; SMS: +5551999685878

### Evidências

**Client Timestamp** Wed Jul 22 2020 13:40:03 GMT-0300 (Horário Padrão de Brasília)

**Geolocation** Latitude: -30.0676918 Longitude: -51.055101199999996 Accuracy: 30

**IP** 189.6.151.45

**Assinatura:**



**Hash Evidências:**

106D839C07F4027FAE26848B9E9C37A44352F9515C8AFCE956F3B2636A172F02

- Lucas de Francisco Carvalho (Signatário) - 319.218.918-58 em 22/07/2020 13:23 UTC-03:00

**Tipo:** Assinatura Eletrônica

**Identificação:** Por email: lucas.carvalho@usf.edu.br; SMS: +5511981028284

**Evidências**

**Client Timestamp** Wed Jul 22 2020 13:22:57 GMT-0300 (Horário Padrão de Brasília)

**Geolocation** Location not shared by user.

**IP** 187.87.250.226

**Assinatura:**



**Hash Evidências:**

8A3E9F9C5BF0F66C1CFA8EA16C607DDC2777F56281A6F4AD4F13D95C876CEDFB

- Nelson Hauck Filho (Signatário) - 010.833.960-21 em 21/07/2020 17:19 UTC-03:00

**Tipo:** Assinatura Eletrônica

**Identificação:** Autenticação de conta; SMS: +5511956553005

**Evidências**

**Client Timestamp** Tue Jul 21 2020 17:19:12 GMT-0300 (Brasilia Standard Time)

**Geolocation** Latitude: -22.857387499999998 Longitude: -47.0536161 Accuracy: 40

**IP** 201.82.49.196

**Assinatura:**



**Hash Evidências:**

B3E5B779384E813CB2C5C47D68467AB3F1C584B2D7883C503DD90A6FAE47D3B0

Felipe Valentini (Signatário) - 004.475.570-84 em 20/07/2020 13:47 UTC-03:00

**Tipo:** Assinatura Eletrônica

**Identificação:** Por email: felipe.valentini@usf.edu.br; SMS: +5521999925634

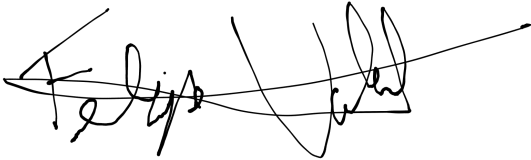
### Evidências

**Client Timestamp** Mon Jul 20 2020 13:47:56 GMT-0300 (Horário Padrão de Brasília)

**Geolocation** Latitude: -22.857803699999998 Longitude: -47.048581 Accuracy: 955

**IP** 187.106.43.162

**Assinatura:**

A handwritten signature in black ink, appearing to read 'Felipe Valentini', with a long horizontal stroke extending to the right.

**Hash Evidências:**

DB106C092CC3C4A92D3A2D9FE74D3C4F08757DE0659BA7A2096BF3DB189B35E6



## Agradecimentos

Antes de qualquer pessoa que está a meu redor, eu quero agradecer ao mundo espiritual que é invisível aos meus olhos, mas cujas mãos sempre guiaram meus passos, me acalentaram nos momentos que pensei que nada fosse dar certo, que me mostraram como superar meus medos e ansiedades. Obrigada aos meus orixás Iemanjá, Oxum e Okolum que me banharam em suas águas abençoadas. Obrigada a meus pretos velhos, Pai João de Aruanda, Preto Velho de Angola, Preto Velho das Águas, Vovó Marilu das Cachoeiras, Vovó Candinha por sua sabedoria, por me ensinarem a importância da disciplina e pelo amor de vocês. Agradeço aos exus, meu padrinho Tranca Rua das Almas, meu pai Exú Tiriri, Exú Zé Pilintra, as pombas, Dona Sete Saias das Sete Encruzilhadas, Dona Maria Molambo do Lixo, minha mãe Maria Padilha e ao Exú Mirim Luiz Arruda por me ensinarem que nem tudo que está na luz é bondade, nem tudo na escuridão é maldade, que gratidão é o melhor presente. Obrigada aos ibejis, Cachoerinha, Pequeno Pajé, minha linda Joaquina, por ensinar a aproveitar as pequenas coisas da vida e aprender a sorrir para a vida. Agradeço aos caboclos, Caboclo das Matas Virgens, meu pai Caboclo das Sete Flechas por me ensinarem a ser forte. Agradeço aos léguas, Seu Léguas Bugibua, por sua sabedoria e ensinamentos sobre os caminhos da vida. Agradeço a corrente cigana, a Cigana Esmeralda, Cigana Soraya, minha mãe Cigana Sulamyta por afastarem os males com sua sabedoria, danças, sorrisos e beleza.

No plano terreno, agradeço aos anjos que a espiritualidade colocou em meu caminho na forma de família. Obrigada mãe, Ruthnéia Vieira Lima, por seu apoio incondicional e pela bolsa-mãe, obrigada por nunca me deixar desistir, por sempre me amar e me guiar na vida, obrigada por aceitar a missão de ser minha mãe. Obrigada pai, Luiz Neto, por me amar, por sempre acreditar no meu potencial e dizer que eu só podia voltar para casa se fosse doutora, e obrigada por ter aceitado a missão de ser meu pai. Obrigada minha madrinha, Osana, que a

espiritualidade colocou em nossos caminhos, por nunca deixar eu pensar menos do que o melhor de mim, por sempre “adivinhar” quando eu mais precisava de você, obrigada por ter aceitado ser minha madrinha, minha segunda mãe na Terra (não que eu tenha te dado escolha). Obrigada a minha irmã, Amanda, minha inteligente e linda Pocahontas, por me mostrar a beleza da vida, por seu orgulho e suporte como irmã mais velha. Obrigada a meu irmão, Luiz Filho, meu príncipe que sempre chorou em cada despedida, que demonstra seu amor e apoio diariamente. Amo vocês!

Também quero agradecer a família que ganhei no Programa de Pós-Graduação. Agradeço a meu orientador, Nelson Hauck Filho, por me ajudar desde o mestrado, me explicando regressão mil vezes (juro que já aprendi!), por sua paciência, suas palavras motivadoras e pela preocupação (prometo não mandar mais e-mail com o título de Socooooooooorro), prometo continuar enchendo sua paciência. Agradeço ao meu orientador na Universidade de Lund, Martin Bäckström, pelo apoio em todos os momentos, por ouvir minhas ideias e por me receber. Agradeço a todos os professores, que se tornaram mais que professores, se tornaram amigos, obrigada Fabián, Ana Paula, Acácia, Rodolfo, Lucas, Ricardo, Makilim e ao meu quase coorientador, Felipe. Agradeço a vocês pelos conselhos, pelas aprendizagens, paciência e gargalhadas.

Meus irmãos de pesquisa, de vida e de escolha, também são e foram essenciais nessa família da pós-graduação e na jornada de se tornar doutora. Obrigada minha companheira de apartamento, Ana Deyvis, por suportar minhas chatices, por ouvir todas as fofocas de famosos quando eu estava com preguiça de estudar, por me dizer que tudo ia dar certo. Obrigada Catarina por sempre ter tempo de me ouvir, de me ajudar nas descobertas das análises, pelos sorrisos (não vamos comentar Belo Horizonte, kkkkkkkkkkk). Obrigada, minhas maravilhosas Leilane e Fernanda, por me aguentarem, por literalmente me darem um teto quando resolvi ir para o outro lado do mundo, pelos cafés da manhã deliciosos, pelo cuidado, amor e risadas,



vocês foram amigas que demorei a descobrir, primeiro nos detestamos, mas foi só um prólogo do que seria esse grande amor e amizade em que surgimos uma na vida das outras no momento certo. Obrigada a meus irmãos de orientação, ‘As Mina da Tríade’ (para referência é nome do grupo no WhatsApp), Bruno, meu feitor acadêmico que sempre me botou para estudar mesmo com preguiça ou com febre (não me dava liberdade por nada!! A gente reclama porque a gente gosta!), Samanta e Natália pelo suporte mútuo, pelas estratégias para falar com o ‘ori’ (Nelson, não se preocupe, gostamos muito de você!). Um agradecimento especial ao meu irmão de orientação, Pedro Cortez (Pedrito), que pareceu tão estranho no primeiro olhar (não que tenha mudado, kkkkk), mas que insistiu que seria meu amigo e hoje me orgulho em chamá-lo de amigo e irmão, obrigada por não desistir de mim! Aos meus amigos Lucas, Gabriela, Felipe, Ana Celi, Mayara, Leonardo, Adriana pelas gargalhadas na pós, pelas pipocas e chocolates. Amo vocês, que nossa amizade se prolongue pela eternidade. Obrigada minha amiga Taís, que eu atormento toda vez que estou ansiosa querendo gastar o que não tenho, obrigada pelos seus conselhos, por ser minha pessoa, te amo!!! Agradeço aos amigos que fiz na Suécia, Aslihan, Mei, Meng, Nathaniel e Rebecka por me ajudarem a me adaptar em uma cultura diferente, pelos sábados visitando todas as cafeterias de Lund.

Um agradecimento todo especial para meu professor e amigo, Lucas Guimarães, quem me apresentou e quase me obrigou ir fazer a prova de mestrado na USF, que disse que só me aceitava se fosse doutora, deu certo, meu lindo! Obrigada, Lucas, por sempre torcer pelo meu sucesso e me empurrar para os caminhos certos, obrigada a espiritualidade por ter me dado você de presente! Te amo!!!

O projeto de me tornar doutora começou aos 12 anos, quando descobri o que era esse negócio ‘doutorado’ e disse para minha mãe que me tornaria doutora até os 27 anos. Deu certo, galera!!! Obrigada por todos que me ajudaram a tornar esse sonho realidade!

*“O melhor caminho é aquele que construímos”*

(Preto Velho de Angola)

Costa, A. R. L. (2020) *O controle da desejabilidade social via diferentes formatos de resposta: Avaliação da tríade sombria*. (Tese de Doutorado). Campinas: Universidade São Francisco.

## Resumo

A avaliação por meio do autorrelato apresenta diversas vantagens, demandando um treinamento mais simples tanto para a aplicação quanto para a correção dos escores. Entretanto, esse formato também tem suas desvantagens, e uma que se destaca é a vulnerabilidade a vieses de resposta, em especial viés de desejabilidade social. Esse viés acontece quando a pessoa responde de forma exageradamente positiva ou otimista sobre si mesmo, exacerbando suas qualidades pessoais ou omitindo suas fraquezas. Diversas características contribuem para a existência desse viés: (a) o contexto de avaliação: alguns situações (ex., uma entrevista de emprego) eliciam mais esse tipo de comportamento do que outras; (b), os itens: alguns itens podem favorecer respostas socialmente desejáveis mais do que outros (ex., itens que empregam palavras com carga valorativa extremamente positiva ou negativa); e (c) o formato de resposta: a maneira como o indivíduo é instruído a pontuar o item pode favorecer a ocorrência de respostas socialmente desejáveis (escalas tipo Likert são mais fáceis de serem falseadas do que itens de escolha forçada); (d) características da pessoa: alguns indivíduos podem estar mais propensos do que outros a responder de forma socialmente desejável. Enquanto o primeiro e o último elemento são mais difíceis de controlar, o segundo e o terceiro apresentam algumas possibilidades. Com isso, este trabalho teve por objetivo avaliar a relação entre o formato de resposta dos itens, formato tipo Likert versus escolha forçada, e a ocorrência do viés de desejabilidade social em escalas que avaliam a tríade sombria da personalidade usando esses dois tipos de métodos de respostas. Para isso, foram realizados três estudos, com 1.057 pessoas, com mais de 18 anos, sendo a maioria mulheres com nível superior. O primeiro estudo apresenta aspectos estruturais necessários para construir escalas com formato de resposta de em escolha forçada e Likert, indicando uma complexidade maior do que a esperada na seleção das opções de resposta. O segundo estudo apresenta os passos envolvidos na construção de escalas de autorrelato, desde investigação teórica até a escolha dos itens para a versão final, indicando que existe uma dificuldade natural na elaboração de itens discriminativos de cada fator, especialmente Maquiavelismo e Psicopatia. O terceiro estudo apresenta estudos sobre a validação dos instrumentos construídos no Estudo 2, incluindo investigação da estrutura interna, consistência interna e correlação entre os fatores encontrados, os resultados sugerem que a escala em formato de escolha forçada discriminou melhor os fatores da Tríade e relação negativa com os fatores de desejabilidade social, indicando que o método de escolha forçada pode ser usado para a avaliação de traços indesejados, ainda que demande mais tempo do pesquisador.

**Palavras-chaves:** Likert; escolha forçada; manipulação de instrução; viés moralista; viés egoísta.

Costa, A. R. L. (2020) *The control of social desirability via different response formats: Assessment of Dark Triad traits*. (Doctoral Dissertation). Campinas: Universidade São Francisco.

### Abstract

A self-report assessment has several advantages, requiring less straightforward training for both application and correction of scores. However, this format also has its disadvantages, and one that stands out is its vulnerability to response bias, especially social desirability bias. This bias happens when the person responds in an exaggeratedly positive or optimistic way about herself/himself, exacerbating her/his personal qualities, or omitting weaknesses. Several characteristics contribute to the existence of this bias: (a) the context of evaluation: some situations (e.g., a job interview) elicit this type of behavior more than others; (b), the items: some items may favor socially desirable responses more than others (e.g., items that use words with an extremely positive or negative value); and (c) the response format: how individuals are instructed to score the item (Likert scales are easier to falsify than items of forced-choice); (d) person characteristics: some individuals may be more likely than others to respond in a socially desirable way. While the first and last elements are more difficult to control, the second and third have some possibilities. Thus, this work aimed to evaluate the relationship between the items' response format, Likert-type format versus forced-choice, and the occurrence of social desirability bias on scales that assess the dark triad personality using these two types of answers. For this, three studies were carried out, with 1,057 people, over 18 years old, most of them women with higher education. The first study presents structural aspects necessary to build scales with a forced-choice and Likert response format, indicating a complexity greater than expected in the selection of response options. The second study presents the steps involved in the construction of self-report scales, from theoretical investigation to the choice of items for the final version, indicating that there is an inherent difficulty in the elaboration of discriminating items for each factor, especially Machiavellianism and Psychopathy. The third study presents investigations on the validation of the instruments built in Study 2, including investigation of the internal structure, internal consistency, and correlation between the factors found. The results suggest that the scale in a forced-choice format better discriminated between the factors of the Triad and had a negative correlation with the factors of social desirability, indicating that the forced-choice method can be used for the assessment of undesirable traits, even though it demands more time from the researcher.

**Keywords:** Likert; forced choice; manipulation of instruction; moralistic bias; selfish bias.

Costa, A. R. L. (2020). *Control de la deseabilidad social a través de diferentes formatos de respuesta: Evaluación de la tríada oscura*. (Tesis de Doctorado). Campinas: Universidade São Francisco.

### Resumen

Una evaluación de autoinforme tiene varias ventajas, ya que requiere una capacitación menos directa tanto para la aplicación como para la corrección de puntajes. Sin embargo, este formato también tiene sus desventajas, y una que destaca es su vulnerabilidad al sesgo de respuesta, especialmente el sesgo de deseabilidad social. Este sesgo ocurre cuando la persona responde de manera exageradamente positiva u optimista sobre sí misma, exacerbando sus cualidades personales u omitiendo debilidades. Varias características contribuyen a la existencia de este sesgo: (a) el contexto de la evaluación: algunas situaciones (por ejemplo, una entrevista de trabajo) provocan este tipo de comportamiento más que otras; (b), los ítems: algunos ítems pueden favorecer respuestas socialmente deseables más que otros (por ej., ítems que usan palabras con un valor extremadamente positivo o negativo); y (c) el formato de respuesta: cómo se instruye a los individuos para calificar el ítem (las escalas Likert son más fáciles de falsificar que los ítems de elección forzada); (d) características de la persona: algunas personas pueden ser más propensas que otras a responder de una manera socialmente deseable. Mientras que el primer y el último elemento son más difíciles de controlar, el segundo y el tercero tienen algunas posibilidades. Por lo tanto, este trabajo tuvo como objetivo evaluar la relación entre el formato de respuesta de los ítems, el formato tipo Likert versus la elección forzada, y la ocurrencia de sesgo de deseabilidad social en escalas que evalúan la personalidad de la tríada oscura usando estos dos tipos de respuestas. Para ello, se realizaron tres estudios, con 1.057 personas, mayores de 18 años, la mayoría mujeres con educación superior. El primer estudio presenta los aspectos estructurales necesarios para construir escalas con un formato de elección forzada y respuesta Likert, lo que indica una complejidad mayor de lo esperado en la selección de opciones de respuesta. El segundo estudio presenta los pasos involucrados en la construcción de escalas de autoinforme, desde la investigación teórica hasta la elección de ítems para la versión final, indicando que existe una dificultad inherente en la elaboración de ítems discriminatorios para cada factor, especialmente el maquiavelismo y la psicopatía. El tercer estudio presenta investigaciones sobre la validación de los instrumentos construidos en el Estudio 2, incluida la investigación de la estructura interna, la consistencia interna y la correlación entre los factores encontrados. Los resultados sugieren que la escala en un formato de elección forzada discriminó mejor entre los factores de la Tríada y tuvo una correlación negativa con los factores de conveniencia social, lo que indica que el método de elección forzada puede usarse para la evaluación de rasgos indeseables, incluso aunque exige más tiempo del investigador.

**Palabras claves:** Likert; elección forzada; manipulación de la instrucción; sesgo moralista; sesgo egoísta.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. (This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001).

## Sumário

Introdução .....	2
Características contextuais.....	5
Características dos itens.....	9
Características dos formatos de resposta .....	12
Características pessoais.....	15
Objetivos.....	<b>Erro! Indicador não definido.</b>
Construindo escalas de autorrelato: o que fazer? (Artigo 1).....	21
Desafios e alternativas na construção de escalas de autorrelato: Likert e escolha-forçada (Artigo 2).....	53
Measuring Dark traits using single and multiple stimulus strategies (Artigo 3) .....	84
Considerações finais .....	113
Referências.....	114

## Introdução

A vida social é organizada por normas que atuam como reguladoras das relações sociais. Tais normas, por sua vez, se pautam em valores que norteiam as escolhas dos sujeitos para se comportarem de uma maneira específica. Os principais fatores delineados pela literatura são os individuais (*agency*) e comunitários (*communion*). Os valores individuais se relacionam ao sujeito se destacar do grupo e, conseqüentemente, sobrepor os próprios interesses em detrimento do grupo, buscando poder que garanta sua diferenciação. Os valores comunitários se relacionam a indivíduos que se veem como parte de um conjunto e, conseqüentemente, buscam a união e intimidade com o outro (Wiggins, 1991).

Tais valores influenciam não somente nas interações sociais diretas, mas também em situações sociais com interação indireta. Um exemplo de interação indireta é quando o sujeito é requisitado a se auto avaliar a fim de conquistar um benefício pessoal, como em seleções de emprego ou admissão em universidade. Essa autoavaliação pode ser feita de forma verbal, por meio de uma entrevista (interação direta), ou por meio de respostas a instrumentos de autorrelato (interação indireta). Nas duas modalidades, o sujeito deve pesar o quanto quer revelar de si mesmo e o quanto ser totalmente verdadeiro poderá ser uma vantagem. Logo, situações de testagem cujo resultado é importante para o sujeito, um sentimento de autopreservação pode levá-lo a escolhas de respostas pouco ou nada acuradas. O foco deste estudo está na decisão que o sujeito faz quando responde a instrumentos de autorrelato, se ele opta pela honestidade ou se opta por se autoprotger manipulando as informações fornecidas da forma como acha mais desejável. Respostas que são manipuladas positivamente a fim de as enquadrar nas expectativas sociais percebidas pelo sujeito são chamadas de respostas socialmente desejáveis.



A desejabilidade social (DS) refere-se a comportamentos que são percebidos como esperados, aceitáveis e importantes considerando os valores de uma dada cultura. Desta feita, as respostas manipuladas nessa direção são chamadas de respostas socialmente desejáveis (RSD). Em instrumentos de autorrelato, as respostas socialmente desejáveis são vistas como uma fonte de variância não-aleatória, que não é o atributo de interesse e que afeta os escores nos testes. Assim, podem comprometer a validade de instrumentos de autorrelato, uma vez que mascaram o escore verdadeiro da pessoa (Georg Krammer, Sommer, & Arendasy, 2017).

Estudos de metanálise demonstraram que RSD inflam a média dos escores em pesquisas que manipularam as instruções (Viswesvaran & Ones, 1999) e em processos reais de seleção de emprego (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). Na metanálise realizada por Viswesvaran e Ones (1999), eles analisaram o resultado de tamanho de efeito de 51 estudos, não sendo identificada variabilidade no falseamento das respostas por dimensões da personalidade avaliadas (Cinco Grandes Fatores), sendo que todos os fatores foram igualmente manipulados. No estudo de metanálise realizado por Birkeland et al. (2006) eles analisaram o desempenho de candidatos em uma seleção de empregos e de não-candidatos em 33 estudos em que verificaram que os candidatos pontuaram mais alto que os não-candidatos nos traços de conscienciosidade ( $d = 0,45$ ) e estabilidade emocional ( $d = 0,44$ ), sugerindo que em situações reais de testagem o traço a ser avaliado é manipulado de acordo com a percepção do sujeito da importância das características desse traço na obtenção de um resultado positivo.

Estudos posteriores encontraram resultados semelhantes (Donovan & Dwright, 2014; Ziegler & Buehner, 2009). Por exemplo, no estudo de Ziegler e Buehner (2009), ao testarem um modelo de controle de respostas socialmente desejáveis verificaram a influência desse tipo de resposta na média e na estrutura de covariância de um instrumento que avaliava os cinco grandes fatores. Enquanto no estudo de Donovan e Dwright (2014),

eles comparam um grupo de pessoas durante o processo de seleção de emprego e seis meses após a contratação. Os resultados indicaram que 49.7% dos candidatos falsearam suas respostas a um instrumento que avaliava orientação de meta (*goal orientation*). Além disso, os resultados também mostraram alteração nos valores de consistência interna - na aplicação na seleção variou entre 0,80 e 0,86 e, seis meses após a consistência interna variou entre -0,16 e 0,57. Todos esses estudos favorecem a perspectiva da influência negativa que as respostas socialmente desejáveis têm sobre as propriedades psicométricas de instrumentos de autorrelato.

Entretanto, isso não é um consenso na literatura, alguns autores apresentam evidências em sentido contrário e consideram espúria a influência da desejabilidade social em autorrelato de personalidade (McGrath, Mitchell, Kim, & Hough, 2010; Morey, 2012; Ones, Viswesvaran, & Reiss, 1996). Ainda que não haja um consenso sobre o real efeito desse viés em autorrelato de personalidade, no presente estudo assumimos que as respostas socialmente desejáveis são um viés de resposta que sofrem influência de vários aspectos e que podem ser parcialmente controladas.

Uma das influências primárias desse viés é a forma como a pessoa processa as características avaliadas pelo instrumento (Paulhus, 1984). Esse processamento pode ser consciente ou não-consciente. No primeiro, a pessoa, propositadamente, maneja suas respostas de forma a parecer ser melhor do que é; no segundo, a pessoa se auto engana e, de fato, acredita ter altos níveis de características consideradas desejáveis.

Os valores e a forma como a pessoa processa o conteúdo avaliado serviram como base para o desenvolvimento de um modelo teórico que busca compreender a desejabilidade social (Paulhus & John, 1998). Nesse modelo a desejabilidade social engloba duas dimensões, chamadas de viés egoísta e moralista. O primeiro está diretamente relacionado aos valores individuais e, o segundo, aos valores comunitários. Cada dimensão, por sua vez, é constituída por duas facetas, uma relacionada ao

processamento consciente e outra ao processamento não-consciente. As facetas do viés egoísta são a administração egoísta da conduta (consciente) e autoengano positivo (não-consciente). De forma geral, esse viés é caracterizado por pessoas que se acham excepcionalmente inteligentes, talentosas, corajosas, confiantes e sociáveis e que tendem a enaltecer quaisquer mínimas qualidades positivas que possuem. As facetas do viés moralista são administração moralista da conduta (consciente) e autoengano negativo (não-consciente). No geral, esse viés é caracterizado por pessoas que se acham exemplos de cidadãos, seguidores das normas sociais de forma rígida, que nunca comentem erros e tendem a negar quaisquer mínimas qualidades negativas (Paulhus, 2002).

Esse modelo teórico explica o conteúdo e as motivações de respostas socialmente desejáveis. Entretanto, para que elas ocorram uma série de elementos influenciam na decisão do sujeito. As RSD ocorrem por meio da interação de características da pessoa, do contexto de avaliação, do item e do instrumento. Assim, quando a pessoa se depara com um item, ela avalia o conteúdo e o compara com os valores da sociedade na qual está inserida, considera as consequências de sua resposta, se existe a necessidade e a possibilidade de falsear de forma a parecer ter características em níveis mais positivos do que de fato tem. Portanto, para RSD acontecerem, o respondente precisa ter valores que tornem aceitáveis a alteração de suas respostas, um contexto de alta motivação (e.g. seleção de emprego), o conteúdo do item ser considerado desejável, o formato de resposta do instrumento aparentar ser fácil de ser manipulado e, por fim, acreditar ter competência para manipular suas respostas (Ellingson, 2012). Para fins didáticos, tais aspectos serão detalhados separadamente nos próximos tópicos.

### **Características contextuais**

Responder a um instrumento de autorrelato que avalia personalidade pode se comparar a responder ao mesmo item diante de uma audiência, no caso uma audiência

composta por avaliadores (Hogan, 1991). Algumas audiências são mais ameaçadoras, por exemplo, responder a um teste, em sala de aula, como parte de uma pesquisa aleatória é menos ameaçador que responder ao mesmo instrumento como parte de um processo seletivo. Na primeira situação, as consequências das respostas têm pouco ou nenhum impacto na vida da pessoa o que acarreta baixa motivação para falsear respostas. Na segunda situação, por outro lado, a pessoa, a depender do resultado, tem muito a ganhar ou perder o que acarreta alta motivação para falsear respostas. Com isso, o contexto de avaliação é um elemento que contribui para a possibilidade de respostas socialmente desejáveis.

Contextos de avaliações compulsórias (e.g., seleção de emprego, porte de arma, psicotécnico) são uma das principais preocupações dos pesquisadores em relação ao efeito da desejabilidade social nas propriedades psicométricas de instrumentos de autorrelato de personalidade. Isso se deve ao fato de que tais avaliações acarretam resultados importantes para os examinandos, seja conseguir um emprego, receber carteira de motorista ou adquirir direito a porte de arma. Avaliações desse tipo tendem a ser mais propensas ao falseamento, já que são contextos de alta motivação. Para identificar o efeito de RSD, estudos comparam as respostas a testes aplicados em contextos de baixa e alta motivação, comparando grupos de candidatos e não candidatos em avaliações “reais” ou manipulando as instruções em pesquisas de laboratório (Anglim, Bozic, Little, & Lievens, 2018; Birkeland et al., 2006; Georg Krammer et al., 2017).

Em uma metanálise, Birkeland et al. (2006) investigaram o quanto candidatos em processos seletivos falseavam suas respostas a testes de personalidade que avaliavam os cinco grandes fatores (extroversão, conscienciosidade, socialização, neuroticismo e abertura). Eles identificaram que candidatos pontuaram mais alto que os não-candidatos nos cinco fatores em situações reais de seleção, principalmente nos fatores conscienciosidade ( $d = 0,45$ ) e estabilidade emocional ( $d = 0,44$ ). Birkeland et al.

também analisaram o efeito moderador do tipo de cargo (vendedor e gerente) e de teste (medida direta ou indireta) nos cinco fatores. As médias nos traços de extroversão e socialização foram moderadas pela variável 'ser candidato a cargo de vendedor'. O tipo de teste teve um efeito moderador sobre todos os traços, exceto extroversão, sendo que medidas diretas (i.e., que avaliavam os cinco fatores explicitamente, como o NEO-PI) desses traços apresentaram médias maiores que de medidas indiretas (i.e., avaliam características de personalidade relacionadas aos cinco fatores, como o 16-PF). Assim, os autores concluíram que candidatos em processos seletivos manipulam suas respostas em função do traço que é avaliado, do cargo e do tipo de teste.

Krammer et al. (2017) investigaram a estabilidade de respostas durante e após um processo seletivo para uma universidade. Seis meses após o processo seletivo, os aprovados foram divididos, aleatoriamente, em três grupos. Cada grupo respondeu aos mesmos testes iniciais sob diferentes instruções: responder de forma honesta (condição honesta), responder como se estivesse passando pelo mesmo processo seletivo (condição de reprodução) e responder para parecer melhor (condição de falseamento). A variância das medidas foi avaliada, evidenciando uma relativa estabilidade de variância configural, porém com violações a nível métrico e escalar, sugerindo que os efeitos de RSD não devem ser negligenciados. Os resultados, similares aos de Birkeland et al., sugerem que em situação real de seleção os candidatos tendem a exagerar suas respostas, mas que alteram esse comportamento em função do conteúdo do instrumento.

No estudo de Anglim et al (2018) foi comparado o desempenho de candidatos a residência médica com os de residentes, estudantes de medicina e amostra normativa em relação aos cinco grandes fatores da personalidade avaliados pelo NEO-PI. Consoante com os estudos anteriores, os candidatos (C) responderam de forma mais socialmente desejável do que os não-candidatos (NC). Os candidatos obtiveram maiores médias em socialização ( $d = 1,38$ ), conscienciosidade ( $d = 1,14$ ), abertura ( $d = 0,84$ ) e extroversão

( $d = 0,64$ ), e menor média em neuroticismo ( $d = -1,14$ ) que os não-candidatos, sugerindo que contexto de processo de admissão e o conteúdo do item direcionam a manipulação das respostas.

Nos estudos relatados, a maioria utilizou como estratégia parcial a manipulação de instrução. Apesar de não ser explicitada as instruções, as condições honestas tendem a ser similares, em que é enfatizado que devem responder da forma mais honesta possível, sendo que não são informados que o objetivo é avaliar manipulação de respostas (e.g. Krammer et al., 2017; van Hooft & Born, 2012). Um exemplo de instrução honesta:

*In the next screens you will be presented with 105 questions with five response options. Please answer the questions **as honestly as possible**. Your answers will remain completely confidential and anonymous, and will be used for research purposes only. For this study we are interested in how you really are. Therefore it is very important that you answer the following questions **as accurately and honestly as you can**. (grifo dos autores, van Hooft & Born, 2012, pp.305)*

As instruções de falseamento variam de acordo com a ênfase e os interesses da pesquisa. Nesse sentido alguns pesquisadores solicitam que as pessoas imaginem que estão participando de um processo de seleção de emprego qualquer, assim são orientados a causar uma boa impressão de forma geral, pois trata-se de um contexto de alta motivação para falsear dada a importância do resultado (e.g. van Hooft & Born, 2012) ou são mais específicos a um tipo de cargo (Shoss & Strube, 2011), em que há uma expectativa de como tal profissional deva ser. Na condição de falseamento os participantes ficam cientes que o objetivo é avaliar a capacidade de manipulação de respostas. Exemplos de instruções para falsear são:

*Please imagine that you are graduated and are applying for a job. As part of the selection procedure you are presented with the following 105 questions with five response options. Please answer the questions such that you will come across as **the ideal employee**. For this study we are not interested in what your real answers for each question would be. Instead, for each question please select the answer that you feel will give you the **best rank** and make you look like **the most suitable job applicant**. (grifo dos autores, van Hooft & Born, 2012, pp.305)*

*Please answer the questions as if you were the ideal candidate for a position in public relations. According to the top public relations firms, an ideal candidate has the following qualities: outgoing, confident, credible, creative, takes initiative, and exercises good judgment. (Shoss & Strube, 2011, pp.166)*

### **Características dos itens**

Os itens costumam descrever comportamentos, pensamentos e sentimentos que a pessoa pode apresentar em situações específicas ou de forma geral no seu cotidiano. A configuração do item pode influenciar diretamente na resposta das pessoas — principalmente em contextos de alta motivação— e, conseqüentemente na validade do instrumento (Johnson, 2004). Assim, a escrita dos itens deve considerar o significado cultural das palavras escolhidas (desejabilidade social).

Itens com conteúdo socialmente desejável podem ser considerados como ameaças a validade dos instrumentos, pois, a depender da motivação do sujeito, ele tenderá a endossar positivamente tais itens. Em um estudo que verificou a interação entre motivação para falsear respostas em autorrelato e características de itens que avaliavam personalidade, foi verificado que essa interação influencia no processo de resposta e na autoavaliação (Hauenstein, Bradley, O’Shea, Shah, & Magill, 2017). Neste estudo, os participantes foram solicitados a verbalizar seus pensamentos a medida que respondiam a um conjunto de itens que avaliavam diferentes traços de personalidade, essa prática foi feita com instrução honesta e instrução para eles parecerem o melhor possível. Os itens foram controlados quanto a desejabilidade (positiva, negativa e neutra) e relacionados com quatro estratégias de processamentos cognitivos de resposta: processamento orientado ao comportamento, ao traço analisado, processamento condicional e análise semântica do significado dos itens.

Considerando os dois contextos induzidos pelas instruções, a frequência de cada estratégia variou. Na instrução honesta, a estratégia mais utilizada foi de processamento condicional ( $M = 15,76$ ), em que o sujeito relativizava sua resposta de acordo com o

contexto. Na instrução de falseamento, a estratégia mais utilizada foi a análise semântica dos itens (70,3% dos participantes), que na aplicação de itens de personalidade está relacionada a desejabilidade do conteúdo. Além disso, os resultados também mostraram que itens com desejabilidade neutra e mais transparentes (i.e., com mais validade de face) favoreceram o processamento condicional e ao orientado ao traço, nas duas instruções (Hauenstein et al., 2017). Isso evidencia a importância do conteúdo valorativo das palavras que compõe os itens.

Os itens são, portanto, formados por conteúdo descritivo e valorativo. O primeiro é referente a características do traço de interesse e, o segundo é referente ao valor cultural do que está sendo avaliado. O ideal, seria que o item fosse apenas formado, ou pelo menos que predominasse o conteúdo descritivo. Uma forma de reduzir a influência do conteúdo valorativo, isto é, reduzir a desejabilidade do item é por meio da técnica de neutralização (Bäckström, Björklund, & Larsson, 2012). Esta técnica sugere que os itens sejam reescritos usando termos que tenham um valor cultural mais neutro, evitando o uso de termos com valor muito positivo ou negativo, por exemplo em vez de “Gosto de me gabar das minhas capacidades” é melhor dizer “Gosto que saibam sobre minhas capacidades”.

De forma prática, alguns passos devem ser seguidos a fim de garantir que o conteúdo valorativo seja controlado. Posterior à reescrita, os itens devem ser avaliados quanto à sua desejabilidade por pessoas da comunidade geral, pode-se usar uma escala de 9 pontos, em que 1 indica que o conteúdo do item é muito indesejável e 9 que o conteúdo do item é muito desejável, escala sugerida pelos autores que desenvolveram a técnica (Bäckström et al., 2012). Assim, o objetivo é que o conteúdo de cada item tenha uma média próxima a 5 o que indicaria uma avaliação mais neutra. Os itens devem ser reescritos e reavaliados até que atinjam uma média próxima a 5.

A principal vantagem do método de itens neutralizados é sua simplicidade, enquanto que a desvantagem é que pode diminuir a heterogeneidade do traço avaliado



(MacCann, Ziegler, & Roberts, 2012) . Em relação ao uso, para demonstrar a facilidade de aplicação, Bäckström e Björklund (2013) pediram a um grupo de calouros de um curso de psicologia que neutralizassem um conjunto de itens que avaliavam os Cinco Grandes Fatores da Personalidade e que depois trocassem os itens entre os grupos para que o nível de desejabilidade social de cada item pudesse ser avaliado e uma média pudesse ser encontrada, ao que os alunos relataram não terem tido dificuldade. Os itens neutralizados apresentaram uma correlação entre 0,60 e 0,84 com os itens originais, indicando que avaliavam aspectos similares.

Uma segunda possibilidade de controle da desejabilidade social é o uso de instrumentos desenvolvidos para a avaliar esse viés (Crowne & Marlowe, 1960; Paulhus, 2006; Vecchione, Alessandri, & Barbaranelli, 2013). Seguindo a perspectiva teórica da desejabilidade social apresentada neste trabalho, podem ser identificadas algumas escalas, como a *Comprehensive Inventory of Desirable Response* (CIDR; Paulhus, 2006), composta por 80 itens que avaliam as quatro facetas dos vieses egoísta e moralista, a *Egoistic and Moralistic Self-Enhancement Scale* (EMS; Vecchione, Alessandri, & Barbaranelli, 2013), composta por 14 itens que avalia os dois vieses de DS, e a *Bidimensional Impression Management Index* (BIMI; Blasberg, Rogers, & Paulhus, 2014), composta por 20 itens que avalia as duas facetas relacionadas ao autogerenciamento (autogerenciamento da atuação social e autogerenciamento do eu). Apesar de úteis, essas escalas não podem ser consideradas medidas puras do viés de desejabilidade, visto que também captam variância de traços de personalidade verdadeiro (Pauls & Stemmler, 2003). Nesta pesquisa optou-se pela EMS por ser mais curta e englobar os dois vieses do modelo de Paulhus e John (1998).

Outros aspectos devem ser considerados na construção dos itens, além do controle do conteúdo valorativo a fim de garantir uma otimização na compreensão global dos itens, e reduzir possíveis distratores. Existem manuais que orientam a escrita de itens que visam

a avaliação de traços de personalidade (Bandalos, 2018b; Kline, 2015). As principais recomendações são escrever frases curtas, com linguagem simples e direta, evitar duplas negativas (“Eu nunca penso em não sair para uma festa”), frases que são ou pareçam uma opinião (“Os políticos são corruptos”), referências ao passado (“Costumava sair mais com meus amigos”). Ainda se sugere que cada frase avalie apenas uma característica (“Sou uma pessoa alegre”, “Sou uma pessoa comunicativa” em vez de “Sou uma pessoa alegre e comunicativa”). Tais orientações conduziram a elaboração dos itens que irão compor as escalas a serem desenvolvidas neste projeto.

### **Características dos formatos de resposta**

Como abordado anteriormente, o efeito de respostas socialmente desejáveis nas propriedades psicométricas em instrumentos de autorrelato ainda é incerto. Essa controvérsia, junto com o desenvolvimento de métodos estatísticos (Brown & Maydeu-Olivares, 2012), fez com que pesquisadores retornassem o interesse por formatos de respostas diferentes do tipo Likert. Um formato que voltou a ser pesquisado foram as escalas de escolha-forçada. Este formato se caracteriza pela apresentação de blocos de itens em que o sujeito deve escolher, dentre as opções disponíveis, o item que melhor lhe descreva. Ou seja, esta opção se baseia no fato de que pelo sujeito ser obrigado a escolher uma alternativa em meio a itens com níveis semelhantes de desejabilidade social, em vez de indicar o grau em que o item melhor lhe descreve, será mais difícil para ele conseguir manipular suas respostas (Christiansen, Burns, & George, 2005).

O formato de resposta de escolha forçada foi baseado na Lei de Julgamento Comparativo, proposto por Thurstone (1928). A ideia partiu da compreensão de que as atitudes estariam distribuídas em um contínuo com distâncias regulares, assim ao apresentar dois estímulos a pessoa tenderia a escolher o que julgasse estar mais próximo da forma que se percebe (i.e., utilidade do item). Dessa forma, para avaliar uma pessoa,

seria preciso que os itens cobrissem todo o contínuo do traços de interesse, isto é, a escala precisaria de itens que avaliassem características do extremo positivo ao negativo, com uma zona neutra ou intermediária (Thurstone, 1928).

Por um tempo, esse formato foi deixado de lado pelos pesquisadores (Drasgow, Chernyshenko, & Stark, 2010). Isso ocorreu pelo fato de que construir instrumentos com itens em escolha forçada é mais complexo e demorado (Heggestad, Morrison, Reeve, & McCloy, 2006), além de que impossibilitava a comparação interindividual devido a forma como os escores eram calculados que geravam escores ipsativos, isto é, a soma dos escores era constante entre os sujeitos. Porém, avanços em modelos estatísticas retornaram o interesse nesse formato, uma vez que as dificuldades foram sendo superadas e esse formato se mostrou mais vantajoso no controle de vieses de respostas (Cheung & Chan, 2002; Wetzel & Greiff, 2018).

As escalas de resposta em formato tipo Likert são uma técnica não comparativa de escalonamento de respostas (Willits, Theodori, & Luloff, 2016). Esse formato proporciona a avaliação absoluta de um item por vez, em que as pessoas indicam a valência (e.g.: se concordam ou discordam) e a intensidade (e.g. pouco, moderado e muito) com que as afirmativas as descrevem. Esse formato foi inicialmente proposto por Likert (1932) que buscava uma forma de melhor compreender atitudes e comportamentos por meio de uma métrica comum e passível de tratamento estatístico, além de ser mais fácil de construir em comparação a proposta de Thurstone (1928). Na proposta inicial de Likert, a escala teria duas valências (desaprovação e aprovação) e cinco categorias de intensidade, em que 1 indicaria total desaprovação e 5 total aprovação. Apesar de aparentemente simples, esse formato de resposta requer que se considere a polaridade, numeração de categorias, rótulos, e quantidade de categorias.

Uma forma de verificar qual formato (Likert e escolha forçada) produz respostas mais confiáveis e qual é menos suscetível à desejabilidade social é por meio de estudos

comparativos, empíricos ou de simulação. Esses estudos costumam apresentar dois instrumentos que avaliam o mesmo traço latente; porém, um é em formato de estímulo-único e, o outro, em formato de escolha forçada. Além disso, manipulam-se as instruções de forma que, no primeiro momento, os sujeitos são solicitados a responderem de forma sincera. Em um segundo momento, os mesmos instrumentos são apresentados, e é dada uma instrução para que os sujeitos manipulem suas respostas, considerando algum contexto pré-determinado. (Christiansen et al., 2005; Usami, Sakamoto, Naito, & Abe, 2016; Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006).

Em estudos de comparação, o formato de escolha forçada apresentou resultados mais favoráveis na prevenção de respostas socialmente desejáveis (Christiansen et al., 2005; Usami et al., 2016; Vasilopoulos et al., 2006). No estudo de Christiansen et al., (2005) foram comparadas duas versões de um instrumento que avalia traços de extroversão e conscienciosidade, sendo demonstradas evidências favoráveis ao uso de escolhas forçadas no controle de desejabilidade social. Neste estudo, foi solicitado aos sujeitos que primeiro respondessem aos itens de forma sincera e, depois, como se estivessem concorrendo a um cargo de vendedor. Houve mudanças na relação entre desejabilidade social e os traços de personalidades nas duas versões e situações. Quando realizada uma regressão, os autores verificaram que o escore proveniente do formato Likert explicou 35% da variância da desejabilidade social, enquanto o escore proveniente do formato de escolha forçada explicou somente 1% dessa variância, sugerindo que este último formato é mais adequado para prevenir respostas socialmente desejáveis, apesar de não ser infalível.

Outro estudo avaliando diferentes aspectos da personalidade (persistência, oscilação emocional, introversão social, desânimo e autoconfiança) com procedimento similar também favoreceu escolhas forçadas como uma alternativa viável à prevenção de respostas desejáveis (Usami et al., 2016). Neste estudo foram aplicados dois

instrumentos, um em formato Likert e outro em escolha forçada, sob duas instruções, honesta e de falsificação. Quatro tipos de informações resultantes da combinação método e instrução foram comparados: Likert/honesta, Likert/falsificada, escolha forçada/honesta e escolha forçada/falsificada. As correlações entre escolha forçada/honesta e escolha forçada/falsificada (amplitude:  $r = 0,31$  e  $0,67$ ,  $M = 0,54$ ) foram maiores que entre Likert/honesta e Likert/falsificada (amplitude:  $r = 0,18$  e  $0,52$ ,  $M = 0,33$ ). Ainda, foi verificada uma correlação entre Likert/honesta e escolha forçada/falsificada (amplitude:  $r = 0,21$  e  $0,49$ ,  $M = 0,39$ ) maior que entre Likert/honesta e Likert/falsificada. A correlação entre escolha forçada/honesta e escolha forçada/falsificada foi maior que entre Likert/falsificada e escolha forçada/honesta (amplitude:  $r = 0,06$  e  $0,50$ ,  $M = 0,26$ ). Todos estes resultados indicaram que testes baseados em escolha forçada estimam melhor o traço latente e são mais resistentes ao falseamento de respostas, independentemente do tipo de instrução e do traço avaliado (Usami et al., 2016).

Christiansen et al. (2005) orientou que os participantes imaginassem que estivessem concorrendo a um cargo de vendedor, sendo apresentada a descrição do cargo. Vasilopoulos et al. (2006) pediram que imaginassem estarem concorrendo na primeira etapa de um processo de admissão na faculdade que eles realmente queriam entrar. Usami et al., (2016) pediram que as participantes imaginesse uma situação de recrutamento de uma empresa em que eles quisessem muito trabalhar, sendo que eles deveriam se apresentar de uma forma que maximizasse a chance de serem contratados. Assim, as instruções têm em comum o fato de especificarem um contexto, ou um cargo, o que serve como pista para a pessoa do quê e em que intensidade deve falsear suas respostas.

### **Características pessoais**

A etapa de julgamento, em que o respondente deve optar por responder de forma socialmente desejável está relacionado com os traços de personalidade que ela possui.

Alguns traços, considerados como socialmente indesejáveis, podem fazer com que pessoas que tem alto nível deles manipulem mais suas respostas que pessoas que tem baixo nível. Esse é um dos pontos de interesse desta pesquisa. Por não ser possível investigar todos os traços de personalidade, optou-se por focar nos que são vistos, por alguns, como os mais prejudiciais a interação interpessoal, são os traços de maquiavelismo, psicopatia e narcisismo que formam a chamada Triade Sombria da Personalidade. Esses traços foram reunidos por compartilharem algumas características, como comportamentos desonestos, dominantes, insensíveis e agressivos. De forma mais específica, o maquiavelismo é caracterizado pela manipulação interpessoal e estratégica; a psicopatia pela impulsividade, falta de empatia e ansiedade; e o narcisismo pelo senso de grandiosidade e superioridade (Paulhus & Williams, 2002).

Pessoas com altos níveis nesses traços tendem a mentir mais, seja por razões inespecíficas ou por almejam algum ganho pessoal (Baughman, Jonason, Lyons, & Vernon, 2014; Jonason, Lyons, Baughman, & Vernon, 2014). O estudo de Baughman et al. (2014) investigou a relação entre os traços da tríade com a probabilidade de mentir em relações afetivas e acadêmicas. Em relações afetivas, psicopatia ( $r = 0,13, p < 0,01$ ) e maquiavelismo ( $r = 0,10, p < 0,01$ ) tiveram relações significativas. Em relações acadêmicas, maquiavelismo apresentou maior relação com probabilidade de mentir ( $r = 0,25, p < 0,01$ ), seguido pela psicopatia ( $r = 0,19, p < 0,01$ ) e narcisismo ( $r = 0,14, p < 0,01$ ). Quanto ao esforço cognitivo envolvido na elaboração da mentira, nas relações afetivas quem apresentou maior elaboração foram os que pontuaram mais alto em narcisismo ( $r = 0,18, p < 0,01$ ), depois os com mais traços de maquiavelismo ( $r = 0,15, p < 0,01$ ) e psicopatia ( $r = 0,14, p < 0,01$ ). Nas relações acadêmicas o maior esforço foi daqueles com mais traços de maquiavelismo ( $r = 0,28, p < 0,01$ ) e psicopatia ( $r = 0,10, p < 0,01$ ), sendo que narcisismo apresentou uma relação não significativa ( $r = 0,08, p > 0,05$ ).

No estudo de Jonason et al. (2014), eles relacionaram os traços da Tríade Sombria da Personalidade com a quantidade e a finalidade de mentiras contadas nos últimos sete dias. Os resultados mostraram uma relação positiva entre os níveis elevados dos três traços com a quantidade de mentiras ( $r = 0,21, p < 0,01$  com maquiavelismo,  $r = 0,21, p < 0,01$  com psicopatia e  $r = 0,10, p < 0,05$  com narcisismo) e a quantidade de pessoas que mentiram ( $r = 0,20, p < 0,01$  com maquiavelismo,  $r = 0,25, p < 0,01$  com psicopatia e  $r = 0,07, p > 0,05$  com narcisismo). Quanto ao objetivo, mentiras brandas foram mais explicadas por níveis mais elevados de traços de maquiavelismo ( $\beta = 0,14, p < 0,01$ ), mentir por nenhuma razão específica foi mais explicado por níveis mais elevados de traços de psicopatia ( $\beta = 0,27, p < 0,01$ ) e, mentir para ganhos pessoais foi mais explicado por níveis mais elevados de traços de narcisismo ( $\beta = 0,18, p < 0,01$ ). Além disso, quando investigado o quanto as pessoas acreditavam na sua habilidade de contar mentiras, a presença de traços de psicopatia ( $\beta = 0,40, p < 0,01$ ) e narcisismo ( $\beta = 0,14, p < 0,01$ ) explicaram mais essa crença. Esses resultados indicam a necessidade de um olhar cuidadoso para a resposta a instrumentos de autorrelato de pessoas que apresentam níveis mais elevados nos traços da Tríade Sombria da Personalidade. Tais estudos evidenciam que as pessoas falsificam suas respostas quando o desejam, ainda que a manipulação seja consciente ou não-consciente, ela depende da interação entre motivação gerada pelo contexto, conteúdo dos itens, formato do teste e características das pessoas.

No que concerne à Tríade Sombria da Personalidade eles tendem a comportamentos hedonísticos e socialmente antagônicos (Jonason, Strosser, Kroll, Duineveld, & Baruffi, 2015). Assim, esses traços tendem a apresentar uma relação com elevados níveis de valores individuais e baixos níveis de valores comunitários (Jonason & Fletcher, 2018). Em se tratando desses valores, a nível de desejabilidade social, os que apresentam mais traços de maquiavelismo e psicopatia tendem a manipulações conscientes (gerenciamento do eu e gerenciamento da atuação social), enquanto que os

que apresentam mais traços narcisistas tendem a manipulação não-consciente (autoengano positivo e autoengano negativo) (Furnham, Richards, & Paulhus, 2013).

Outro aspecto é que o conteúdo dos itens usados para avaliar a Triade Sombria da Personalidade tendem a ser socialmente indesejáveis. Eles focam no espectro mais negativo desses traços, podendo ser considerados como mais vulneráveis ao falseamento, uma vez que as pessoas tendem a ter mais dificuldade em relatar aspectos negativos da própria personalidade do que aspectos positivos, principalmente em contextos de alta motivação. Existem diversos instrumentos que avaliam esses traços de forma isolada ou em conjunto (e.g. Raskin & Hall, 1979; Jonason & Webster, 2010). Assim, considerando o uso de diferentes formatos de resposta, realizou-se uma rápida investigação para verificar a existência de escalas em formato de escolha forçada e tipo Likert que avaliassem esses traços, em conjunto ou isoladamente. A pesquisa foi feita nas 10 primeiras páginas do Periódicos CAPES e *Google Scholar*, no período de maio de 2017, usando como descritores as palavras “*psychopath*”, “*narcissism*”, “*machiavellianism*”, “*forced-choice*”, “*ipsative*”, “*ranking scale*” “*Likert*”, “*single-stimulus*” e “*rating scale*”.

Dentre os traços que compõem a tríade sombria, foram encontradas escalas em formato de escolha forçada somente para os traços maquiavelismo e narcisismo. Para o primeiro traço, pode ser citada a escala *Mach-V* (Christie & Geis, 1970) que é composta por 20 *triplets* (i.e. bloco com três itens), pontuada no tipo MOLE (i.e. seleciona o item que mais e menos tem a ver consigo) em que a pessoa recebe um ponto se marca como ‘mais a ver comigo’ o item que avalia maquiavelismo ou se o item intermediário em desejabilidade fosse escolhido. Para o segundo traço pode-se citar o *Narcissistic Personality Inventory* (NPI; Raskin & Hall, 1979), que é composto por 40 *duplets* (i.e. bloco com dois itens) e com uma chave de resposta, em que a pessoa recebe um ponto para cada vez que seleciona a resposta “correta” (i.e., que avalia alto nível de narcisismo).



Em relação ao traço de psicopatia, não foram encontradas escalas que usassem formato de escolha forçada.

Escalas que avaliam a Tríade Sombria da Personalidade e que usam o formato de resposta tipo Like foram encontradas tanto para avaliação individual dos traços como eles em conjunto. Dois exemplos de escalas que avaliam os traços em conjunto são a *Short Dark Triad* (Daniel N Jones & Paulhus, 2014) e a *Dirty Dozen* (Jonason & Webster, 2010), elas usam escala de 5 pontos, com todas as categorias rotuladas de 1 = “discordo totalmente” até 5 = “concordo totalmente”. Alguns exemplos de instrumentos que avaliam cada traço individualmente são: o Mach-IV (Christie & Geis, 1970), composto por 40 itens sobre maquiavelismo respondidos em uma escala de 5 pontos; a versão em formato Likert do NPI (Ackerman, Donnellan, Roberts, & Fraley, 2016), com 80 itens sobre narcisismo de 5 pontos com todas as categorias rotuladas; o *Pathological Narcissistic Inventory* (PNI; Pincus et al., 2009), composto por 52 itens de 6 pontos, cuja ausência do ponto neutro é justificada como uma estratégia para diminuir o efeito do viés de resposta do ponto mediano; e a *Self-Report Psychopathy IV* (SRP 4; Paulhus, Neumann, Hare, Williams, & Hemphill, 2016), que avalia psicopatia e apresenta versões de 29 e 64 itens.

Como foi revisado, existem vários aspectos que interagem para aumentar a chance de ocorrência de uma resposta socialmente desejável. Esses aspectos envolvem característica do contexto, que podem ser fontes de maior ou menor estresse de acordo com as consequências advindas dos resultados da testagem. A presença de RSD também envolve aspectos relacionados das pessoas, que podem ter traços que naturalmente as façam engajar em mais comportamentos de manipulação e mentira, como as pessoas com altos níveis de Maquiavelismo, Psicopatia e Narcisismo. Um terceiro aspecto está relacionado a forma como os itens são elaborados, se são usados muitos termos que evocam um julgamento moral do conteúdo, tirando o foco do aspecto descritivo do que se pretende

avaliar. Por fim, envolve aspectos relacionados as opções de resposta dadas ao sujeito, que podem favorecer ou não uma manipulação mais direta e fácil dos itens, como o caso das escalas Likert que permitem que cada item seja avaliado isoladamente, ou ainda formatos de resposta que demandem mais do indivíduo, forçando-o a tomar decisões que refletem o cotidiano de suas vidas, que é o caso do formato de escolha forçada. Diante disso, a presente pesquisa apresenta dois objetivos principais. O primeiro é desenvolver escalas que avaliem traços da tríade sombria, uma em formato de escolha forçada e outra em formato de resposta Likert (Estudo 1). O segundo é realizar os estudos iniciais das propriedades psicométricas de cada escala (estrutura interna, consistência interna, relação com outra escala de tríade sombria) (Estudo 2).

## **Construindo escalas de autorrelato: o que fazer? (Artigo 1)**

Ariela Raissa Lima Costa

Nelson Hauck Filho

Bruno Bonfá Araujo

*Universidade São Francisco*

**Resumo:** Escalas de autorrelato são comuns no cotidiano de pesquisadores, porém detalhes, que as vezes parecem de pouca importância, acabam sendo desconsiderados. Autorrelato são caracterizados pela pessoa de interesse ser a própria fonte de informação do pesquisador, assim a forma em que as possibilidades de resposta são apresentadas é de suma importância. Os formatos de resposta mais comuns são tipo Likert e escolhas forçadas que vem ganhando popularidade nos últimos anos. Neste artigo são apresentados os elementos (e.g., quantidade de categorias, rótulos, correção) que devem ser observados quando se constrói ou se escolhe escalas que possuam esses dois formatos de resposta. Dessa forma, pretende-se contribuir para a psicologia, em especial a area de construção de instrumentos de autorrelato.

**Palavras-chaves:** estímulo único; múltiplos estímulos; construção de instrumento; questionário.

**Abstract:** Self-report scales are common in the daily lives of researchers, but details, which sometimes seem of little importance, end up being disregarded. Self-reports are characterized by the person of interest being the researcher's own source of information, so the way in which the possibilities of response are presented is importance. The most common response formats are Likert type and forced-choices, that have been gaining popularity in recent years. This article presents the elements (e.g., number of categories,

labels, correction) that must be observed when building or choosing scales that have these two response formats. Thus, it is intended to contribute to psychology, especially the area of construction of self-report instruments.

**Keywords:** single stimulus; multiple stimulus; instrument construction; questionnaire.

**Resumen:** Las escalas de autoinforme son comunes en la vida de los investigadores, pero los detalles, que a veces parecen tener poca importancia, terminan siendo ignorados. Los autoinformes se caracterizan porque la persona de interés es la fuente de información del investigador, por lo que la forma en que se presentan las posibilidades de respuesta es de suma importancia. Los formatos de respuesta más comunes son Likert y elecciones forzadas, que han ido ganando popularidad en los últimos años. Este artículo presenta los elementos (por ejemplo, número de categorías, etiquetas, corrección) que deben observarse al crear o elegir escalas que tengan estos dos formatos de respuesta. Por lo tanto, se pretende contribuir a la psicología, especialmente al área de construcción de instrumentos de autoinforme.

**Palabras claves:** estímulo único; estímulo múltiple; construcción de instrumentos; cuestionario.

As formas de coletar informações sobre a personalidade são diversas. Cattell (1958) distinguiu três modalidades, as quais ele denominou de registro de vida (*Life record observation* ou L-data), questionário (*Questionnaire data* ou Q-data) e testes objetivos (*Objective data* ou T-data). No primeiro, formato L-, o sujeito é avaliado por um ou mais observador externo, de modo que a investigação é feita no ambiente natural do sujeito, por meio de suas relações familiares e sociais e pode ser registrado, por exemplo, pela frequência que determinado comportamento aparece. O segundo, formato

Q-, recolhe informações por meio da auto-observação, autoavaliação e autorrelato que o sujeito faz de si; são os questionários, inventários e escalas. O último, formato T-, avalia o sujeito por meio de sua reação a diferentes estímulos em um cenário controlado, usando estímulos aparentemente ambíguos, o que dificulta para a pessoa saber o que de fato está sendo avaliado, como testes situacionais e associações implícitas.

Cada formato de coleta determina o tipo de informação que deve ser acessada pelo avaliador. Tais informações estão distribuídas entre as que são acessíveis ao outro e/ou ao sujeito. Luft e Ingham (1961) elaboraram um esquema, denominado Janela de Johari, em que mostram quais informações estão disponíveis para avaliação direta ou indireta de pesquisadores e clínicos. Na Figura 1 são apresentados os quadrantes desenvolvidos por esses autores. No quadrante ‘arena’ as informações sobre a personalidade são percebidas pelo próprio sujeito e pelos outros, por exemplo uma pessoa extrovertida, que percebe em si e é visível para terceiros a facilidade para interagir com outras pessoas em diversos contextos.

No quadrante ‘ponto cego’, os outros observam algo que a própria pessoa não consegue perceber. Como exemplo, pode-se pensar em uma pessoa que considera seus comportamentos adequados a situação, mas para os outros ela é percebida como alguém hostil. No quadrante ‘fachada’, a pessoa, deliberadamente esconde informações sobre si, por exemplo, ela evita expressar pensamentos íntimos que ache que possa lhe causar embaraço. E por fim, o quadrante ‘oculto’, engloba o que é desconhecido, em que existem influências na forma de se comportar, pensar e sentir da pessoa que não são percebidas por ela nem pelos outros, por exemplo, uma pessoa que tem preconceito contra algum grupo e não demonstra isso para outros e não percebe em si (Luft & Ingham, 1961).

	Conhecido pelo eu	Desconhecido pelo eu
Conhecido pelos outros	<b>Arena</b>	Ponto cego
Desconhecido pelos outros	<b>Fachada</b>	Oculto

*Figura 1.* Modelo da Janela de Johari (Luft & Ingham, 1961)

Ao relacionar a proposta de Cattell (1958) e Luft e Ingham (1961) é possível identificar um formato de resposta para cada quadrante. As informações da ‘arena’ e ‘ponto cego’ podem ser acessadas por meio do formato L-, uma vez que este se baseia na observação externa. O formato Q- possibilita coletar informações disponíveis no quadrante ‘arena’ e ‘fachada’, posto que depende da percepção que o sujeito tem de si e de informações que, em sua maior parte, só podem ser acessadas por ele mesmo. O formato T- pode ser útil para acessar informações contidas no quadrante ‘oculto’, uma vez que o sujeito é avaliado por meio de estímulos aparentemente ambíguos que facilita que conteúdo desconhecidos ao sujeito se manifestem, e dificulta que responda baseado apenas em informações acessíveis diretamente, como testes de associação implícita.

O formato Q- (ou autorrelato, como será tratado no decorrer do texto) é muito utilizado na avaliação da personalidade (Chan, 2009) e o foco deste artigo. Isso ocorre, pois, o formato L- demanda mais tempo do pesquisador ou clínico e mais recursos pessoais e financeiros, pois o pesquisador vai a campo e passa um tempo amplo relatando os comportamentos da pessoa em diferentes situações, além de precisar entrar em contato com pessoas que fazem parte do círculo social daquela que está sendo avaliada. Do mesmo modo o formato T-, pois além da questão financeira e de tempo, ele é um método complexo tanto para a elaboração dos estímulos, como para a codificação e interpretação dos dados coletados (Cattell, 1958).

Além do tipo de informação que se pretende obter é preciso decidir de que forma o sujeito irá fornecê-las. No autorrelato, a forma de responder tem implicações da

computação de escores e nas análises estatísticas usadas para tratar cada tipo de dado. O formato de resposta mais comum é o tipo Likert, porém outro formato vem ganhando atenção dos pesquisadores é o de escolha forçada (Weijters, Cabooter, & Schillewaert, 2010). Neste último, a pessoa é apresentada a dois ou mais itens e deve escolher o item que mais tem a ver consigo (Thurstone, 1928). No tipo Likert, são apresentadas categorias ordenadas de respostas, para que a pessoa indique a intensidade com que concorda ou discorda do conteúdo do item em relação a presença ou ausência daquela característica em si (Likert, 1932). As escalas tipo Likert surgiram com a ideia de simplificar o processo de construção, já que a construção de escalas de escolha forçada demanda mais tempo para o pesquisador, pois há necessidade de criar mais itens e que tenham níveis similares de desejabilidade social e diferentes em termos de dificuldade.

São várias as decisões a serem tomadas na construção de escalas de autorrelato, seja no formato de escolha forçada ou escala Likert. Essas decisões influenciam no processo cognitivo envolvido no ato de responder a uma escala de autorrelato. Para responder a tais testes, o sujeito primeiro precisa compreender o que o item afirma (etapa de compreensão), o que implica em clareza de instrução, de escrita dos itens e o *desing* do teste em si (Tourangeau, Rips, & Rasinski, 2000), daí a importância dos elementos que serão apresentados a seguir. Para respeitar uma sequência histórica, primeiro serão apresentados os elementos de uma escala de escolha forçada, em seguida serão apresentados os elementos de uma escala com formato tipo Likert.

### **Elementos de escalas de escolha forçada**

O formato de resposta de escolha forçada foi baseado na Lei de Julgamento Comparativo, proposto por Thurstone (1928). A ideia partiu da compreensão de que as atitudes estariam distribuídas em um contínuo com distâncias regulares, assim ao apresentar dois estímulos a pessoa tenderia a escolher o que julgasse estar mais próximo

da forma que se percebe (i.e., utilidade do item). Dessa forma, para avaliar uma pessoa, seria preciso que os itens cobrissem todo o contínuo do traços de interesse, isto é, a escala precisaria de itens que avaliassem características do extremo positivo ao negativo, com uma zona neutra ou intermediária (Thurstone, 1928). As recomendações de construção de escalas em formato de escolha forçada sugerem que elas devem ser elaboradas considerando seis elementos: número de itens por bloco, a quantidade de blocos, tipos de opções de respostas, dimensionalidade dos blocos, polaridade dos itens e o sistema de correção (Brown & Maydeu-Olivares, 2011, 2016; Heggstad et al., 2006).

### **Número de itens**

Os blocos devem conter no mínimo dois itens, os chamados *duplets*. O recomendado é que os itens tenham extensão curta, evitando negativas diretas, e apresentem o máximo de quatro itens por bloco, pois quanto maior a quantidade de itens maior será a complexidade cognitiva, além de cansar os sujeitos e contribuir para respostas aleatórias (Brown & Maydeu-Olivares, 2011).

### **Quantidade de blocos**

A ideia de que mais itens mais informação sobre os sujeitos é válida no formato de escolha forçada. No caso da escolha forçada, cada bloco é considerado um item e, assim, quanto maior a quantidade de blocos melhores informações serão obtidas. Brown e Maydeu-Olivares (2011) compararam 24 (12 blocos com 2 itens em cada) e 48 (24 blocos com 2 itens em cada) itens que avaliavam dois fatores, ambos forneceram medidas precisas dos valores esperados quando usados itens positivos e negativos, no modelo de 12 blocos a confiabilidade dos dados variou entre 0,645 e 0,756 e no modelo com 24 blocos a variação foi entre 0,812 e 0,877. Também compararam 60 itens que avaliavam cinco fatores divididos em 15 (4 itens por bloco), 20 (3 itens por bloco) e 30 blocos (2



itens por bloco), e verificaram que todos esses modelos fornecem medidas precisas, com itens positivos e negativos. Em outros estudos, foram comparados instrumentos com 18 ( $r = 0,70-0,77$ ) e 36 blocos, o último teve um melhor desempenho ( $r = 0,73-0,83$ ) em recuperar o escore verdadeiro (Hontangas, Leenen, Torre, Ponsoda, & Morillo, 2016; Hontangas et al., 2015). Assim, a escolha da quantidade de blocos depende da quantidade de traços avaliados e se os itens são positivos ou mistos (positivos e negativos).

### **Opções de repostas**

Ao construir uma escala de formato de escolha forçada as opções de resposta são os formatos chamados PICK, MOLE e RANK (ver Figura 2). No formato PICK, termo que vem da palavra em inglês “*pick*” que significa “escolher”, o sujeito deve escolher somente um item por bloco, sendo mais indicado para blocos com dois itens. No formato MOLE, termo que surgiu da junção da primeira sílaba de duas palavras em inglês “*most*” e “*least*”, que significam, respectivamente, “o mais” e “o menos”, o sujeito indica em cada bloco qual item tem mais a ver consigo e o que tem menos a ver consigo, sendo indicado para blocos com três ou quatro itens. No formato RANK, termo que vem da palavra em inglês “*rank*” que significa “classificar”, o sujeito deve classificar em ordem crescente o quanto cada item lhe descreve, sendo recomendado para blocos que contenham três ou mais itens. Em estudo de simulação, considerando o tamanho e discriminação do bloco, polaridade dos itens e variabilidade de parâmetros os modelos RANK ( $r = 0,67-0,74$ ), e MOLE ( $r = 0,66-0,73$ ) apresentaram desempenhos similares para estimar o escore verdadeiro, com uma diferença menor que 0,013 (Hontangas, Leenen, Torre, et al., 2016; Hontangas et al., 2015).

<b>PICK:</b> Selecione o item que mais tem a ver com você.	
a. Gosto de manipular as pessoas.	
b. Sou uma pessoa insensível.	
<b>MOLE:</b> Selecione o item que MAIS tem a ver com você e o que MENOS tem a ver com você.	
a. Gosto de manipular as pessoas.	
b. Sou uma pessoa insensível.	
c. Mereço tratamento especial dos outros.	
<b>RANK:</b> Ordene os itens de 1 a 4, sendo que 1 indica o item que mais tem a ver com você e 4 o item que tem menos a ver com você.	
a. Gosto de manipular as pessoas.	( )
b. Sou uma pessoa insensível.	( )
c. Mereço tratamento especial dos outros.	( )
d. Sinto alegria na maior parte do tempo.	( )

Figura 2. Exemplo de tipos de opções de respostas em escalas de escolha forçada

### Dimensão do atributo avaliado

O atributo pode ser uni ou multidimensional. Os instrumentos multidimensionais têm blocos em que cada item se refere a uma dimensão independente, e que a distinção no conteúdo dos itens seja clara. Sendo sugerido que quanto maior o número de atributos e menor e/ou negativa for a correlação entre eles melhor a avaliação, pois a informação será influenciada majoritariamente por um traço (Brown & Maydeu-Olivares, 2011).

### Polaridade dos itens

Dentro de cada bloco, a polaridade dos itens pode ser positiva, invertida (negativa) ou mista (positivos e negativos). Essa decisão está ligada à dimensionalidade do atributo avaliado. Os itens positivos avaliam a diferença entre dois traços, e itens invertidos a soma da influência de dois fatores, sendo que a combinação (itens mistos) deles permite localizar o item no contínuo do traço. Estudos de simulação sugerem que itens escritos em direções opostas (positivos e negativos) fornecem melhores dados ( $r = 0,80 - 0,91$ ) que itens apenas direcionados positivamente para estimar o escore verdadeiro ( $r = 0,67 - 0,74$ ) quando avaliada em relação a quantidade de itens, tipos de opções de resposta e

dimensionalidade (Brown & Maydeu-, 2011; Hontangas et al., 2015; Hontangas, Leenen, Torre, Ponsoda, & Morillo, 2016).

### **Sistema de correção**

Refere-se a como o escore de cada sujeito será contabilizado. Duas abordagens para corrigir escalas de escolha forçada foram identificados por Brown e Maydeu-Olivares (2016). O primeiro é baseado na abordagem clássica ou medida por “decreto”, em que simplesmente somam-se os escores de cada item, isto é, escalas de escolha forçada são tratadas como escalas normativas o que faz com que sejam produzidos escores ipsativos ou quase-ipsativos, em que não há variabilidade na pontuação entre diferentes pessoas. O segundo é baseado em abordagens que usam a teoria de resposta ao item (TRI) como uma forma de contornar a formação de escores ipsativos, eles fazem uso de métodos de comparação binária de forma a possibilitar comparações interindividuais, pois permitem variação na pontuação das pessoas. Na tabela 1 pode-se visualizar a correção de alguns itens com base na abordagem clássica e uma baseada em TRI.

A forma de correção clássica gera escores ipsativos. O termo ‘ipsativo’ foi usado inicialmente por Cattell para designar escores de uma escala que são constantes para todas as pessoas (Brown & Maydeu-Olivares, 2011; Brown & Maydeu-Olivares, 2013). Esse tipo de dado compromete a qualidade psicométrica da escala, uma vez que a pontuação em um atributo é dependente de outro atributo que está sendo avaliado, isto é, uma pessoa que pontua alto em um traço necessariamente vai pontuar baixo em outro, dessa forma a pessoa só pode ser comparada com ela mesma, a aplicação desse método pode ser visto na Tabela 1. Também distorce as evidências de validade do instrumento, por exemplo, devido à falta de variância a relação com outras escalas seja enviesada e dificulta a estimação da fidedignidade, pois viola os pressupostos de técnicas estatísticas usadas para calcular esse coeficiente, como a ideia de linearidade entre escore observado e escore

verdadeiro no traço avaliado (Brown & Maydeu-Olivares, 2011; Brown & Maydeu-Olivares, 2013; Chan, 2003; Wang, Qiu, Chen, Ro, & Jin, 2017)

O sistema de correção Thurstonian baseado na teoria de resposta ao item apresenta a vantagem de poder ser usado tanto para instrumentos com estrutura unidimensionais e multidimensionais. Por meio dele é possível analisar blocos com qualquer quantidade de itens. Essa proposta de correção foi baseada na teoria de Thurstonian em 1931, em que ele dizia que uma pessoa ao responder a um teste comparava a utilidade (valor psicológico) de cada item com o seu nível individual no traço avaliado para decidir qual selecionar. Na prática são usadas comparações binárias entre pares de itens que formam um bloco. A quantidade de comparações obtidas é calculada pela fórmula  $\tilde{n} = n(n-1)/2$  em que  $n$  corresponde ao número de itens por bloco, um exemplo de aplicação desse método pode ser visualizada na Tabela 1. O modelo estima carga fatorial, variância de erro, correlação entre os atributos, o valor de *thresholds* (probabilidade de 50% da pessoa selecionar uma categoria de resposta em detrimento a outra), e permite a formulação de um escore para cada sujeito (Brown & Maydeu-Olivares, 2012).

Tabela 1.

*Comparação entre pontuação clássica (PC) e pontuação pelo modelo Thurstonian de TRI em um bloco de item de escolha forçada com tipo MOLE de opção de resposta*

Itens	Pessoa A			
	Mais	Menos	PC	Thurstonian
a.Gosto de manipular as pessoas	x		2	{A,B} = 0
b.Sou uma pessoa insensível			1	{A,C} = 1
c.Considero-me melhor do que os outros		x	0	{B,C} = 1
Total			3	2
Itens	Pessoa B			
	Mais	Menos	PC	Thurstonian
a.Gosto de manipular as pessoas		x	0	{A,B} = 0
b.Sou uma pessoa insensível	x		2	{A,C} = 0
c.Considero-me melhor do que os outros			1	{B,C} = 1
Total			3	1

A Tabela 1 traz o exemplo de duas pessoas (A e B) que responderam ao mesmo teste e selecionaram diferentes itens, porém obtiveram o mesmo escore (coluna PC) em um item de escolha forçada com tipo MOLE de resposta. Na pontuação clássica, o item que mais tem a ver com o sujeito recebe 2 pontos, o que menos tem a ver recebe 0 e os outros recebem 1. Assim, independentemente dos itens selecionados e da pessoa a pontuação será sempre 3, a única comparação possível é entre a pessoa com ela mesma, isto é, um escore ipsativo. Quando o escore é calculado por uma abordagem baseada na TRI, por exemplo o modelo Thurstonian de TRI, em que se parte de combinações binárias, o item A é comparado com o item B e C, e o item B com o item C, se o primeiro item for selecionado em vez do segundo dá-se um ponto, em caso contrário dá-se zero. Elas permitem que haja variância nos escores e a comparação interpessoal se torna possível, logo fica perceptível que a Pessoa A tem mais traços da tríade sombria do que a Pessoa B. Assim, neste trabalho será considerado o sistema de correção Thurstonian que se baseia na TRI.

### **Elementos de escalas de formato tipo Likert**

As escalas de resposta em formato tipo Likert são uma técnica não comparativa de escalonamento de respostas (Willits et al., 2016). Esse formato proporciona a avaliação absoluta de um item por vez, em que as pessoas indicam a valência (e.g.: se concordam ou discordam) e a intensidade (e.g. pouco, moderado e muito) com que as afirmativas as descrevem. Esse formato foi inicialmente proposto por Likert (1932) que buscava uma forma de melhor compreender atitudes e comportamentos por meio de uma métrica comum e passível de tratamento estatístico, além de ser mais fácil de construir em comparação a proposta de Thurstone (1928). Na proposta inicial de Likert, a escala teria duas valências (desaprovação e aprovação) e cinco categorias de intensidade, em que 1

indicaria total desaprovação e 5 total aprovação. Apesar de aparentemente simples, esse formato de resposta requer que se considere a polaridade, numeração de categorias, rótulos, e quantidade de categorias.

### Polaridade da escala

A escala pode ser unipolar (i.e., apresenta apenas um item em um extremo da escala) ou bipolar, também conhecido como formato de diferencial semântico (i.e., apresenta um item em cada extremo da escala), como apresentado na Figura 3. Esse aspecto é explicado pelo efeito da simetria, em que os respondentes, ao se depararem com as categorias, percebem escalas bipolares como mais simétricas e mais fáceis de responder. No caso das escalas unipolares, elas têm o significado das categorias dividida de uma forma assimétrica, demandando um esforço cognitivo maior por parte do respondente para selecionar a categoria que melhor lhe representa (Cabooter, Weijters, Geuens, & Vermeir, 2016).

a.Unipolar positivo	Sou uma pessoa sensível	Discordo totalmente	1	2	3	4	5	Concordo totalmente				
b.Bipolar positivo/negativo	Sou uma pessoa insensível	Discordo totalmente	-2	-1	0	1	2	Concordo totalmente	Sou uma pessoa sensível			
c.Unipolar positivo/negativo	Sou uma pessoa sensível	Discordo totalmente	1	Discordo	2	Nem discordo, nem concordo	3	Concordo	4	5	Concordo totalmente	
d.Bipolar positivo	Sou uma pessoa insensível	Discordo totalmente	-2	Discordo	-1	Nem discordo, nem concordo	0	Concordo	1	2	Concordo totalmente	Sou uma pessoa sensível

Figura 3. Exemplos de formatos de escalas tipo Likert. Fonte: adaptado de Cabooter et al. (2016). Nota. Os formatos ‘b’ e ‘d’ também são chamados de diferencial semântico.

### Numeração das categorias

Esta pode ser com todos os itens positivos e apresentados em ordem crescente (Figura 3, itens 'a' e 'd'), ou metade dos itens negativos e outra metade dos itens positivos (Figura 3, itens 'b' e 'c'). A influência desse aspecto pode ser explicada pelo efeito de intensidade, em que categorias extremas são percebidas como mais intensas do que as outras. A presença de itens positivos e negativos torna mais claro o significado dos pontos, principalmente dos pontos extremos por dar mais intensidade a eles, fazendo com que as pessoas escolham as categorias com mais cautela, o que pode contribuir para redução do viés de respostas extremas (Cabooter et al., 2016).

No estudo de Cabooter et al. (2016) eles verificaram a relação entre os tipos de polaridade quanto a efeito de simetria e intensidade, e a relação delas com viés de concordância e respostas extremas. Os resultados mostram que escalas unipolares positivas geraram menos concordância em relação aos outros três formatos, e menos respostas extremas em comparação com escalas unipolares positivas/negativas ( $F(1330) = 18,24, p < 0,01$ ). As escalas bipolares positivas/negativas ( $m = 0,17$ ) geraram menos respostas extremas em relação às escalas bipolares positivas ( $m = 0,23$ ), possivelmente pelo efeito de intensidade que foi maior na primeira ( $F(1330) = 7,76, p < 0,01$ ). Cada tipo apresentou limitações que podem interferir na distribuição de resposta, assim os autores recomendam que seja considerado o atributo que está sendo avaliado para decidir qual formato de escala se ajusta melhor.

## **Rótulos**

Eles indicam o tipo de julgamento a ser feito pela pessoa (Parker, Vannest, & Davis, 2013). Os mais comuns são os que vão desde “discordo totalmente” até “concordo totalmente”. Mesmo que a simples tradução de categorias pareça algo comum, a intensidade percebida para um dado advérbio e/ou substantivo é diferente para cada cultura (e.g. o termo 'very' do inglês pode ser traduzido como 'muito' ou 'bastante' e

esses dois termos do português podem ser percebidos com intensidade variadas, em que alguns podem entender que muito > bastante ou bastante > muito). Consequentemente, isso pode influenciar na distribuição das respostas devido a diferenças na compreensão de intensidade dos rótulos (Dolnicar & Grün, 2013). O ideal é a realização de investigações com pessoas de cada cultura para entender a ordem de intensidade dada a advérbios como ‘muito’, ‘bastante’, ‘extremamente’, ‘moderadamente’, ‘pouco’, ‘às vezes’, ‘sempre’, entre outros que são encontrados em instrumentos de autorrelato que usam escalas de resposta tipo Likert (Weijters, Geuens, & Baumgartner, 2013).

A posição dos rótulos é outro ponto importante. A dúvida recai sobre rotular todas as categorias de respostas (Figuras 3, itens ‘a’ e ‘b’) ou apenas os extremos (Figuras 3, itens ‘c’ e ‘d’). Embora o assunto pareça de pouca importância, esse detalhe de construção da escala pode criar vieses de respostas, como investigado por Weijters, Cabooter e Schillewaert (2010). No estudo, os autores investigaram a influência de rótulos completos ou nos extremos em favorecer respostas aquiescentes (RA), extremas (RE) e respostas descuidadas a itens invertidos (RD). Por meio de análises de moderação e regressão, os resultados mostraram que escalas com rótulos completos tendiam a inflar RA, porém com um tamanho de efeito pequeno ( $\beta = 0,168$ ,  $t = 6,32$ ,  $p < 0,001$ ), diminuir RE ( $\beta = 0,436$ ,  $t = -9,67$ ,  $p < 0,001$ ) e RD ( $\beta = -0,490$ ,  $t = -5,81$ ,  $p < 0,001$ ), pois a diferença entre categorias positivas e negativas é destacada, assim como destacado o ponto intermediário. Outro aspecto é que assim se demanda um menor esforço cognitivo do respondente, já que o significado de todas as categorias está disponível. Por outro lado, o uso de rótulos apenas nos extremos teve valores consistentemente altos de variância explicada em todos os itens (entre 1,4 e 2,6). Isso sugere que esse formato proporciona melhores estimativas em modelos lineares, isto é, em casos que se queira investigar e estimar a relação linear entre duas variáveis (e.g.: correlações, regressões lineares) (Weijters, Cabooter, & Schillewaert, 2010)



A familiarização com os rótulos é outro ponto a se considerar ao construir uma escala do tipo Likert. Quando a pessoa já respondeu testes de diversas temáticas, mas que usam o mesmo formato de resposta, no momento em que se depara com tal formato ela tende a interpretar as categorias (rótulos) baseadas em orientações passadas (Weijters et al., 2013). Isso pode fazer com que as pessoas respondam de forma descuidada, pois não há uma preocupação em avaliar o que está sendo afirmado antes de responder, como também pode acarretar informações equivocadas, pois o novo teste pode ter uma interpretação das categorias diferente das quais ela está habituada. Por exemplo, considerando escalas com cinco categorias, se a pessoa aprende que 1 indica discordância e 5 concordância ela tenderá a responder com essas interpretações, mas caso responda a um teste em que o pesquisador resolveu trocar para que 1 indique concordância e 5 discordância e a pessoa se abstém de ler as orientações, todas as repostas estarão erradas e a avaliação da pessoa apresentará outra interpretação.

### **Quantidade de categorias de resposta**

Em um estudo comparativo Lee e Paek (2014) não identificaram diferenças nas propriedades psicométricas no uso de 4, 5 e 6 categorias, porém houve um decréscimo na qualidade quando consideradas 2 e 3 categorias. Weijters, Cabooter e Schillewaert (2010) verificaram que escalas de 5 pontos com rótulos completos tendem a diminuir respostas descuidadas para itens invertidos (RD), e que escalas com 5 ou 7 pontos com rótulos nos extremos estimam melhor a relação linear entre variáveis. Weng (2004) investigou a relação entre a quantidade de categorias e o coeficiente alfa. Ele verificou uma relação positiva entre essas variáveis, sugerindo que aumentar a quantidade de categoria aumentaria o coeficiente alfa da escala. Não obstante, o autor destaca que a escolha dessa quantidade deve considerar a habilidade discriminativa da amostra investigada, pois quanto mais categorias mais esforço cognitivo é exigido dos respondentes.

A presença ou ausência do ponto neutro é outro aspecto a ser considerado na decisão da quantidade de categorias. A interpretação desse ponto varia e pode ser compreendido, segundo Baka e Figgou (2012), em três categorias: (a) falta de conhecimentos ou indiferença, quando a pessoa desconhece sobre o assunto investigado, quando tem dificuldade de compreender o conteúdo do item ou não se importa com o que está sendo avaliado; (b) dilema e ambivalência, essa interpretação é reflexo de uma verdadeira neutralidade do sujeito em relação ao que está sendo avaliado ou porque considera o sentido do que está sendo avaliado como eticamente conflitante; (c) oposição a afirmativa, quando a pessoa percebe o ponto neutro como uma forma de se opor ao que está sendo avaliado. Achados similares foram encontrados por Nadler, Weston e Voyles (2015), que perguntaram as pessoas qual o sentido do ponto neutro para elas, os significados mais frequentes foram não ter uma opinião (15%), não se importar (14%), incerteza da resposta (13%) e como sendo um ponto intermediário (10%).

O mais comum é que tais decisões sejam tomadas de forma aleatória e sem uma preocupação com o impacto na qualidade da avaliação das pessoas. O desenho de escalas tipo Likert pode afetar na distribuição das respostas (i.e. vieses de resposta), na avaliação da homogeneidade ou heterogeneidade da amostra, na média, na correlação entre itens e na relação com variáveis demográficas, afetando a qualidade psicométrica e as conclusões geradas pelo escore do teste (Cabooter et al., 2016; Parker, Vannest, & Davis, 2013). Assim é importante considerar o construto que está sendo avaliado e a população a que se destina para verificar qual a melhor configuração de uma escala em que se pretende usar formato de resposta tipo Likert.

### **Boas práticas na construção de itens**

Até agora foram apresentados elementos dos formatos de resposta, sendo preciso também focar na qualidade da escrita dos itens que irão compor a escala. Uma vez

escolhido o formato (i.e., Likert, escolha forçada) para ser utilizado no instrumento, bem como as especificidades do construto a ser mensurado, certas práticas podem facilitar o trabalho do pesquisador ao construir os itens. Na Tabela 2 são apresentados comparativos no que tangem alguns elementos de escalas de autorrelato.

Tabela 2.

*Boas práticas no processo de construção de itens*

	<b>O que evitar</b>	<b>Como pode ser feito</b>
Evite itens com conteúdo complexo	Sou lépido em festividades Tenho tido uma conduta melancólica	Sou animado em festas Tenho me sentido triste
Evite itens com expressões regionais	Costumo ser o bode expiatório Sou mão-de-vaca	Costumo ser culpado por coisas que não fiz Sou avarento
Evite itens com dupla negativas	Eu nunca falho com meus compromissos Eu nunca falho com minhas atividades	Sou comprometido Eu sempre faço minhas atividades
Evite itens com duas ideias	Sou extrovertido e amável Me sinto bem quando estou com meus colegas e meus familiares	Sou extrovertido Sou amável Me sinto bem com meus colegas Me sinto bem com meus familiares
Evite itens com a palavra não	Não sou feliz Não sou organizado	Sou infeliz Sou desorganizado

Ao propor um instrumento, os responsáveis pela pesquisa devem estar atentos e evitar certas práticas que tendem a exigir um maior esforço cognitivo dos respondentes, bem como influenciar negativamente as análises realizadas após a coleta de dados (Artino, Gehbach, & During, 2011; Furr, 2011). Itens com conteúdo complexo são inacessíveis a maior parte da população, bem como desmotivam os respondentes; itens bem construídos devem expressar de maneira clara e concisa o conteúdo. Expressões regionais podem representar uma escapatória para construtos de difícil definição, contudo pesquisadores devem estar cientes de que a maior parte destas não se aplicam a todos os

grupos sociais e tornam o instrumento restrito à um pequeno grupo (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014).

Nessa mesma conformidade, itens com duplas negativas e que mensuram duas ideias são armadilhas que podem dificultar o processo de análises. Além de exigirem mais dos respondentes, o processo de mensuração se torna incerto, visto que as respostas atribuídas não podem ser consideradas totalmente fiéis. Por último, ao serem propostos instrumentos balanceados (i.e., a mesma quantidade de itens com direção positiva e negativa) para estimativas estatísticas como a aquiescência; deve-se evitar apenas incluir a palavra “não” para inverter o conteúdo de um item. Dado que a inserção dessa palavra dificulta o processo de resposta com chaves que vão desde discordo totalmente até concordo totalmente (Furr, 2011).

### **Considerações finais**

Os elementos apresentados são importantes no processo de construção de instrumentos de autorrelato, eles influenciam na quantidade e na qualidade de informação que os pesquisadores irão obter. Muitas vezes são tratados como aspectos secundários no processo de construção, porém como os estudos demonstraram cada elemento influencia na compreensão e julgamento que o sujeito faz antes de optar por uma resposta. Com isso, pretendeu-se deixar mais acessível esses conhecimentos e contribuir para a área de construção de instrumentos, especialmente, instrumentos que avaliam construtos psicológicos. Deve-se estar principalmente atento ao processo de construção dos itens, se estes retratam o construto de forma genuína, como também considerar práticas que visem a proporcionar uma melhor clareza e concisão para o respondente.

Este estudo não está isento de limitações, assim como não pretendeu esgotar a temática de desenvolvimento de instrumentos. Contudo, teve como principal

recomendação colocar em foco uma prática, algumas vezes, menosprezada ao propor o desenvolvimento de uma ferramenta que auxilia os profissionais. É oportuno destacar tópicos não explorados aqui, que também podem influenciar na resposta do examinando, como a apresentação visual dos itens e o processo de formatação de instrumentos (ver Artino & Gehlbach, 2012).

## Referências

- American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME] (2014). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Artino, A. R., Jr., Gehlbach, H., & Durning, S. J. (2011). AM Last Page: Avoiding five common pitfalls of survey design. *Academic Medicine: Journal Of The Association Of American Medical Colleges*, 86(10), 1327-1327. <https://doi.org/10.1097/ACM.0b013e31822f77cc>
- Artino, A. R., Jr., & Gehlbach, H. (2012). AM last page: Avoiding four visual-design pitfalls in survey development. *Academic Medicine: Journal Of The Association Of American Medical Colleges*, 87(10), 1452. <https://doi.org/10.1097/ACM.0b013e31826ac042>
- Ackerman, R. A., Donnellan, M. B., Roberts, B. W., & Fraley, R. C. (2016). The effect of response format on the psychometric Properties of the Narcissistic Personality Inventory: Consequences for item meaning and factor structure. *Assessment*, 23(2), 203–220. <https://doi.org/10.1177/1073191114568113>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC, US: American Psychological Association.
- Anglim, J., Bozic, S., Little, J., & Lievens, F. (2018). Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings. *Advances in Health Sciences Education*, 23(2), 311–321. <https://doi.org/10.1007/s10459-017-9796-8>

- Artino, A. R. J., Gehlbach, H., & Durning, S. J. (2011). AM Last Page: Avoiding Five Common Pitfalls of Survey Design. *Academic Medicine*, 86(10), 1327. <https://doi.org/10.1097/ACM.0b013e31822f77cc>
- Bäckström, M., & Björklund, F. (2013). Social desirability in personality inventories: symptoms, diagnosis and prescribed cure. *Scandinavian Journal of Psychology*, 54(2), 152–159. <https://doi.org/10.1111/sjop.12015>
- Bäckström, M., & Björklund, F. (2014). *Social Desirability in Personality Inventories: The Nature of the Evaluative Factor*. <https://doi.org/10.1027/1614-0001/a000138>
- Bäckström, M., Björklund, F., & Larsson, M. R. (2012). Social desirability in personality assessment. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 201–213). New York, NY: Oxford University Press, Inc.
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation*, 19(5). <https://doi.org/https://doi.org/10.1146/annurev.psych.53.100901.135239>.
- Baka, A., & Figgou, L. (2012). “Neither agree, nor disagree”: A critical analysis of the middle answer category in Voting Advice Applications. *International Journal of Electronic Governance*, 5(3/4), 244–263. <https://doi.org/10.1504/IJEG.2012.051306>
- Bandalos, D. L. (2018a). The test development process. In D. L. Bandalos (Ed.), *Measurement Theory and Applications for the Social Sciences* (pp. 41–84). New York, NY: The Guilford Press.
- Bandalos, D. L. (2018b). Writing noncognitive items. In D. L. Bandalos (Ed.), *Measurement Theory and Applications for the Social Sciences* (pp. 85–119). New York, NY: The Guilford Press.
- Baughman, H. M., Jonason, P. K., Lyons, M., & Vernon, P. A. (2014). Liar liar pants on fire: Cheater strategies linked to the Dark Triad. *Personality and Individual Differences*

*Differences*, 71(71), 35–38. <https://doi.org/10.1016/j.paid.2014.07.019>

- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14(4), 317–335. <https://doi.org/10.1111/j.1468-2389.2006.00354.x>
- Blasberg, S. A., Rogers, K. H., & Paulhus, D. L. (2014). The bidimensional impression management index (BIMI): Measuring agentic and communal forms of impression management. *Journal of Personality Assessment*, 96(5), 523–531. <https://doi.org/10.1080/00223891.2013.862252>
- Brown, A., & Maydeu-olivares, A. (2016). Modeling forced-choice response formats. In P. Irwing, T. Booth, & D. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing* (pp. 1–64). London: John Wiley & Sons, Inc.
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460–502. <https://doi.org/10.1177/0013164410375112>
- Brown, Anna. (2016). Item response models for forced-choice questionnaires: A common framework. *Psychometrika*, Vol. 81, pp. 135–160. <https://doi.org/10.1007/s11336-014-9434-9>
- Brown, Anna, & Maydeu-Olivares, A. (2012). Fitting a thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods*, 44(4), 1135–1147. <https://doi.org/10.3758/s13428-012-0217-x>
- Brown, Anna, & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, Vol. 18, pp. 36–52. <https://doi.org/10.1037/a0030641>
- Cabooter, E., Weijters, B., Geuens, M., & Vermeir, I. (2016). Scale format effects on response option interpretation and use. *Journal of Business Research*, 69(7), 2574–



2584. <https://doi.org/10.1016/j.jbusres.2015.10.138>

Cattell, R. B. (1958). What is “objective” in “objective personality tests?” *Journal of Counseling Psychology*, 5(4), 285–289. <https://doi.org/10.1037/h0046268>

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309–336). New York, NY: Routledge/Taylor & Francis Group.

Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, 30(1), 99–121. <https://doi.org/10.2333/bhmk.30.99>

Christiansen, N. D., Burns, G. N., & George, E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, 18(3), 267–307. [https://doi.org/10.1207/s15327043hup1803\\_4](https://doi.org/10.1207/s15327043hup1803_4)

Christie, R., & Geis, F. L. (1970). *Studies in machiavellianism*. <https://doi.org/10.1016/B978-0-12-174450-2.50008-7>

Collison, K. L., Vize, C. E., Miller, J. D., & Lynam, D. R. (2018). Development and preliminary validation of a Five Factor Model measure of Machiavellianism. *Psychological Assessment*, 30(10).

Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349–354. <https://doi.org/10.1037/h0047358>

Cunha, C. M., Almeida Neto, O. P., & Stackfleth, R. (2016). Principais métodos de avaliação psicométrica da confiabilidade de instrumentos de medida. *Revista de Atenção à Saúde*, 14(49), 98–103. <https://doi.org/10.13037/rbcs.vol14n49.3671>

Dolnicar, S., & Grün, B. (2013). “Translating” between survey answer formats. *Journal of Business Research*, 66(9), 1298–1306. <https://doi.org/10.1016/j.jbusres.2012.02.029>

- Donovan, J., & Dwright, S. A. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal Of Business and Psychology*, 29(September 2013), 479–493. <https://doi.org/10.1007/s10869-013-9318-5>
- Dragow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology*, 21(3), 465–476. <https://doi.org/http://dx.doi.org/10.5380/psi.v21i3.53054>
- Dubas, J. S., Baams, L., Doornwaard, S. M., & Aken, M. A. G. Van. (2017). Dark Personality Traits and Impulsivity Among Adolescents: Differential Links to Problem Behaviors and Family Relations. *Journal of Abnormal Child Psychology*, 126(7), 877–889.
- Ellingson, J. E. (2012). People fake only when they need to fake. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 19–33). New York, NY: Oxford University Press, Inc.
- Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social and Personality Psychology Compass*, Vol. 7, pp. 199–216. <https://doi.org/10.1111/spc3.12018>
- Geel, M. Van, Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences*, 106, 231–235. <https://doi.org/10.1016/j.paid.2016.10.063>
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96.
- Hauenstein, N. M. A., Bradley, K. M., O’Shea, P. G., Shah, Y. J., & Magill, D. P. (2017). Interactions between motivation to fake and personality item characteristics:

- Clarifying the process. *Organizational Behavior and Human Decision Processes*, 138, 74–92. <https://doi.org/10.1016/j.obhdp.2016.11.002>
- He, J., van de Vijver, F. J. R., Dominguez Espinosa, A., Abubakar, A., Dimitrova, R., Adams, B. G., ... Villieux, A. (2015). Socially Desirable Responding. *Cross-Cultural Research*, 49(3), 227–249. <https://doi.org/10.1177/1069397114552781>
- Heggstad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, Vol. 91, pp. 9–24. <https://doi.org/10.1037/0021-9010.91.1.9>
- Hervé, H. (2017). Psychopathy across the ages: A history of the Hare psychopath. In H. Hugues & J. C. Yuille (Eds.), *The Psychopath: Theory, Research, and Practice* (pp. 31–55). Routledge.
- Hogan, R. (1991). Personality and personality measurement. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 873–919). Palo Alto, CA: Consulting Psychologists Press.
- Hontangas, P. M., Leenen, I., de la Torre, J., Ponsoda, V., Morillo, D., & Abad, F. J. (2016). Traditional scores versus IRT estimates on forced-choice tests based on a dominance model. *Psicothema*, Vol. 28, pp. 76–82. Hontangas, Pedro M.: Facultad de Psicología, Universidad de Valencia, Valencia, Spain, 46011, [hontanga@uv.es](mailto:hontanga@uv.es): Colegio Oficial de Psicólogos del Principado de Asturias.
- Hontangas, P. M., Leenen, I., Torre, J. De, Ponsoda, V., & Morillo, D. (2016). Traditional scores versus IRT estimates on forced-choice tests based. *Psicothema*, 28(1), 76–82. <https://doi.org/10.7334/psicothema2015.204>
- Hontangas, P. M., Torre, J. De, Ponsoda, V., Leenen, I., Morillo, D., & Abad, F. J. (2015). Comparing traditional and IRT scoring of forced-choice tests. *Applied Psychological Measurement*, 39(8), 598–612. <https://doi.org/10.1177/0146621615585851>

- Johnson, J. A. (2004). The impact of item characteristics on item and scale validity. *Multivariate Behavioral Research*, 39(2), 273–302. <https://doi.org/10.1207/s15327906mbr3902>
- Jonason, P. K., & Fletcher, S. A. (2018). Agentic and communal behavioral biases in the Dark Triad traits. *Personality and Individual Differences*, 130(August), 76–82. <https://doi.org/10.1016/j.paid.2018.03.044>
- Jonason, P. K., Lyons, M., Baughman, H. M., & Vernon, P. A. (2014). What a tangled web we weave: The dark triad traits and deception. *Personality and Individual Differences*, 70, 117–119. <https://doi.org/10.1016/j.paid.2014.06.038>
- Jonason, P. K., Strosser, G. L., Kroll, C. H., Duineveld, J. J., & Baruffi, S. A. (2015). Valuing myself over others: The Dark Triad traits and moral and social values. *Personality and Individual Differences*, 81, 102–106. <https://doi.org/10.1016/j.paid.2014.10.045>
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, Vol. 22, pp. 420–432. <https://doi.org/10.1037/a0019265>
- Jones, D. N., & Paulhus, D. (2009). Machiavellianism. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 93–108). The Guilford Press.
- Jones, Daniel N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3). *Assessment*, 21(1), 28–41. <https://doi.org/10.1177/1073191113514105>
- Jones, Daniel N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, 2(1), 28–41. <https://doi.org/10.1177/1073191113514105>
- Kline, P. (2015). Making test reliable II: Personality inventories. Item writing. In P. Kline (Ed.), *A handbook of test construction: Introduction to psychometric design* (pp.

24–76). Routledge/Taylor & Francis Group.

Kowalski, C. M., Vernon, P. A., & Schermer, J. A. (2019). The Dark Triad and facets of personality. *Current Psychology*, (November). <https://doi.org/10.1007/s12144-019-00518-0>

Krammer, G., Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment*, *99*(5), 510–523. <https://doi.org/https://doi.org/10.10>

Krammer, Georg, Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment*, *99*(5), 510–523. <https://doi.org/10.1080/00223891.2017.1285781>

Krizan, Z., & Herlache, A. D. (2017). The narcissism spectrum model: A synthetic view of narcissistic personality. *Personality and Social Psychology Review*, 1–29. <https://doi.org/10.1177/1088868316685018>

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22* 140, 55. <https://doi.org/2731047>

Luft, J., & Ingham, H. (1961). The johari window. *Human Relations Training News*, *5*(1), 6–7.

MacCann, C., Ziegler, M., & Roberts, R. D. (2012). Faking in personality assessment: reflections and recommendations. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 309–329). New York, NY: Oxford University Press, Inc.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, *136*(3), 450–470. <https://doi.org/10.1037/a0019216>

- Morey, L. C. (2012). Detection of Response Bias in Applied Assessment: Comment on McGrath et al. (2010). *Psychological Injury and Law*, 5(3–4), 153–161. <https://doi.org/10.1007/s12207-012-9131-x>
- Müller, G., Welter, R., & Capitão, C. G. (2007). Medidas ipsativas na avaliação psicológica. *Avaliação Psicológica*, 6(2), 157–165.
- Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology*, 142(2), 71–89. <https://doi.org/10.1080/00221309.2014.994590>
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81(6), 660–679. <https://doi.org/10.1037/0021-9010.81.6.660>
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2013). Reliability of multi-category rating scales. *Journal of School Psychology*, 51(2), 217–229. <https://doi.org/10.1016/j.jsp.2012.12.003>
- Paulhus, D. L. (2002). Socially desirable responding: the evolution of a construct. In H. I. Brown, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.
- Paulhus, D. L. (2006). *Comprehensive Inventory of Desirable Responding (CIDR)*. New Orleans.
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, 66(6), 1025–1060. <https://doi.org/10.1111/1467-6494.00041>
- Paulhus, D. L., Neumann, C. S., Hare, R. D., Williams, K. M., & Hemphill, J. F. (2016). *Self-report Psychopathy Scale - 4* (4th ed.). Toronto, ON, Canada: Multi-Healthy Systems.
- Paulhus, D. L. (1984). Two-Component Models of Socially Desirable Responding.

*Personality Processes and Individual Differences*, 46(3), 598–609.

- Paulhus, Delroy L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://doi.org/https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/https://doi.org/10.1016/S0092-6566(02)00505-6)
- Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences*, 35(2), 263–275. [https://doi.org/10.1016/S0191-8869\(02\)00187-3](https://doi.org/10.1016/S0191-8869(02)00187-3)
- Persson, B. N., Kajonius, P. J., & Garcia, D. (2017). Revisiting the structure of the Short Dark Triad. *Assessment*, 1–14. <https://doi.org/10.1177/1073191117701192>
- Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K. N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychological Assessment*, 21(3), 365–379. <https://doi.org/10.1037/a0016530>
- Raskin, R. N., & Hall, C. S. (1979). A Narcissistic Personality Inventory. *Psychological Reports*, 45(2), 590–590. <https://doi.org/10.2466/pr0.1979.45.2.590>
- Shoss, M. K., & Strube, M. J. (2011). How do you fake a personality test? An investigation of cognitive models of impression-managed responding. *Organizational Behavior and Human Decision Processes*, 116(1), 163–171. <https://doi.org/10.1016/j.obhdp.2011.05.003>
- Simonet, D. V., Tett, R. P., Foster, J., Angelback, A. I., & Bartlett, J. M. (2017). Dark-Side Personality Trait Interactions: Amplifying Negative Predictions of Leadership Performance. *Journal of Leadership & Organizational Studies*, (September), 1–18. <https://doi.org/10.1177/1548051817727703>
- Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Propriedades psicométricas na avaliação de instrumentos: Avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*, 26(3), 649–659.

<https://doi.org/10.5123/S1679-49742017000300022>

Spector, P. E., Rosen, C. C., Richardson, H. A., Williams, L. J., & Johnson, R. E. (2019).

A new perspective on method variance: A measure-Centric approach. *Journal of Management*, 45(3), 855–880.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing

and scoring pairwise preference items involving stimuli on different dimensions :

The multi-unidimensional pairwise-preference model. *Applied*, 29(3), 184–203.

<https://doi.org/10.1177/0146621604273988>

Sutton, G. (2019). *The Dark Triad in personnel selection: An exploration of narcissism,*

*psychopathy, and Machiavellianism in candidates involved in an organisational*

*selection process*. University of Canterbury.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 101(2),

266–270.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*,

33(4), 529–554.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). An introduction and a point of view.

In R. Tourangeau, L. J. Rips, & K. Rasinski (Eds.), *The psychology of survey response* (pp. 1–22). Cambridge, UK: Cambridge University Press.

Tutz, G., & Schauberger, G. (2018). *Uncertainty as Response Style in Latent Trait*

*Models*.

Usami, S., Sakamoto, A., Naito, J., & Abe, Y. (2016). Developing pairwise preference-

based personality test and experimental investigation of its resistance to faking effect

by item response model. *International Journal of Testing*, 16(4), 288–309.

<https://doi.org/10.1080/15305058.2016.1145123>

van Hooft, E. A. J., & Born, M. P. (2012). Intentional response distortion on personality

tests: Using eye-tracking to understand response processes when faking. *Journal of*



*Applied Psychology*, 97(2), 301–316. <https://doi.org/10.1037/a0025711>

- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2), 195–217. <https://doi.org/10.1093/ijpor/eds021>
- Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance*, 19(3), 175–199. [https://doi.org/10.1207/s15327043hup1903\\_1](https://doi.org/10.1207/s15327043hup1903_1)
- Vecchione, M., Alessandri, G., & Barbaranelli, C. (2013). Measurement and application of egoistic and moralistic self-enhancements. *International Journal of Selection and Assessment*, 21(2), 170–182. <https://doi.org/10.1111/ijsa.12027>
- Viswesvaran, C., & Ones, D. S. (1999). Meta-Analyses of Fakability Estimates: Implications for Personality Measurement. *Educational and Psychological Measurement*, 59(2), 197–210. <https://doi.org/10.1177/00131649921969802>
- Vize, C. E., Lynam, D. R., Collison, K. L., & Miller, J. D. (2018). Differences among dark triad components: A meta-analytic investigation. *Personality Disorders: Theory, Research, and Treatment*, 9(2), 101.
- Wang, W., Qiu, X., Chen, C., Ro, S., & Jin, K. (2017). Item response theory models for ipsative tests with Multidimensional Pairwise Comparison Items. *Applied Psychological Measurement*, 41(8), 1–14. <https://doi.org/10.1177/0146621617703183>
- Watts, A. L., Lilienfeld, S. O., Edens, J. F., Douglas, K. S., Skeem, J. L., Verschuere, B., & LoPilato, A. C. (2016). Does response distortion statistically affect the relations between self-report psychopathy measures and external criteria? *Psychological Assessment*, 28(3), 294–306. <https://doi.org/10.1037/pas0000168>
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010a). The effect of rating scale format

- on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247. <https://doi.org/10.1016/j.ijresmar.2010.02.004>
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010b). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247. <https://doi.org/https://doi.org/10.1016/j.ijresmar.2010.02.004>
- Weijters, B., Geuens, M., & Baumgartner, H. (2013). The Effect of Familiarity with the Response Category Labels on Item Response to Likert Scales. *Journal of Consumer Research*, 40(2), 368–381. <https://doi.org/10.1086/670394>
- Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956–972. <https://doi.org/10.1177/0013164404268674>
- Wetzel, E., & Greiff, S. (2018). The World Beyond Rating Scales. *European Journal of Psychological Assessment*, 34(1), 1–5. <https://doi.org/10.1027/1015-5759/a000469>
- Willits, F. K., Theodori, G. L., & Luloff, A. E. (2016). Another look at Likert scales. *Journal of Rural Social Sciences*, 31(3), 126–139. Retrieved from [http://journalofruralsocialsciences.org/pages/Articles/JRSS\\_2016\\_31/3/JRSS\\_2016\\_31\\_3\\_126-139.pdf](http://journalofruralsocialsciences.org/pages/Articles/JRSS_2016_31/3/JRSS_2016_31_3_126-139.pdf)
- Ziegler, Matthias. (2015). “F\*\*\* you, I won’t do what you told me!”—Response biases as threats to psychological assessment. *European Journal of Psychological Assessment*, 31(3), 153–158. <https://doi.org/10.1027/1015-5759/a000292>
- Ziegler, Matthias, & Buehner, M. (2009). Modeling Socially Desirable Responding and Its Effects. *Educational and Psychological Measurement*, 69(4), 548–565.

## **Desafios e alternativas na construção de escalas de autorrelato: Likert e escolha-forçada (Artigo 2)**

Ariela Raissa Lima Costa

Nelson Hauck Filho

*Universidade São Francisco*

**Resumo:** Os testes psicológicos são bastante usados por psicólogos(os) para avaliar desde personalidade até habilidades cognitivas e motoras. Entretanto, poucos são os profissionais cientes do processo envolvido na construção dos testes e sua importância para a garantia de uma avaliação criteriosa. Assim, este artigo propõe a apresentar o processo de construção de escalas de autorrelato (i.e., especificação do objetivo, elaboração e seleção de itens), usando como exemplo os traços da Tríade Sombria da Personalidade. Para isso a pesquisa consistiu na elaboração de itens e aplicação com pessoas da comunidade geral. Os itens foram avaliados quanto a seu nível de desejabilidade social e selecionados de acordo com sua capacidade de ser discriminado em cada fator. Pode-se perceber que o processo de construção de escalas é iterativo, oneroso e demanda tempo do pesquisador, devendo ser feito com embasamento teórico sólido.

**Palavras-chave:** psicometria; avaliação psicológica; neutralização; psicopatia; narcisismo; maquiavelismo.

**Abstract:** Psychological tests are widely used by psychologists to assess from personality to cognitive and motor skills. However, few professionals are aware of the process involved in the tests construction and its importance for ensuring a careful evaluation. Thus, this article proposes to present the process of building self-report scales (i.e.,

objective specification, elaboration and selection of items), using as an example the Dark Triad traits. Thus this research consisted on the elaboration of items and application with people from the general community. The items were evaluated regarding their level of social desirability and their capacity to be discriminated in each factor. The process of construction of scales is iterative, costly and demands time from the researcher, and should be done with a solid theoretical basis

**Keywords:** psychometric; psychological assessment; neutralization; psychopathy; narcissism; Machiavellianism.

**Resumen:** Los psicólogos utilizan ampliamente las pruebas psicológicas para evaluar todo, desde la personalidad hasta las habilidades cognitivas y motoras. Sin embargo, pocos profesionales conocen el proceso involucrado en la construcción de las pruebas y su importancia para garantizar una evaluación cuidadosa. Por lo tanto, este artículo propone presentar el proceso de construcción de escalas de autoinforme (es decir, especificación objetiva, elaboración y selección de ítems), utilizando como ejemplo los rasgos de la Tríada Oscura de la Personalidad. Para esto, la investigación consistió en la elaboración de artículos y su aplicación con personas de la comunidad general. Los ítems fueron evaluados por su nivel de deseabilidad social y su capacidad de ser discriminado en cada factor. Se puede ver que el proceso de construcción de escalas es interactivo, costoso y exige tiempo del investigador, y debe hacerse con una base teórica sólida

**Palabras clave:** psicometría; evaluación psicológica; neutralización psicopatía; narcisismo Maquiavelismo.

Psicólogos, de várias áreas, costumam usar escalas de autorrelato como parte do seu processo de avaliação psicológica, sejam para avaliar personalidade ou quaisquer outros atributos. Um dos fatores ignorado pelos usuários é o processo envolvido na elaboração dessas ferramentas, que é crucial para garantir a acurácia das avaliações. Esse processo, de acordo com as Normas para Testes Educacionais e Psicológicos (*Standards for Educational and Psychological Testing*; AERA, APA, NCME, 2014), é dividido em quatro grandes etapas, a saber, (a) especificação do objetivo do teste, (b) desenvolvimento e avaliação dos itens, (c) avaliação da escala formada e, (d) desenvolvimento de procedimentos de aplicação, correção e interpretação dos escores. Para cada uma das etapas existem sub-etapas. Elas vão desde a investigação na literatura sobre os conceitos e comportamentos característico do atributo de interesse, elaboração de itens, critérios de seleção de itens que irão compor a versão final da escala, definição do formato de resposta e do tamanho da escala, investigação das propriedades psicométricas até a interpretação dos escores (ver Figura 1).

Construir escalas de autorrelato para avaliar personalidade demandam um cuidado por parte do(a) pesquisador(a) para diminuir a influência de vieses de resposta. Itens de autorrelato são mais suscetíveis a vieses de respostas, em especial, o viés de desejabilidade social (Ziegler, 2015). Respostas socialmente desejáveis (RSD) consistem na seleção de uma resposta em função do quanto o respondente a considera bom ou ruim ter determinada característica quando comparado com as normas sociais do seu contexto de vida, isto é, a resposta é dada em função do aspecto valorativo em vez do aspecto descritivo (i.e., traço de interesse) do conteúdo do item (Paulhus, 2002). A presença de RSD podem interferir no processo de decisão e avaliação do avaliador, pois o escore total não refletirá o verdadeiro nível no traço de interesse (G. Krammer, Sommer, & Arendasy, 2017).

Alguns traços de personalidade são caracterizados por comportamentos indesejados pela sociedade, sendo passíveis de receberem repostas socialmente desejáveis. Neste trabalho serão destacados os traços que compõe a Tríade Sombria da Personalidade, que são Maquiavelismo, psicopatia e narcisismo (Paulhus & Williams, 2002). Os três traços são caracterizados por comportamentos antagonistas, insensibilidade, manipulação e baixa empatia. Todas essas características vão de encontro as normas sociais, em especial quando se considera a cultura brasileira que é conhecida por comportamentos pró-sociais e colaborativos.

Antes de começar o processo de construir uma escala de autorrelato é preciso definir o que se pretende avaliar e o porquê dessa avaliação (Bandalos, 2018a). A construção de uma escala é um processo longo e oneroso, por isso é preciso que o(a) pesquisador(a) defina a relevância da construção de uma nova escala. Em seguida, é preciso definir quem, como e onde se pretende avaliar. O primeiro passo nessa etapa é investigar se já existem escalas que tenham o mesmo propósito e que apresentem boas propriedades psicométricas. Em caso negativo, é mais fácil justificar a construção da escala, sendo necessários apenas indicar de que forma a construção de uma escala irá contribuir para desenvolver o conhecimento sobre o construto de interesse. Em caso positivo, se a escala estiver em outro idioma, deve-se analisar suas qualidades psicométricas e verificar a possibilidade de tradução, adaptação e validação. Caso a escala encontrada apresente limitações psicométricas, pode-se, quando possível, propor melhorias em uma nova escala ou na escala identificada.

A segunda etapa consiste na elaboração e seleção dos itens (questões ou tarefas). Neste momento, o(a) pesquisador(a) precisa ter feito uma busca extensa dos diversos comportamentos, pensamentos e sentimentos típicos do atributo de interesse para escrever itens capazes de avaliar o que se pretende. Goldberg et al. (2006) visando a contribuir para a disseminação de itens para avaliação da personalidade desenvolveu o

site *International Personality Item Pool* ([www.ipip.ori.org](http://www.ipip.ori.org)), em que são disponibilizados itens que avaliam diversos construtos. Na escrita dos itens sugere-se usar frases curtas, evitar a palavra ‘não’, evitar frases que indicam opinião (e.g., “Sou a favor do aborto legalizado”), e que apresentam duas ideias (e.g., “Gosto de ir para festas e conversar com amigos”) (Artino, Gehlbach, & Durning, 2011; Bandalos, 2018b).

Com os itens prontos, deve-se pensar como os itens serão respondidos. As recomendações de construção de escalas em formato de escolha forçada sugerem considerar o número de itens por bloco, tipos de opções de respostas, dimensionalidade dos blocos, polaridade dos itens e o sistema de correção (Brown, 2016; Heggstad, Morrison, Reeve, & McCloy, 2006). Os blocos podem ter dois, três, quatro ou mais itens, quanto maior a quantidade de itens por bloco, mais esforço cognitivo exigido. Nesse formato é solicitado que os respondentes selecionem o item que melhor lhe descreve em caso de bloco de dois itens (modelo PICK). Para blocos com três itens, solicita-se que a pessoa identifique o item que tem mais (*most*) e menos (*least*) a ver com ela (modelo MOLE), e no caso de blocos com quatro ou mais itens, solicita-se que a pessoa classifique os itens em ordem crescente do quanto se identifica com cada frase (modelo RANK). Os blocos podem avaliar um (unidimensional) ou mais fatores (multidimensional). Os instrumentos multidimensionais têm blocos em que cada item se refere a uma dimensão independente com clara distinção no conteúdo dos itens. Sendo sugerido que quanto maior o número de atributos e menor e/ou negativa for a correlação entre eles melhor a avaliação, pois a informação será influenciada majoritariamente por um traço (Brown & Maydeu-Olivares, 2011).

Esse tipo de escala foi evitado por não permitir comparação interpessoal. Porém, foram desenvolvidos sistemas de correção que propõe a solucionar essa questão, neste estudo considerou-se o sistema baseado na Teoria de Resposta ao Item e na teoria de Thurstone, sendo conhecido como modelo Thurstonian (Brown & Maydeu-Olivares,

2012). Na prática são usadas comparações binárias entre pares de itens que formam um bloco. A quantidade de comparações obtidas é calculada pela fórmula  $\tilde{n} = n(n-1)/2$  em que  $n$  corresponde ao número de itens por bloco. O modelo estima carga fatorial, variância de erro, correlação entre os atributos, o valor de *thresholds* (probabilidade de 50% da pessoa selecionar uma categoria de resposta em detrimento a outra), e permite a formulação de um escore para cada sujeito (Brown & Maydeu-Olivares, 2013).

O formato tipo Likert, apesar de aparentemente simples, requer atenção para a polaridade, numeração de categorias, rótulos, e quantidade de categorias de resposta. A polaridade da escala está relacionada ao fato dos itens serem apresentados de forma unipolar (i.e., apresenta apenas um item em um extremo da escala) ou bipolar (i.e., apresenta um item em cada extremo da escala) (Cabooter et al., 2016). A numeração das categorias indica o formato de categorias de respostas serão representadas, se apenas com itens positivos (e.g., 1 2 3 4 5) ou mistos (e.g. -3 -2 -1 0 1 2 3), sendo importante por afetar a interpretação de intensidade com que cada categoria será percebida pelos respondentes (Cabooter et al., 2016). Os rótulos referem-se aos nomes dados a cada categoria de resposta, sendo mais comum o uso das expressões que vão de “discordo totalmente” até “concordo totalmente”. Por fim, deve-se pensar na quantidade de categorias de resposta, que costumam variar de 2 a 10 (Parker et al., 2013). Essa decisão deve se basear na habilidade discriminativa da amostra investigada, pois quanto mais categorias mais esforço cognitivo é exigido dos respondentes (Weng, 2004).

Na terceira etapa são investigadas as propriedades psicométricas –fidedignidade e validade– da escala elaborada. A fidedignidade (ou confiabilidade) refere-se à coerência do teste para medir, sem erro, o atributo de interesse, ou seja, capacidade do teste de avaliar pessoas em diferentes momentos de vida obtendo dados reproduzíveis (AERA, APA, NCME, 2014; Cunha, Almeida Neto, & Stackfleth, 2016). As formas para avaliar a fidedignidade podem ser baseadas na correlação dos itens (i.e., teste-reteste, formas



paralelas e interavaliadores) e na consistência interna (i.e., coeficiente alfa e duas metades). A validade indica a qualidade da interpretabilidade dos resultados com base em evidências (AERA, APA, NCME, 2014). Os tipos de evidências de validade podem se basear no conteúdo do item, na relação com outros instrumentos, na relação com critérios externos, dimensionalidade do instrumento e transculturalidade (AERA, APA, NCME, 2014; Souza, Alexandre, & Guirardello, 2017). A quarta etapa é a demonstração da aplicabilidade da escala para usuários do instrumento por meio da interpretação dos escores latentes, indicado aos usuários o que significa cada pontuação nas escalas.

Com isso, este artigo se propõe a apresentar os dois passos iniciais (i.e., especificação do objetivo e elaboração e seleção de itens) do processo de construção de um teste de autorrelato, por meio de discussão teórica e demonstração prática da elaboração de três escalas para avaliação da Tríade Sombria da Personalidade, de forma a contribuir tanto para os que pretendem seguir esse caminho como para os usuários ficarem atentos aos instrumentos utilizados. Duas escalas serão em formato Likert, uma com e outra sem controle de desejabilidade social, a terceira escala será uma escala em formato de escolha forçada.

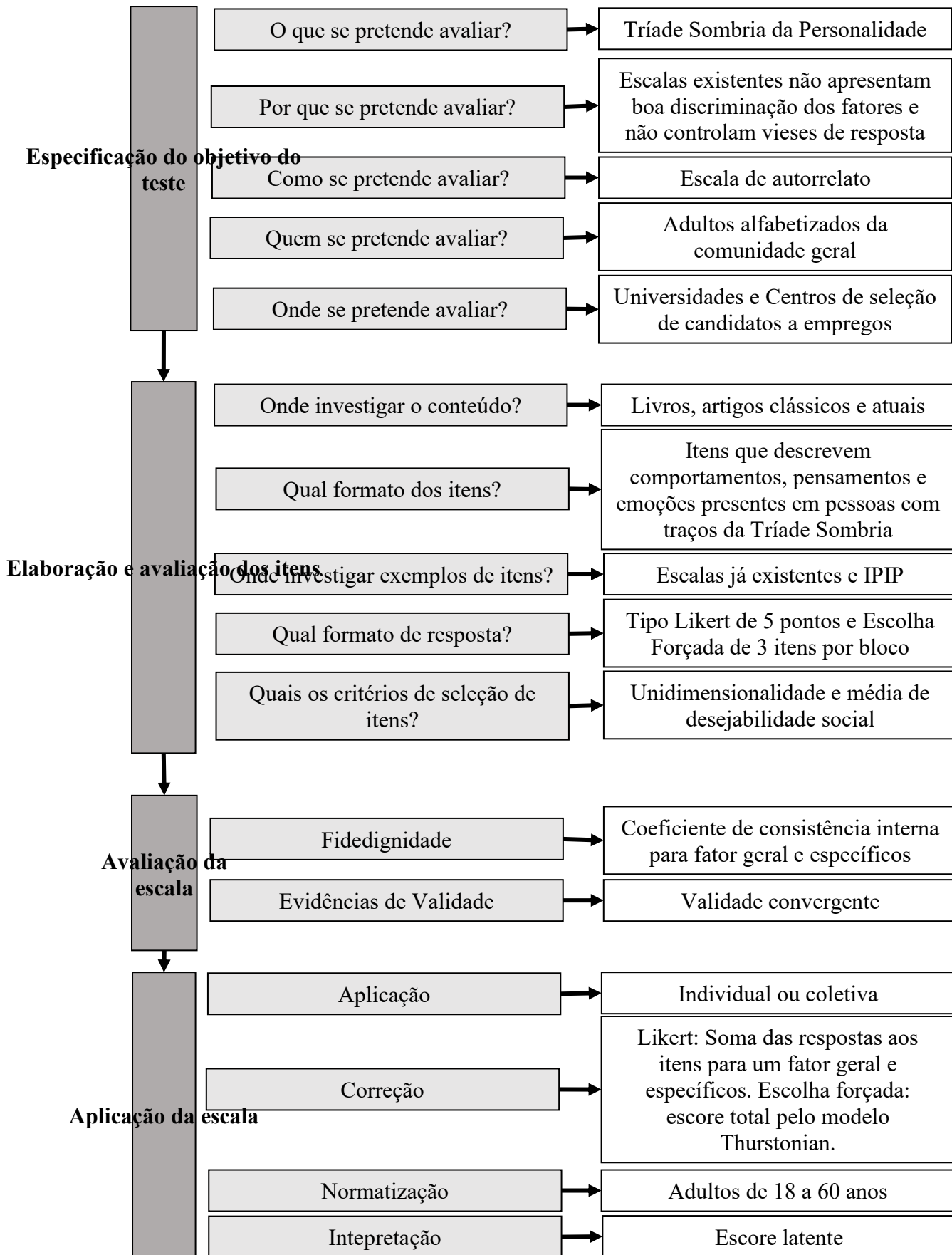


Figura 1. Etapas envolvidas no processo de construção de escalas de autorrelato.

## Método

### Participantes

**Juízes de conteúdo.** Participaram três pesquisadores e doutores da área de personalidade e psicometria.

**Amostra geral.** Na primeira coleta, participaram 417 pessoas, com idades variando entre 18 e 70 anos ( $M = 24,95$ ;  $DP = 8,94$ ), sendo que a maioria da amostra foi composta por mulheres (50,7%) e cursando ensino superior (70,7%). Na segunda coleta, participaram 647 pessoas, com idades variando entre 18 e 59 anos ( $M = 28,76$ ;  $DP = 6,14$ ), sendo que a maioria da amostra foi composta por mulheres (82,8%) e cursando pós-graduação (69,7%).

**Juízes de desejabilidade social.** Na primeira coleta, participaram 26 pessoas, com idades variando entre 18 e 63 anos ( $M = 27,68$ ;  $DP = 11,61$ ), sendo a maioria homens (65,4%) e cursando ensino superior (53,8%). Na segunda coleta, participaram 44 pessoas, com idades variando entre 20 e 63 anos ( $M = 32,25$ ;  $DP = 8,26$ ), sendo a maioria mulheres (75%) e estudantes de pós-graduação (70,5%).

### Instrumentos

Nas duas coletas foi aplicado um questionário de caracterização da amostra, com perguntas sobre a idade e sexo. Em seguida, foram aplicados os itens desenvolvidos que avaliavam os traços da Triade Sombria da Personalidade. Na primeira coleta foram aplicados 112 itens (e.g., “Sei dizer o que as pessoas querem ouvir para conseguir o que quero”) e na segunda 93 itens (e.g., “Com boas mentiras posso alcançar qualquer objetivo”). Nos dois momentos, os participantes responderam aos itens em uma escala Likert de cinco pontos, com rótulos em todas as categorias, de “*Nada a ver comigo*”, “*Pouco a ver comigo*”, “*Moderadamente a ver comigo*”, “*Muito a ver comigo*” e “*Tudo a ver comigo*”. Os juízes analisaram os mesmos itens, porém pontuando o quanto

consideravam que o conteúdo do item era algo socialmente desejável ou indesejável, para isso usaram uma escala tipo Likert de nove pontos, com rótulo apenas nos extremos, sendo 1 = “*Totalmente indesejável*” e 9 = “*Totalmente desejável*”.

### **Procedimentos de avaliação dos juízes de conteúdo**

Os itens foram enviados para juízes pesquisadores da área de personalidade para verificar a relevância do conteúdo dos itens. A avaliação das respostas dos juízes foi realizada de forma qualitativa, sendo excluídos itens que receberam avaliação negativa de todos os juízes, no que resultou em 112 itens para a primeira coleta e 93 para a segunda coleta.

### **Procedimento de aplicação**

A pesquisa foi aprovada por um Comitê de Ética em Pesquisa (CAAE: 84891918.0.0000.5514). A primeira coleta foi feita de forma presencial em uma unidade escolar pública de Educação para Jovens Adultos e em duas universidades públicas. No momento da aplicação foi explicado o objetivo da pesquisa, apresentado o Termo de Consentimento Livre e (TCLE) e solicitado que os participantes assinalassem quaisquer dúvidas que tivessem de compreensão da frase e de termos específicos. A segunda coleta foi feita de forma *online*, em que o link da pesquisa foi disponibilizado em mídias sociais (*Facebook* e *WhatsApp*). Na forma *online* foi disponibilizado um resumo com o objetivo da pesquisa junto ao convite de participação, em seguida foi apresentado o TCLE seguido dos itens. No caso dos juízes de desejabilidade social, na coleta presencial foi explicada a atividade verbalmente e disponibilizada instruções escritas exemplificando a atividade e na coleta *online* foram apresentadas apenas as instruções por escrito.

### **Análise de dados**

Nas duas aplicações foram usadas as mesmas análises. Inicialmente foi realizada análises descritivas para caracterização da amostra geral e dos juízes de desejabilidade social. Em relação as respostas da amostra geral, os dados foram analisados por meio de Análise Fatorial Exploratória, por meio do *software* Mplus, em que foram testados modelos com três e quatro fatores, com rotação oblíqua e ortogonal, com estimador *Weighted Least Squares Mean Variance and Adjusted Estimator* (WLSMV), mais apropriado para dados ordinais (Baglin, 2014). A variedade das análises se deu a fim de determinar quais os melhores itens a serem selecionados para construção das escalas. Foram selecionados itens com cargas fatoriais igual ou maior que 0,40 por serem mais discriminativos de cada dimensão avaliada. Em relação as respostas dos juízes, calculou-se a média para cada item, sendo que itens com média inferior a 4,5 foram considerados indesejáveis, itens com médias entre 4,6 e 5,5 foram considerados neutros e, itens com médias acima de 5,6 foram considerados desejáveis. A carga fatorial dos itens e a média em desejabilidade social dos itens das duas coletas foram usadas como critério para a seleção dos itens.

## **Resultados**

### **Primeiro passo: Especificação do objetivo do teste**

No caso da avaliação da Tríade Sombria, existem dois instrumentos (Jonason & Webster, 2010; Daniel N Jones & Paulhus, 2014) que apresentam problemas de ajuste desde sua criação até a replicação do modelo em outras culturas (e.g., Persson, Kajonius, & Garcia, 2017). Além disso, eles não controlam o viés da desejabilidade social e não foi encontrada versão em escolha forçada de avaliação desses atributos em conjunto, apenas para narcisismo (Raskin & Hall, 1979) e maquiavelismo (Christie & Geis, 1970), o que justifica a elaboração de uma escala em que haja o controle desse viés. Assim optou-se

por construir escalas em diferentes formatos, duas no modelo Likert e uma no formato de escolha forçada, para avaliar o nível desses traços em adultos alfabetizados.

### **Segundo passo: Elaboração e avaliação dos itens**

Para a elaboração dos itens investigou-se, na literatura, características relacionadas a cada um dos traços. Como fonte complementar, foram considerados os itens disponíveis no *International Personality Item Pool* (IPIP; <https://ipip.ori.org/>), que disponibiliza itens de atributos únicos e são liberados para uso em pesquisa. Inicialmente, do site do IPIP, selecionou-se itens que avaliavam os atributos de grandiosidade, agressividade, irresponsabilidade, manipulação, dominância, comportamento de risco, moralidade, impulsividade, sinceridade, altruísmo, vulnerabilidade e violação de normas. A tradução dos itens foi realizada pelos autores desta pesquisa. Os passos do processo de decisão de construção da escala estão apresentados na Figura 1. Para a elaboração dos itens foram considerados os conteúdos mais discriminativos de cada traço. Por exemplo, para o traço de maquiavelismo enfatizou-se itens que avaliassem comportamentos de manipulação (e.g., “Ameaço e intimido as pessoas para garantir meu poder sobre elas”). Em relação ao traço de psicopatia enfatizou-se características de tendência ao tédio, crueldade e impulsividade (e.g., “Não me sinto bem seguindo regras”). Em relação ao traço de narcisismo, enfatizou-se características relacionadas ao senso de grandiosidade e necessidade de atenção (e.g., “Gosto de ser o centro das atenções”).

Neste estudo, optou-se por usar uma escala Likert de cinco pontos, com apresentação de todos os rótulos que foram “*Nada a ver comigo*”, “*Pouco a ver comigo*”, “*Moderadamente a ver comigo*”, “*Muito a ver comigo*” e “*Tudo a ver comigo*”, sem apresentação simultânea de um número associado a um rótulo, isto é, sem indicar, por exemplo, 1 = “*Nada a ver comigo*”. Em relação os itens do formato em escolha forçada, após avaliação dos juízes, foram selecionados itens com níveis similares de desejabilidade

social. Quanto aos blocos, optou-se por blocos com três itens usando o modelo MOLE de resposta, isto é, os respondentes tinham que indicar o item que mais e o que menos lhe representassem. Essa escolha foi devido a estudos de simulação indicarem que esse formato tem um bom desempenho para estimar o escore verdadeiro (Hontangas, Leenen, de la Torre, et al., 2016; Hontangas et al., 2015).

O primeiro conjunto dos itens foram avaliados pelos juízes de conteúdo. Os principais comentários dos juízes foram em relação a escrita mais acessível para o público leigo e rejeição de alguns itens, por serem considerados inadequados para avaliar a Tríade Sombria. Dos 120 itens elaborados, foram mantidos 112 itens para aplicação. Os itens foram aplicados presencialmente em uma universidade particular, universidade pública e em uma escola pública, nas turmas de Educação de Jovens e Adultos (EJA), em São Paulo. Os participantes universitários levaram uma média de 12 minutos para responder aos itens, em que se mostraram engajados e interessados no conteúdo, questionando sobre objetivos, análises e repetição de conteúdo dos itens. Em relação ao público de EJA, a média de duração de aplicação foi de 18 minutos, em que apresentaram mais dificuldade em compreender algumas palavras como “menosprezar” e “negligenciar”, além disso, mostraram-se pouco engajados em responder aos itens.

Os itens da primeira coleta foram analisados por meio de análise fatorial exploratória de três fatores, com cargas fatoriais variando entre 0,01 e 0,82, sendo que diversos itens apresentaram cargas cruzadas. Os itens de maquiavelismo e psicopatia foram os que apresentaram mais cargas cruzadas. Pensando que teoricamente os traços de narcisismo, maquiavelismo e psicopatia apresentam mais de uma dimensão (Collison, Vize, Miller, & Lynam, 2018; Hervé, 2017; Krizan & Herlache, 2017), foi testado um modelo de cinco fatores em que foi possível uma melhor diferenciação dos fatores, sendo selecionados 55 itens. Percebeu-se, durante a análise dos itens, que nem todas as dimensões dos traços tinham sido abordadas nos itens, o que levou os autores a pensar

em um novo conjunto de itens. Os itens da segunda coleta passaram pelas mesmas análises dos itens da primeira coleta, com cargas fatoriais variando entre 0,01 e 0,83, em que a divisão em cinco fatores foi a que possibilitou uma melhor interpretação dos itens, sendo selecionados 67 itens.

A avaliação da desejabilidade social dos itens foi feita separadamente. Na primeira coleta, a maioria dos juízes eram alunos de pós-graduação com experiência na avaliação da desejabilidade social. A aplicação presencial permitiu que uma explicação detalhada da atividade, pois a aplicadora explicava com exemplos como os participantes juízes deveriam responder, assim mesmo os que não completavam a leitura das instruções tinha uma ideia clara da forma que deveria avaliar os itens. Na segunda coleta, toda a aplicação foi *online*, e as médias dos juízes para alguns itens foram incoerentes com o conteúdo apresentado, isto é, itens explicitamente indesejáveis obtiveram médias altas, como “O bem-estar das pessoas pouco me importa” que ficou com média de 6,89. Com isso, presume-se que os participantes ou não entenderam as instruções apresentadas ou não leram as instruções detalhadamente, confundido a avaliação dos itens com a autoavaliação. Na impossibilidade de considerar tais médias, foi solicitado a um grupo de alunos com experiência na atividade para reavaliar os itens, sendo as médias consideradas.

Com tais informações disponíveis, a segunda etapa foi de seleção dos itens para montar três escalas que avaliassem a Tríade Sombria da Personalidade. Duas em formato Likert, sendo uma com itens neutralizados e outra com itens indesejáveis, e uma terceira escala em formato Likert. O primeiro critério de seleção de itens foi ausência de carga cruzada. Para selecionar os itens da versão Likert neutralizada foi usada como critério adicional itens que tivessem médias de desejabilidade social próxima a 5, sendo aceito médias entre 4 e 6 (ver Tabela 1). Para a versão Likert com itens indesejáveis, foram priorizados itens que tivessem cargas superiores a 0,50 em cada fator da Tríade (ver



Tabela 3). Em relação a versão em escolha forçada, para cada bloco (que é considerado um item) foram selecionados itens com média de desejabilidade próxima, independentemente de serem todos desejáveis ou não, e tamanho similar das frases (ver Tabela 2). Devido a composição dos itens finais disponíveis, esses critérios não foram seguidos rigorosamente, especialmente no que diz respeito ao tamanho da frase.

Tabela 1.

*Versões em escala tipo Likert da medida da Tríade Sombria da personalidade*

<b>Versão A – Likert neutralizada</b>		
Item	M <sub>DS</sub>	Dimensão
Em algumas situações, alterar fatos é uma estratégia válida.	5	M+
Evito contar mentiras mesmo que eu tenha prejuízo.	5,5	M-
Tenho talento para influenciar as pessoas.	5,32	M+
Mantenho uma boa relação com pessoas de quem não gosto se for para atingir minhas metas.	4,76	M+
Ajo de forma imediata diante das situações	4,5	P+
Algumas vezes, faço coisas apenas pela adrenalina.	4,6	P+
Gosto de seguir as regras de forma rígida.	4,38	P-
Tenho dificuldade em lidar com pessoas emotivas	4,75	P+
Evito chamar a atenção das pessoas.	4,50	N-
Sempre tenho as melhores ideias.	4,50	N+
Meu medo é não ter o reconhecimento que mereço.	4,13	N+
Sei que serei alguém bem-sucedido.	4,25	N+
<b>Versão B – Likert indesejável</b>		
Item	M <sub>DS</sub>	Dimensão
Com boas mentiras posso alcançar qualquer objetivo.	1,63	M+
Acredito que para vencer na vida preciso enganar as pessoas.	1,5	M+
Agrada-me colocar duas pessoas uma contra a outra.	1,75	M+
Ameaço e intimido as pessoas para garantir meu poder sobre elas.	1,25	M+
Sinto-me desconfortável seguindo regras.	3	P+
Se alguém me desagrada, eu a retiro da minha vida por bem ou por mal.	2,63	P+
Tomo decisões de última hora sem pensar nas consequências.	2,63	P+
Se alguém tenta me impor alguma regra eu ignoro.	2,5	P+
Para mim é difícil lidar com rejeição.	3,25	N+
Gosto de pessoas que reconhecem o quanto sou especial.	3,88	N+
Os outros não conseguem entender o meu valor.	3,63	N+
Irrito-me quando alguém me critica.	3	N+

*Nota.* M<sub>DS</sub> = Média em desejabilidade social; M = Maquiavelismo; P = Psicopatia; N = Narcisismo. Os símbolos + e -, indicam se o item avalia um aspecto positivo ou negativo do traço de personalidade.

Um desafio adicional que se percebeu, no momento da seleção dos itens, foi a existência de diversos itens com conteúdo semelhante. Com isso, os autores escreveram itens adicionais sem passar pelas etapas anteriores de avaliação, mas que a literatura demonstra serem representativos de cada traço. Além disso, para testar a qualidade dos blocos de escolha forçada foram adicionados itens *bugs*, que são itens não relacionados

com os atributos avaliados. Os itens *bugs* avaliam Abertura a Experiência, pois evidências sugerem uma relação não significativa com os traços da Tríade (e.g., Vize, Lynam, Collison, & Miller, 2018).

Tabela 2.

*Versões em escala de Escolha-Forçada da medida da Tríade Sombria da personalidade*

	Frases	M <sub>DS</sub>	Dimensão
Item 1	Finjo estar preocupado com as pessoas para conseguir algo.	2,13	M
	Para mim os bens materiais são mais importantes do que as pessoas.	2,13	P
	Sou a melhor pessoa para ditar como as coisas devem ser.	3,25	N
Item 2	Altero os fatos de propósito.	2,25	M
	Detesto seguir ordens.	3,00	P
	Mereço tratamento especial dos outros.	3,00	N
Item 3	Relaciono-me apenas com pessoas que me beneficiam.	3,88	M
	A aprovação dos outros é muito importante para mim.	3,63	N
	A maioria das pessoas me entendia.	3	P
Item 4	Fico irritado quando alguém me critica.	3	N
	Mentir é a melhor estratégia para se livrar de problemas.	3,25	M
	Geralmente me arrependo por não pensar antes de agir	4	P
Item 5	Arranjo inimigos com mais facilidade do que amigos.	2,88	P
	Uso os outros para meus próprios objetivos.	2,68	M
	Irrito-me quando alguém discorda de mim.	2,92	N
Item 6	Gosto apenas de atividades que me façam sentir adrenalina e fortes emoções.	4,30	P
	Sinto dificuldade em aceitar quando alguém fala que tenho defeitos.	4,36	N
	Identifico a pessoa mais influente em um lugar para ser próximo e obter benefícios.	4,38	M
Item 7	As pessoas deveriam agradecer por ter a minha companhia.	3,25	N
	Tenho dificuldade para acatar ordens de superiores e outras pessoas de autoridade.	3,65	P
	Consigo facilmente fazer uma pessoa se sentir culpada.	3,00	M
Item 8	Nem sempre estou satisfeito comigo mesmo.	4,00	N
	Sem perceber, acabo repetindo erros que cometi.	4,40	P
	Decoro pequenas informações para parecer mais inteligente.	3,38	M
Item 9	Falo o que as pessoas querem ouvir para conseguir o que eu quero delas.	2,38	M
	Viajar é perda de tempo	2,75	A
	Divirto-me com piadas cruéis.	2,43	P
Item 10	Costumo mentir quando é oportuno.	2,25	M
	Se alguém tenta me impor alguma regra eu ignoro.	2,50	P
	Sou tão brilhante que deveria ser pago só por existir.	2,75	N
Item 11	Acredito que vale de tudo para manter o poder.	2,63	M
	Tenho dificuldade em seguir regras.	2,75	P
	Gosto de ser o centro das atenções.	3,25	N
Item 12	Detesto discussões filosóficas	3,00	A
	Para mim é difícil lidar com rejeição.	3,25	N
	Não me sinto bem seguindo regras.	3,00	P
Item 13	Sou bom em usar os pontos fracos das pessoas para ganho próprio.	2,13	M
	Detesto aprender coisas novas	2,25	A
	Irrito-me quando alguém critica meu trabalho.	3,13	N
Item 14	Tenho pouco interesse em ter ideias originais	3,25	A
	Admiro pessoas mestres em manipulação.	2,38	M
	Planejar é perda de tempo.	3,25	P
Item 15	Evito ler livros complexos	2,75	A
	Para mim é insuportável ser esquecido pelos outros.	3,13	N

	Penso que honestidade raramente é vantajosa.	2,25	M
Item 16	Gosto de criar coisas novas, mesmo que inúteis	4,25	A
	Sei que serei alguém bem-sucedido.	4,25	N
	Gosto de seguir as regras de forma rígida.	4,38	P

*Nota.* M<sub>DS</sub> = Média em desejabilidade social; M = Maquiavelismo; P = Psicopatia; N = Narcisismo; A =

Abertura a experiência.

### **Terceiro e quarto passo: Avaliação e Aplicação da escala**

No presente estudo, foi verificado apenas a dimensionalidade dos itens para selecionar os mais representativos dos atributos avaliados, sem informações sobre a interpretação dos escores latentes. Entretanto, de forma geral, entende-se que quanto maior a pontuação da pessoa em cada escala, maior o nível de traços da Tríade Sombria.

### **Discussão**

O objetivo deste trabalho foi demonstrar parte do processo envolvido na construção de escalas de autorrelato. Para tanto, foram apresentadas as etapas desenvolvidas na construção de três escalas que visavam a avaliação dos traços da Tríade Sombria da Personalidade, maquiavelismo, psicopatia e narcisismo. Cada escala desenvolvida teve seu próprio critério de seleção de itens para a versão Likert neutralizada selecionou-se itens com desejabilidade social neutra (i.e., média entre 4,5 e 5,5), para versão Likert indesejável selecionou-se itens com cargas fatoriais mais discriminativas de cada fator e, na versão de escolha forçada selecionou-se itens com médias similares em desejabilidade social.

A escolha das palavras e a forma dos itens podem interferir no engajamento do participante para responder ao instrumento. Como foi observado, pessoas com menor nível de educação formal tiveram mais dificuldade de abstrair o significado de algumas palavras, bem como consideraram enfadonha a atividade de endossar os itens. Por outro lado, estudantes de nível superior demonstraram maior curiosidade sobre o processo de elaboração dos itens, bem como dos possíveis resultados, provavelmente por estarem

mais próximos de um ambiente em que a pesquisa é atividade comum. Assim, o público-alvo do instrumento a ser construído deve ser considerado com prioridade no momento de decidir que tipo de item e de escala de resposta para garantir um maior engajamento e diminuir indecisão nas respostas, ou seja, diminuir a probabilidade de respostas descuidadas (Tutz & Schauburger, 2018). Ao comparar Likert e escolhas-forçadas, estudos demonstram que este último demanda maior esforço cognitivo (Christiansen et al., 2005). Outro exemplo, crianças tem mais facilidade de responder a imagens gráficas do que em frases que indicam o quanto elas concordam/discordam dos itens.

De forma geral, este estudo se propôs a demonstrar de maneira prática como se dá o processo de construção de escalas em formato Likert e escolha-forçada. As etapas finais desse processo, de validação e normatização das escalas, serão apresentadas nos estudos seguintes. Essa discussão é importante pela pouca aproximação que os profissionais da prática têm com o processo de construção dos instrumentos utilizados. Entretanto, alguns pontos-chaves nesse processo, como vieses de respostas (para mais informações consultar Van Vaerenbergh & Thomas, 2013) e aplicação piloto, foram poucos explorados em favor de pontos mais genéricos. Porém, isso apenas identifica a necessidade de produção científica que demonstre, de forma prática, como considerar tais elementos relacionados a construção das escalas.

O processo de construção de instrumentos de autorrelato é desafiante, pois há necessidade de garantir a qualidade dos itens. Assim, apesar de os passos serem bem definidos, a construção de um teste é um processo iterativo. Isso significa que em várias ocasiões são necessárias revisões e retorno a alguma das etapas (Bandalos, 2018a). Mesmo testes que já existem e são validados para uso, seja para pesquisa ou avaliação profissional, precisam ser revisados. Isso acontece pelo fato de o conhecimento ser dinâmico e alterações culturais influenciarem na forma como as pessoas respondem aos itens, por isso a importância de se investigar como foi o processo de elaboração dos

instrumentos utilizados para avaliação, principalmente, se existe uma finalidade diagnóstica ou de seleção.

### Referências

- Ackerman, R. A., Donnellan, M. B., Roberts, B. W., & Fraley, R. C. (2016). The effect of response format on the psychometric Properties of the Narcissistic Personality Inventory: Consequences for item meaning and factor structure. *Assessment, 23*(2), 203–220. <https://doi.org/10.1177/1073191114568113>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC, US: American Psychological Association.
- Anglim, J., Bozic, S., Little, J., & Lievens, F. (2018). Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings. *Advances in Health Sciences Education, 23*(2), 311–321. <https://doi.org/10.1007/s10459-017-9796-8>
- Artino, A. R. J., Gehlbach, H., & Durning, S. J. (2011). AM Last Page: Avoiding Five Common Pitfalls of Survey Design. *Academic Medicine, 86*(10), 1327. <https://doi.org/10.1097/ACM.0b013e31822f77cc>
- Bäckström, M., & Björklund, F. (2013). Social desirability in personality inventories: symptoms, diagnosis and prescribed cure. *Scandinavian Journal of Psychology, 54*(2), 152–159. <https://doi.org/10.1111/sjop.12015>
- Bäckström, M., & Björklund, F. (2014). *Social Desirability in Personality Inventories The Nature of the Evaluative Factor*. <https://doi.org/10.1027/1614-0001/a000138>
- Bäckström, M., Björklund, F., & Larsson, M. R. (2012). Social desirability in personality assessment. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives*

- on faking in personality assessment* (pp. 201–213). New York, NY: Oxford University Press, Inc.
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation, 19*(5). <https://doi.org/10.1146/annurev.psych.53.100901.135239>.
- Baka, A., & Figgou, L. (2012). “Neither agree, nor disagree”: A critical analysis of the middle answer category in Voting Advice Applications. *International Journal of Electronic Governance, 5*(3/4), 244–263. <https://doi.org/10.1504/IJEG.2012.051306>
- Bandalos, D. L. (2018a). The test development process. In D. L. Bandalos (Ed.), *Measurement Theory and Applications for the Social Sciences* (pp. 41–84). New York, NY: The Guilford Press.
- Bandalos, D. L. (2018b). Writing noncognitive items. In D. L. Bandalos (Ed.), *Measurement Theory and Applications for the Social Sciences* (pp. 85–119). New York, NY: The Guilford Press.
- Baughman, H. M., Jonason, P. K., Lyons, M., & Vernon, P. A. (2014). Liar liar pants on fire: Cheater strategies linked to the Dark Triad. *Personality and Individual Differences, 71*(71), 35–38. <https://doi.org/10.1016/j.paid.2014.07.019>
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*(4), 317–335. <https://doi.org/10.1111/j.1468-2389.2006.00354.x>
- Blasberg, S. A., Rogers, K. H., & Paulhus, D. L. (2014). The bidimensional impression management index (BIMI): Measuring agentic and communal forms of impression management. *Journal of Personality Assessment, 96*(5), 523–531. <https://doi.org/10.1080/00223891.2013.862252>
- Brown, A., & Maydeu-olivares, A. (2016). Modeling forced-choice response formats. In

- P. Irwing, T. Booth, & D. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing* (pp. 1–64). London: John Wiley & Sons, Inc.
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*(3), 460–502. <https://doi.org/10.1177/0013164410375112>
- Brown, Anna. (2016). Item response models for forced-choice questionnaires: A common framework. *Psychometrika, Vol. 81*, pp. 135–160. <https://doi.org/10.1007/s11336-014-9434-9>
- Brown, Anna, & Maydeu-Olivares, A. (2012). Fitting a thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods, 44*(4), 1135–1147. <https://doi.org/10.3758/s13428-012-0217-x>
- Brown, Anna, & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods, Vol. 18*, pp. 36–52. <https://doi.org/10.1037/a0030641>
- Cabooter, E., Weijters, B., Geuens, M., & Vermeir, I. (2016). Scale format effects on response option interpretation and use. *Journal of Business Research, 69*(7), 2574–2584. <https://doi.org/10.1016/j.jbusres.2015.10.138>
- Cattell, R. B. (1958). What is “objective” in “objective personality tests?” *Journal of Counseling Psychology, 5*(4), 285–289. <https://doi.org/10.1037/h0046268>
- Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309–336). New York, NY: Routledge/Taylor & Francis Group.
- Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika, 30*(1), 99–121. <https://doi.org/10.2333/bhmk.30.99>
- Christiansen, N. D., Burns, G. N., & George, E. (2005). Reconsidering forced-choice item

- formats for applicant personality assessment. *Human Performance*, *18*(3), 267–307.  
[https://doi.org/10.1207/s15327043hup1803\\_4](https://doi.org/10.1207/s15327043hup1803_4)
- Christie, R., & Geis, F. L. (1970). *Studies in machiavellianism*.  
<https://doi.org/10.1016/B978-0-12-174450-2.50008-7>
- Collison, K. L., Vize, C. E., Miller, J. D., & Lynam, D. R. (2018). Development and preliminary validation of a Five Factor Model measure of Machiavellianism. *Psychological Assessment*, *30*(10).
- Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal Of Consulting Psychology*, *24*(4), 349–354.  
<https://doi.org/10.1037/h0047358>
- Cunha, C. M., Almeida Neto, O. P., & Stackfleth, R. (2016). Principais métodos de avaliação psicométrica da confiabilidade de instrumentos de medida. *Revista de Atenção à Saúde*, *14*(49), 98–103. <https://doi.org/10.13037/rbcs.vol14n49.3671>
- Dolnicar, S., & Grün, B. (2013). “Translating” between survey answer formats. *Journal of Business Research*, *66*(9), 1298–1306.  
<https://doi.org/10.1016/j.jbusres.2012.02.029>
- Donovan, J., & Dwright, S. A. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal Of Business and Psychology*, *29*(September 2013), 479–493. <https://doi.org/10.1007/s10869-013-9318-5>
- Dragow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology*, *21*(3), 465–476.  
<https://doi.org/http://dx.doi.org/10.5380/psi.v21i3.53054>
- Dubas, J. S., Baams, L., Doornwaard, S. M., & Aken, M. A. G. Van. (2017). Dark Personality Traits and Impulsivity Among Adolescents: Differential Links to Problem Behaviors and Family Relations. *Journal of Abnormal Child Psychology*,



126(7), 877–889.

- Ellingson, J. E. (2012). People fake only when they need to fake. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 19–33). New York, NY: Oxford University Press, Inc.
- Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social and Personality Psychology Compass*, Vol. 7, pp. 199–216. <https://doi.org/10.1111/spc3.12018>
- Geel, M. Van, Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences*, 106, 231–235. <https://doi.org/10.1016/j.paid.2016.10.063>
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96.
- Hauenstein, N. M. A., Bradley, K. M., O’Shea, P. G., Shah, Y. J., & Magill, D. P. (2017). Interactions between motivation to fake and personality item characteristics: Clarifying the process. *Organizational Behavior and Human Decision Processes*, 138, 74–92. <https://doi.org/10.1016/j.obhdp.2016.11.002>
- He, J., van de Vijver, F. J. R., Dominguez Espinosa, A., Abubakar, A., Dimitrova, R., Adams, B. G., ... Villieux, A. (2015). Socially Desirable Responding. *Cross-Cultural Research*, 49(3), 227–249. <https://doi.org/10.1177/1069397114552781>
- Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, Vol. 91, pp. 9–24. <https://doi.org/10.1037/0021-9010.91.1.9>
- Hervé, H. (2017). Psychopathy across the ages: A history of the Hare psychopath. In H.

- Hugues & J. C. Yuille (Eds.), *The Psychopath: Theory, Research, and Practice* (pp. 31–55). Routledge.
- Hogan, R. (1991). Personality and personality measurement. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 873–919). Palo Alto, CA: Consulting Psychologists Press.
- Hontangas, P. M., Leenen, I., de la Torre, J., Ponsoda, V., Morillo, D., & Abad, F. J. (2016). Traditional scores versus IRT estimates on forced-choice tests based on a dominance model. *Psicothema*, Vol. 28, pp. 76–82. Hontangas, Pedro M.: Facultad de Psicología, Universidad de Valencia, Valencia, Spain, 46011, hontanga@uv.es: Colegio Oficial de Psicólogos del Principado de Asturias.
- Hontangas, P. M., Leenen, I., Torre, J. De, Ponsoda, V., & Morillo, D. (2016). Traditional scores versus IRT estimates on forced-choice tests based. *Psicothema*, 28(1), 76–82. <https://doi.org/10.7334/psicothema2015.204>
- Hontangas, P. M., Torre, J. De, Ponsoda, V., Leenen, I., Morillo, D., & Abad, F. J. (2015). Comparing traditional and IRT scoring of forced-choice tests. *Applied Psychological Measurement*, 39(8), 598–612. <https://doi.org/10.1177/0146621615585851>
- Johnson, J. A. (2004). The impact of item characteristics on item and scale validity. *Multivariate Behavioral Research*, 39(2), 273–302. <https://doi.org/10.1207/s15327906mbr3902>
- Jonason, P. K., & Fletcher, S. A. (2018). Agentic and communal behavioral biases in the Dark Triad traits. *Personality and Individual Differences*, 130(August), 76–82. <https://doi.org/10.1016/j.paid.2018.03.044>
- Jonason, P. K., Lyons, M., Baughman, H. M., & Vernon, P. A. (2014). What a tangled web we weave: The dark triad traits and deception. *Personality and Individual Differences*, 70, 117–119. <https://doi.org/10.1016/j.paid.2014.06.038>
- Jonason, P. K., Strosser, G. L., Kroll, C. H., Duineveld, J. J., & Baruffi, S. A. (2015).

- Valuing myself over others: The Dark Triad traits and moral and social values. *Personality and Individual Differences*, 81, 102–106. <https://doi.org/10.1016/j.paid.2014.10.045>
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, Vol. 22, pp. 420–432. <https://doi.org/10.1037/a0019265>
- Jones, D. N., & Paulhus, D. (2009). Machiavellianism. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 93–108). The Guilford Press.
- Jones, Daniel N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3). *Assessment*, 21(1), 28–41. <https://doi.org/10.1177/1073191113514105>
- Jones, Daniel N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, 2(1), 28–41. <https://doi.org/10.1177/1073191113514105>
- Kline, P. (2015). Making test reliable II: Personality inventories. Item writing. In P. Kline (Ed.), *A handbook of test construction: Introduction to psychometric design* (pp. 24–76). Routledge/Taylor & Francis Group.
- Kowalski, C. M., Vernon, P. A., & Schermer, J. A. (2019). The Dark Triad and facets of personality. *Current Psychology*, (November). <https://doi.org/10.1007/s12144-019-00518-0>
- Krammer, G., Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment*, 99(5), 510–523. <https://doi.org/https://doi.org/10.10>
- Krammer, Georg, Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations

- across response conditions. *Journal of Personality Assessment*, 99(5), 510–523.  
<https://doi.org/10.1080/00223891.2017.1285781>
- Krizan, Z., & Herlache, A. D. (2017). The narcissism spectrum model: A synthetic view of narcissistic personality. *Personality and Social Psychology Review*, 1–29.  
<https://doi.org/10.1177/1088868316685018>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22 140, 55. <https://doi.org/2731047>
- Luft, J., & Ingham, H. (1961). The johari window. *Human Relations Training News*, 5(1), 6–7.
- MacCann, C., Ziegler, M., & Roberts, R. D. (2012). Faking in personality assessment: reflections and recommendations. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 309–329). New York, NY: Oxford University Press, Inc.
- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, 136(3), 450–470. <https://doi.org/10.1037/a0019216>
- Morey, L. C. (2012). Detection of Response Bias in Applied Assessment: Comment on McGrath et al. (2010). *Psychological Injury and Law*, 5(3–4), 153–161.  
<https://doi.org/10.1007/s12207-012-9131-x>
- Müller, G., Welter, R., & Capitão, C. G. (2007). Medidas ipsativas na avaliação psicológica. *Avaliação Psicológica*, 6(2), 157–165.
- Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology*, 142(2), 71–89. <https://doi.org/10.1080/00221309.2014.994590>
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied*

*Psychology*, 81(6), 660–679. <https://doi.org/10.1037/0021-9010.81.6.660>

Parker, R. I., Vannest, K. J., & Davis, J. L. (2013). Reliability of multi-category rating scales. *Journal of School Psychology*, 51(2), 217–229. <https://doi.org/10.1016/j.jsp.2012.12.003>

Paulhus, D. L. (2002). Socially desirable responding: the evolution of a construct. In H. I. Brown, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.

Paulhus, D. L. (2006). *Comprehensive Inventory of Desirable Responding (CIDR)*. New Orleans.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, 66(6), 1025–1060. <https://doi.org/10.1111/1467-6494.00041>

Paulhus, D. L., Neumann, C. S., Hare, R. D., Williams, K. M., & Hemphill, J. F. (2016). *Self-report Psychopathy Scale - 4* (4th ed.). Toronto, ON, Canada: Multi-Healthy Systems.

Paulhus, D. L. (1984). Two-Component Models of Socially Desirable Responding. *Personality Processes and Individual Differences*, 46(3), 598–609.

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)

Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences*, 35(2), 263–275. [https://doi.org/10.1016/S0191-8869\(02\)00187-3](https://doi.org/10.1016/S0191-8869(02)00187-3)

Persson, B. N., Kajonius, P. J., & Garcia, D. (2017). Revisiting the structure of the Short Dark Triad. *Assessment*, 1–14. <https://doi.org/10.1177/1073191117701192>

Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K.

- N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychological Assessment*, 21(3), 365–379. <https://doi.org/10.1037/a0016530>
- Raskin, R. N., & Hall, C. S. (1979). A Narcissistic Personality Inventory. *Psychological Reports*, 45(2), 590–590. <https://doi.org/10.2466/pr0.1979.45.2.590>
- Shoss, M. K., & Strube, M. J. (2011). How do you fake a personality test? An investigation of cognitive models of impression-managed responding. *Organizational Behavior and Human Decision Processes*, 116(1), 163–171. <https://doi.org/10.1016/j.obhdp.2011.05.003>
- Simonet, D. V., Tett, R. P., Foster, J., Angelback, A. I., & Bartlett, J. M. (2017). Dark-Side Personality Trait Interactions: Amplifying Negative Predictions of Leadership Performance. *Journal of Leadership & Organizational Studies*, (September), 1–18. <https://doi.org/10.1177/1548051817727703>
- Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Propriedades psicométricas na avaliação de instrumentos: Avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*, 26(3), 649–659. <https://doi.org/10.5123/S1679-49742017000300022>
- Spector, P. E., Rosen, C. C., Richardson, H. A., Williams, L. J., & Johnson, R. E. (2019). A new perspective on method variance: A measure-Centric approach. *Journal of Management*, 45(3), 855–880.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model. *Applied*, 29(3), 184–203. <https://doi.org/10.1177/0146621604273988>
- Sutton, G. (2019). *The Dark Triad in personnel selection: An exploration of narcissism, psychopathy, and Machiavellianism in candidates involved in an organisational*

*selection process*. University of Canterbury.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *101*(2), 266–270.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*(4), 529–554.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). An introduction and a point of view. In R. Tourangeau, L. J. Rips, & K. Rasinski (Eds.), *The psychology of survey response* (pp. 1–22). Cambridge, UK: Cambridge University Press.

Tutz, G., & Schauberger, G. (2018). *Uncertainty as Response Style in Latent Trait Models*.

Usami, S., Sakamoto, A., Naito, J., & Abe, Y. (2016). Developing pairwise preference-based personality test and experimental investigation of its resistance to faking effect by item response model. *International Journal of Testing*, *16*(4), 288–309. <https://doi.org/10.1080/15305058.2016.1145123>

van Hooft, E. A. J., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*(2), 301–316. <https://doi.org/10.1037/a0025711>

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217. <https://doi.org/10.1093/ijpor/eds021>

Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance*, *19*(3), 175–199. [https://doi.org/10.1207/s15327043hup1903\\_1](https://doi.org/10.1207/s15327043hup1903_1)

Vecchione, M., Alessandri, G., & Barbaranelli, C. (2013). Measurement and application of egoistic and moralistic self-enhancements. *International Journal of Selection and*

*Assessment*, 21(2), 170–182. <https://doi.org/10.1111/ijjsa.12027>

- Viswesvaran, C., & Ones, D. S. (1999). Meta-Analyses of Fakability Estimates: Implications for Personality Measurement. *Educational and Psychological Measurement*, 59(2), 197–210. <https://doi.org/10.1177/00131649921969802>
- Vize, C. E., Lynam, D. R., Collison, K. L., & Miller, J. D. (2018). Differences among dark triad components: A meta-analytic investigation. *Personality Disorders: Theory, Research, and Treatment*, 9(2), 101.
- Wang, W., Qiu, X., Chen, C., Ro, S., & Jin, K. (2017). Item response theory models for ipsative tests with Multidimensional Pairwise Comparison Items. *Applied Psychological Measurement*, 41(8), 1–14. <https://doi.org/10.1177/0146621617703183>
- Watts, A. L., Lilienfeld, S. O., Edens, J. F., Douglas, K. S., Skeem, J. L., Verschuere, B., & LoPilato, A. C. (2016). Does response distortion statistically affect the relations between self-report psychopathy measures and external criteria? *Psychological Assessment*, 28(3), 294–306. <https://doi.org/10.1037/pas0000168>
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010a). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247. <https://doi.org/10.1016/j.ijresmar.2010.02.004>
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010b). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247. <https://doi.org/https://doi.org/10.1016/j.ijresmar.2010.02.004>
- Weijters, B., Geuens, M., & Baumgartner, H. (2013). The Effect of Familiarity with the Response Category Labels on Item Response to Likert Scales. *Journal of Consumer Research*, 40(2), 368–381. <https://doi.org/10.1086/670394>



- Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64*(6), 956–972. <https://doi.org/10.1177/0013164404268674>
- Wetzel, E., & Greiff, S. (2018). The World Beyond Rating Scales. *European Journal of Psychological Assessment, 34*(1), 1–5. <https://doi.org/10.1027/1015-5759/a000469>
- Willits, F. K., Theodori, G. L., & Luloff, A. E. (2016). Another look at Likert scales. *Journal of Rural Social Sciences, 31*(3), 126–139. Retrieved from [http://journalofruralsocialsciences.org/pages/Articles/JRSS\\_2016\\_31/3/JRSS\\_2016\\_31\\_3\\_126-139.pdf](http://journalofruralsocialsciences.org/pages/Articles/JRSS_2016_31/3/JRSS_2016_31_3_126-139.pdf)
- Ziegler, Matthias. (2015). “F\*\*\* you, I won’t do what you told me!”—Response biases as threats to psychological assessment. *European Journal of Psychological Assessment, 31*(3), 153–158. <https://doi.org/10.1027/1015-5759/a000292>
- Ziegler, Matthias, & Buehner, M. (2009). Modeling Socially Desirable Responding and Its Effects. *Educational and Psychological Measurement, 69*(4), 548–565.

## Measuring Dark traits using single and multiple stimulus strategies (Artigo 3)

Ariela Raissa Lima Costa

Nelson Hauck Filho

*Universidade São Francisco*

**Abstract:** The common measures of Dark Triad traits use Likert scales that are well known for being vulnerable to social desirability bias (SD). In this article, we propose new measures to assess Dark Triad Traits, including alternative methods of control of SD, like neutralized items and a forced-choice measure. For this, 614 participated ( $M = 29.8$  years;  $SD = 6.49$ ), being 75.9% women. The results suggest that a two-factor solution is adequate to explain the dark triad, where Machiavellianism and Psychopathy collapsed into one factor. Also, it demonstrates that forced-choice is a useful tool to assess undesirable traits, but it requires more efforts from the researcher.

**Keywords:** ipsative; single stimulus; personality; dark triad; psychometrics.

**Resumo:** As medidas mais comuns da Tríade Sombria usam escalas Likert, que são conhecidas por serem vulneráveis ao viés de desejabilidade social (DS). Neste artigo, propomos novas medidas para avaliar as características da tríade sombria, incluindo métodos alternativos de controle de DS, como itens neutralizados e uma medida de escolha forçada. Para isso, 614 pessoas participaram ( $M = 29,8$  anos;  $DP = 6,49$ ), sendo 75,9% mulheres. Os resultados sugerem que uma solução de dois fatores é mais adequada para explicar os traços da tríade sombria, sendo que Maquiavelismo e psicopatia formaram um único fator. Além disso, é demonstrado que o uso de escolha forçada é uma ferramenta útil para avaliar características indesejáveis, mas esse formato requer mais atenção do pesquisador.

**Palavras-chave:** ipsativa; estímulo único; personalidad; tríade sombria; psicometria.

**Resumen:** Las medidas comunes de los rasgos de la Tríada oscura usan escalas Likert que son bien conocidas por ser vulnerables al sesgo de deseabilidad social (DE). En este artículo, proponemos nuevas medidas para evaluar los Rasgos de la Tríada Oscura, incluidos métodos alternativos de control de SD, como elementos neutralizados y una medida de elección forzada. Para esto, participaron 614 ( $M = 29.8$  años;  $SD = 6.49$ ), siendo el 75.9% mujeres. Los resultados sugieren que una solución de dos factores es adecuada para explicar la tríada oscura, donde el maquiavelismo y la psicopatía formaron un factor. Además, demuestra que la elección forzada es una herramienta útil para evaluar los rasgos indeseables, pero requiere más atención por parte del investigador.

**Palabras llave:** ipsativo; estímulo único; personalidad; tríada oscura psicometría

Personality self-report inventories have been used in diverse contexts, such as educational assessment and organizational settings. Self-reports are popular given their ease of application and correction, cost- and time-effectiveness, and its capability to predict several outcomes (e.g., Dubas, Baams, Doornwaard, & Aken, 2017; Simonet, Tett, Foster, Angelback, & Bartlett, 2017). Nonetheless, methodological researches demonstrate fragilities in this method of data collection (Ziegler, 2015). These fragilities are mainly focused on self-reports that use Likert-type scales as response format and its vulnerability to response bias (He et al., 2015). The social desirability bias refers to the manipulation of the answer in accordance with perceived social values, which means that the examinee answers the inventories according to her/his perception about what is socially valued in the context of assessment (Paulhus, 2002). This growing concern with

social desirability led researchers to try new ways of writing items, like neutralization process (Bäckström & Björklund, 2013) and recovering an old response format, the forced-choice (FC) (Brown & Maydeu-Olivares, 2016).

The neutralization process consists of re-writing the items controlling for its evaluative aspect. Most items in self-reports are written without the concern about how the selected words can be interpreted within the culture and assessment context of the examinee (Bäckström & Björklund, 2014; Bäckström et al., 2012). So this procedure suggests that when creating or re-writing an item, the research should observe how desirable or undesirable each word is, for example, “I am disorganized,” the term ‘disorganized’ bring a negative aspect, is something that is not good to admit. Still, if the researcher writes “I tend to let my stuff out of place” the attribute is preserved, and the evaluative aspect is reduced, which led the examinee to focus more on the content instead of on the desirability of the item.

Historically, forced-choice scales emerged before Likert scales, which were created as an easy alternative do FC. The FC was developed by Thurstone (1927), he proposed the idea that all psychological attributes are distributed in a continuum and, when someone reads an item, she/he compares her/his own level on the specific attribute with the one represented in the item. Thus, a FC response format was created, where someone’s personality is assessed by presenting items with a different level on the continuum, but with similar evaluative content (i.e., similar levels of social desirability) in a block. Each block must have two or more items, and the examinee must select which is closest to her/his level in the assessed attribute or rank the items from the one that most describe her/him to the one that least describes. So, the advantage of this format is that the examinee analyzes the items as a group; they are not analyzed individually, which reduces the possibility of manipulation of the answers.

A methodological concern regarding this format, and the main reason why it was ignored for a long time, is because it generates ipsative scores. This means that all examinees receive the same score, regardless of their item's choice, which allows only intraindividual comparisons and making impossible the formulation of normative tables. However, due to technological and theoretical advances, today is possible to find statistical analysis, based on Item Response Theory (IRT), that make possible generate scores that allow interindividual comparisons (e.g., Brown & Maydeu-Olivares, 2012; Stark, Chernyshenko, & Drasgow, 2005).

Some personality traits reinforce researchers' worries, such as the Dark Triad (DT) traits (i.e., Machiavellianism, Psychopathy, and Narcissism) (Paulhus & Williams, 2002); these traits raise a twofold concern. First, the content of items used to assess these traits usually is undesirable, which increases the probability that respondents do not provide an honest answer, particularly in high stakes contexts (Sutton, 2019). Second, people with high levels of these traits might fake more than people with low levels, considering that they tend to manipulate and lie even without reason (Baughman et al., 2014). Hence, in this article is proposed two new measures of the DT traits, one in Likert format with neutralized and non-neutralized items and other on the forced-choice format.

## **Method**

### **Participants**

A total of 614 people answered the research protocol, with ages between 18 and 66 ( $M = 29.8$  years;  $SD = 6.49$ ), being 75.9% women. Concerning race, 63.3% declare themselves as white, 24.7% as brown, and 9.8% as black. Concerning education, 71% claimed having completed a graduate degree, 21% completed undergraduate degree, and 7.6% completed high-school degree. Concerning the civil state, 52.1% declared themselves as single, 18,5% as married, and 16.7% living with a partner. The sample had

representatives from 25 Brazilian states, being 2.9% from North, 19.2% from Northeast, 5.8% from Midwest, 51.9% from Southeast, and 20.2% from South.

## **Instruments**

**Short Dark Triad** (Daniel N. Jones & Paulhus, 2014). The SD3 assess Machiavellianism, Psychopathy, and Narcissism, with nine items per trait, totalizing 27 items. These traits were considered to share some common characteristics like lack of empathy and manipulation. In this study, the internal consistency value of the total scale was  $\alpha = .856$  ( $\omega = .868$ ), and for each subscale was  $\alpha = .828$  ( $\omega = .834$ ) for Machiavellianism,  $\alpha = .693$  ( $\omega = .743$ ) for Psychopathy, and  $\alpha = .730$  ( $\omega = .733$ ) for Narcissism.

**Experimental Dark Traits Scale – Likert version.** This scale was constructed by the authors, following the process presented in Article 2. For this scale, each dark trait, Machiavellianism, Psychopathy, and Narcissism, was composed of 8 items, being four items controlled for social desirability and 4 with no control. The items are presented on Results, for an English version see Appendix. The internal consistency values of the total scale was  $\alpha = .771$  ( $\omega = .793$ ), and for each subscale was  $\alpha = .662$  ( $\omega = .728$ ) for Machiavellianism,  $\alpha = .533$  ( $\omega = .558$ ) for Psychopathy, and  $\alpha = .674$  ( $\omega = .686$ ) for Narcissism.

**Experimental Dark Traits Scale – Forced-choice version.** As the Likert version, this scale was constructed by the authors following the process presented in Article 2. This scale is composed of 16 blocks with three items to be responded in MOLE (MOst and Least like me) format, where respondents are asked to choose which phrase better describes her/him in a full raking format. The items from the same block had their social desirability level balanced. Each item is related to one of the dark traits, summing up 14 items for each of the traits, Machiavellianism, Psychopathy, and Narcissism. As an

additional measure of control, in 6 blocks, one of the items was replaced with an item related to Openness, for present low correlation with the three assess traits (e.g., Geel, Goemans, Toprak, & Vedder, 2017; Kowalski, Vernon, & Schermer, 2019).

**Egoistic and Moralistic Scale (EMS; Vecchione, Alessandri, & Barbaranelli, 2013. Annex 7).** EMS is a self-report scale composed of 14 items on a 5-point Likert scale, with 1 = "strongly disagree" and 5 = "strongly agree". It assesses traits of social desirability based on egoistic bias (VE; 7 items) and moralistic bias (VM; 7 items). The Brazilian version adapted by Costa (2017), with authorization from the authors of the original, and found internal consistency equal to 0.81 for VE and 0.79 for VM, with a factorial structure equal to the original.

## **Procedures**

An Ethics Committee approved the study protocol under the CAAE 91793718.7.0000.5514. The instruments and Consent Forms were available in Survey Monkey (Online protocol), the link was distributed via social media (Facebook and E-mail). An invitation and a brief explanation of the study was available on the first page, followed by the Consent Form and instruments. To guarantee that participants understood how to answer the forced-choice scale, a control item was used, where people should put in crescent order the numbers 8, 10, and 15 that were disposed of as 8,15 and 10. Those who failed on this task were excluded from the dataset, corresponding to 56 persons (4.56%).

## **Data analysis**

**Likert scale.** We performed an Exploratory Factor Analysis (EFA) to assess the factor structure of the Experimental Dark Traits – Likert version. We run the analysis using a polychoric matrix and as estimator the Weighted Least Squares Means and

Variance (WLSMV). To decide the number of factors, we used Parallel Analysis with a random permutation of raw data, and with an oblique rotation, Promin method. As measures of goodness of the model, we considered Chi-square ( $\chi^2$ ), Root Mean Square Error Approximation (RMSEA), Comparative Fit Index (CFI), and Taylor-Lewis Index (TLI) indexes. We perform this analysis using Mplus and Factor software, the second was also considered for providing additional information that helped to decide the factor number. As additional information from Factor program, we assessed factor stability with H index, which measures how good a set of items represents a common factor (ref). Values from H index vary between 0 to 1, values greater than 0.80 suggest a stable latent variable that has a high probability of being replicated in future studies.

**Forced-Choice scale.** To verify the factor loadings of the items in each factor, we used the Thurstonian Item Response Theory Model (IRT-T; Brown & Maydeu-Olivares, 2011), recommended for Confirmatory Factor Analysis using forced-choice items. To build the syntax, we used the Excel macro available in Anna Brown's website (<http://annabrown.name/software>) that automates Mplus syntax creation (for more details, see Brown & Maydeu-Olivares, 2012). The IRT-T requires that all answers be transformed into binary variables. In this study, three binary variables were created for each block: {A, B}, {A, C}, and {B, C}, items' response was replaced by 1 when the first item was preferred over the second, and 0 otherwise. For example, if in a block (A, B, C), a respondent rank the items as A = 2, B = 3, C = 1, the corresponding dummy variables are {A, B} = 1, {A, C} = 0; and {B, C} = 0.

We fitted the model to tetrachoric correlations, defined the items for each of the three factors, and did not assume a previous correlation between factors before the analysis. The items referent to Openness factor were fixed as having a zero correlation with the traits. The first analysis considered the 16 blocks, which had had iteration problems. To identify a reduced version with a good fit for the model, we tested several



methods by excluding blocks with the highest residual variance. Then, we tested the remaining blocks and the excluded blocks separately. To verify the fitness of the model, we considered goodness-of-fit indexes Chi-square ( $\chi^2$ ), Root Mean Square Error Approximation (RMSEA), Comparative Fit Index (CFI), and Taylor-Lewis Index (TLI).

**Latent factor intercorrelation.** When the best fitting model for each response format was identified, we correlated the Likert, forced-choice response formats and social desirability scale.

## Results

### Likert scale

The adequacy of the correlation matrix was tested considering the Bartlett's statistic (3762.5,  $gl = 153$ ,  $p < 0,001$ ) and KMO (.798; CI .786 - .820), that suggested a fair interpretability of the matrix. The advised number of factor when 95 percentiles considered was 3 factors. The fit for the hypothesized models are for the 3-factor model goodness-of-fit indexes were  $\chi^2(102) = 848.076$  ( $p < .001$ ), RMSEA = .092 (95% CI: .086 - .098), TLI = .811 (95% CI: .883 - .921), and CFI = .874 (95% CI: .912 - .945) suggested a poor structure. The factor loadings are shown in Table 1. Based on H index, only Factor 1 and 3, that has more items representative of Narcissism Grandiose narcissism and Vulnerable narcissism, respectively, demonstrated stable latent variables.

Table 1

*Factor loadings of Experimental Dark Triad Traits - Liker version with a model of 3 and 2 factors*

Items	3-factor model			2-factor model	
	F1	F2	F3	F1	F2
1.Em algumas situações, alterar fatos é uma estratégia válida.	0.105	<b>0.553</b>	0.066	<b>0.559</b>	0.082
2.Evito contar mentiras mesmo que eu tenha prejuízo.	0.001	<b>-0.581</b>	0.052	<b>-0.590</b>	0.082
3.Sempre tenho as melhores ideias.	<b>0.822</b>	0.018	0.008	-0.069	<b>0.458</b>
4.Tenho um talento para influenciar as pessoas.	<b>0.615</b>	0.098	-0.036	-0.004	<b>0.358</b>
5.Gosto de seguir as regras de forma rígida.	0.159	<b>-0.591</b>	0.255	<b>-0.598</b>	<b>0.343</b>
6.Meu medo é não ter o reconhecimento que mereço.	0.142	-0.064	<b>0.640</b>	-0.027	<b>0.645</b>
7.Sei que serei alguém bem-sucedido.	<b>0.565</b>	-0.036	0.010	-0.102	<b>0.343</b>
8.Com boas mentiras posso alcançar qualquer objetivo.	0.093	<b>0.826</b>	-0.010	<b>0.830</b>	-0.005
9.Não me sinto bem seguindo regras.	-0.008	<b>0.703</b>	-0.175	<b>0.697</b>	-0.191
10.Para mim é difícil lidar com rejeição.	-0.134	0.021	<b>0.619</b>	0.092	<b>0.464</b>
11.Acredito que para vencer na vida preciso enganar as pessoas.	-0.003	<b>0.808</b>	0.092	<b>0.830</b>	0.032
12.Gosto de pessoas que reconhecem o quanto sou especial.	0.145	0.000	<b>0.683</b>	0.040	<b>0.684</b>
13.Agrada-me colocar duas pessoas uma contra a outra.	0.044	<b>0.491</b>	0.263	<b>0.514</b>	0.228
14.Tomo decisões de última hora sem pensar nas consequências.	-0.129	<b>0.332</b>	0.228	<b>0.362</b>	0.117
15.Os outros não conseguem entender o meu valor.	-0.011	0.210	<b>0.554</b>	0.252	<b>0.473</b>
16.Ameaço e intimido as pessoas para garantir meu poder sobre elas.	0.230	<b>0.419</b>	0.274	<b>0.422</b>	<b>0.344</b>
17.Se alguém tenta me impor alguma regra eu ignoro.	0.050	<b>0.480</b>	0.060	<b>0.486</b>	0.054
18.Irrito-me quando alguém me critica.	-0.044	0.190	<b>0.612</b>	0.243	<b>0.502</b>
H-latent	0.836	0.759	0.904	0.902	0.828
H-observed	0.820	0.714	0.802	0.804	0.826

Note. Factor-loadings greater than 0.30 are in **bold**.

The 2-factor model goodness-of-fit indexes were  $\chi^2(118) = 1701.830$  ( $p < .001$ ), RMSEA = .124 (95% CI: .119 - .130), TLI = .653 (95% CI: .811 - .877), and CFI = .732 (95% CI: .843 - .898) suggested a poor structure. The factor loadings are shown in Table 1. Based on H index, the two factors demonstrated stable latent variables.

### Forced-choice scale

The items that composed the 16 blocks were selected based on their similarity level in social desirability (see Article 1) and distributed in a way that each block had one item of each trait, Machiavellianism, Psychopathy, and Narcissism. In six blocks, one of the items was replaced by an item related to Openness, which has a low relationship with the assess traits (e.g., Geel et al., 2017; Kowalski et al., 2019), to be used as a bug item.

The investigation was performed in several steps, until finding a set of items with a good fit index. In the first step, the 16 blocks were investigated but were not possible to follow because there was no convergence, the number of iterations exceeded. In the second step, we separated the blocks with and without bug items before the analysis.

Considering that the scale was constructed to assess Dark Triad traits, we first test a 3-factor solution with the ten blocks without bug items, and three blocks were removed because they had the highest residual variances, between 28.573 and 227.891, while the variances of the other blocks varied between .227 and 5.418. This step was repeated until all blocks had similar residual variance, which led to the exclusion of four blocks. The same steps were repeated, adding the blocks with bug items, which excluded four more blocks.

In sequence, we investigate the remaining blocks and the excluded blocks to verify if selection were by chance or not. The eight excluded blocks had a no convergence, exceeding the number of iterations. The eight blocks of the final version showed reasonable values of goodness-of-fit, being  $\chi^2(236) = 260.470$  ( $p = .1130$ ), RMSEA = .013 (90% C.I. .000 – .022) CFI = 0.981, and TLI = 0.978, the correlation between the factor were  $r_{MxP} = .967$  ( $p=.000$ ),  $r_{MxN} = .308$  ( $p=.077$ ), and  $r_{PxN} = .621$  ( $p=.000$ ). The factor loadings of each item are presented in Table X.

We observed a correlation between Machiavellianism and Psychopathy near to 1. Due to this result we decided to test a 2-factor solution, which showed a good fit to the model,  $\chi^2(236) = 294.807$  ( $p = .0055$ ), RMSEA = .02 (90% C.I. .011 – .027), CFI = .958, TLI = .951, and correlation between factor of  $r = -.328$  ( $p=.000$ ). The factor loadings of each item are presented in Table 2.

Table 2

*Confirmatory Factor Analysis with 2- and 3-factor models of selected blocks*

Blocks	U	Items	CFA – 3 factors			CFA – 2 factors	
			M	P	N	M/P	N
1	t1	Finjo estar preocupado com as pessoas para conseguir algo.	.741			.796	
	t2	Detesto seguir ordens.		.516		.564	
	t3	Sou a melhor pessoa para ditar como as coisas devem ser.			-.071		.125
2	t4	Altero os fatos de propósito.	.801			.786	
	t5	Para mim os bens materiais são mais importantes do que as pessoas.		.224		.182	
	t6	As pessoas deveriam agradecer por ter minha companhia.			.230		-.122
5	t13	Arranjo inimigos com mais facilidade do que amigos.		.397		.362	
	t14	Uso os outros para meus próprios objetivos.	-.024			-.046	
	t15	Irrito-me quando alguém discorda de mim.			-.745		.447
7	t19	Mereço tratamento especial dos outros.			-.104		.164
	t20	Tenho dificuldade para acatar ordens de chefes, superiores e outras pessoas de autoridade.		.257		.221	
	t21	Sou bom em usar os pontos fracos das pessoas para ganho próprio.	.521			.513	
8	t22	Nem sempre estou satisfeito comigo mesmo.			-.488		.400
	t23	Prefiro viver o agora do que planejar algo.		.171		.223	
	t24	Decoro pequenas informações para parecer mais inteligente.	.600			.736	
9	t25	Costumo mentir quando é oportuno.	.955			.810	
	t26	Se alguém tenta me impor alguma regra eu ignoro.		.776		.897	
	t27	Sou tão brilhante que deveria ser pago só por existir.			.842		-.997
14	t40	Tenho pouco interesse em ter ideias originais.					
	t41	Admiro pessoas mestres em manipulação.	.058			.068	
	t42	Tomo decisões de última hora sem pensar nas consequências.		.240		.282	
16	t46	Gosto de criar coisas novas mesmo que inúteis.					
	t47	Sei que serei alguém bem-sucedido.			.311		-.239
	t48	Faço o que quero sem pensar nas consequências.		.381		.328	

Note. U = utility; CFA = Confirmatory Factor Analysis.

**Evidence based on relations to other variables**

Table 3 shows the correlation between the assessed scales. Based on the results, it is noticeable that the forced-choice version had a higher correlation with Short Dark Triad (SD3) than the Likert version. In both versions, the Machiavellianism subscale had a higher and negative correlation with the Moralistic aspect of social desirability. Narcissism subscales had a higher correlation with the Egoistic aspect of social desirability, but the forced-choice version showed a negative relationship, and the Likert version a positive relationship.

Table 3.

*Correlation between subscales of Forced choice and Likert scales of Dark Triad, Short Dark Triad Scale and Egoistic and Moralistic Scale*

	1	2	3	4	5	6	7	8
1.FC-MP	-							
2.FC-N	.181	-						
3.L-MP	.389***	.085	-					
4.L-N	.035	-.023	.492***	-				
5. M-SD3	.390***	.158	.706***	.617***	-			
6. N-SD3	.025	.474***	.440***	.843***	.440***	-		
7. P-SD3	.344***	.131	.775***	.532***	.842***	.520***	-	
8. ES-EMS	-.153*	-.508***	-.079	.260***	.129**	.483***	.059	-
9. EM-EMS	-.207**	-.122	-.562***	-.089**	-.348***	-.121**	-.402***	.342***

*Note.* FC-MP = Forced-choice version, subscale Machiavellianism and Psychopathy; FC-N = Forced-choice, subscale Narcissism; L-MP = Likert version, subscale Machiavellianism and Psychopathy; L-N = Likert version, subscale Narcissism; M-SD3 = Short Dark Triad, subscale Machiavellianism; N-SD3 = Short Dark Triad, subscale Narcissism; P-SD3 = Short Dark Triad, subscale Psychopathy; ES = Egoistic Subscale; EM = Moralistic Subscale.

A multivariate model was tested with the three subscales of Short Dark Triad and the two subscales of Egoistic and Moralistic as independent variables and the factors extracted from forced choice and Likert versions as dependent variables, see Figure 1. The goodness of fit measures suggested a poor fit to the model, being  $\chi^2(3269) = 6414.544$ , RMSEA = .032 (90% CI: .031 - .033), CFI = .766, TLI = .756. Modification indices suggested residual variance between variables. For the factor from the forced-choice version, the F1 (FC-MP) was predicted by Machiavellianism and Narcissism from SD3 and regarding EMS, only Egoistic subscale predicted F2 (FC-N). Concerning the factor from the Likert version, they were predictor for almost all variables, except that Psychopathy had a non-significative relation with F2 (L-N).

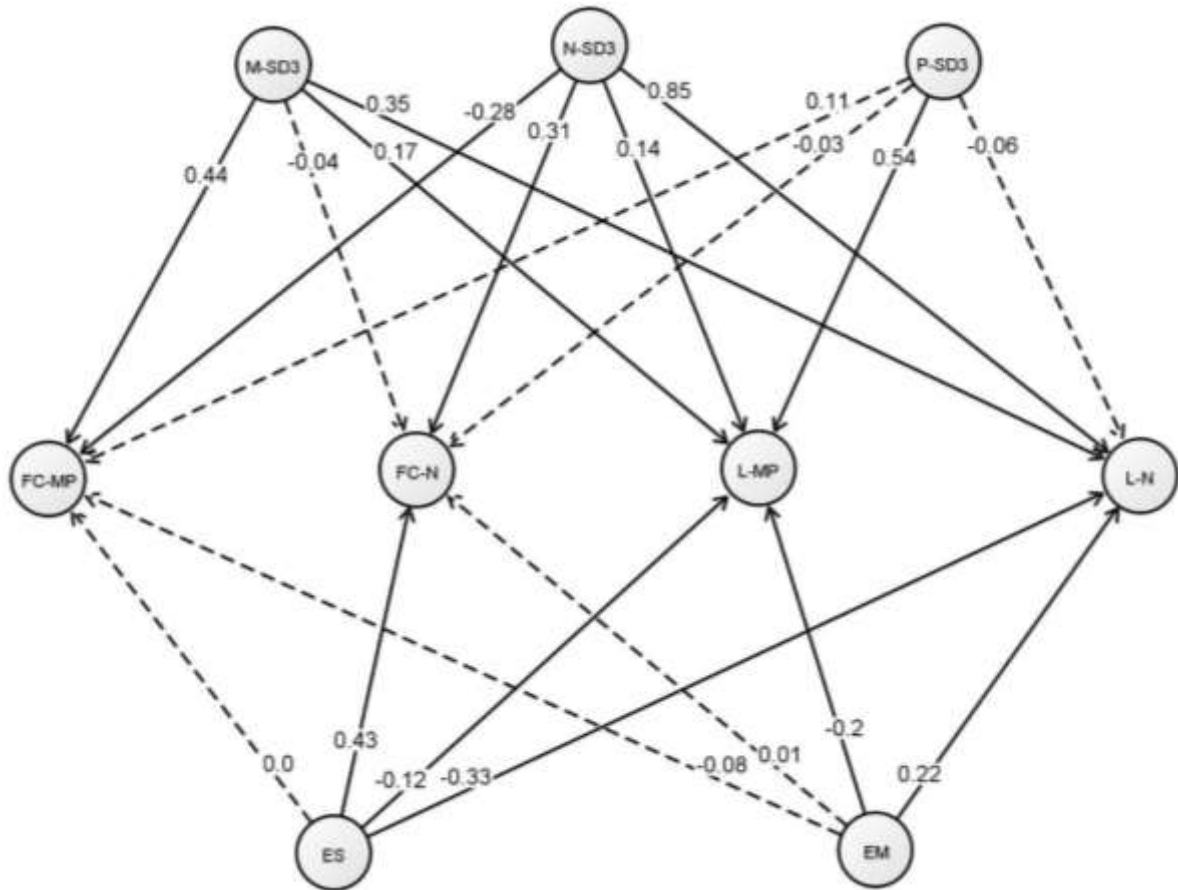


Figure 1. A predictive model of Short Dark Triad (SD3) and Egoistic and Moralistic Scale (EMS) on latent dark traits of forced-choice and Likert version. For parsimony, correlations measures, error measures, and observable variables were suppressed. Dashed lines represent a non-significant relationship. Note. FC-MP = Forced-choice version, subscale Machiavellianism and Psychopathy; FC-N = Forced-choice, subscale Narcissism; L-MP = Likert version, subscale Machiavellianism and Psychopathy; L-N = Likert version, subscale Narcissism; M-SD3 = Short Dark Triad, subscale Machiavellianism; N-SD3 = Short Dark Triad, subscale Narcissism; P-SD3 = Short Dark Triad, subscale Psychopathy; ES = Egoistic Subscale; EM = Moralistic Subscale.

## Discussion

Psychologists interested in personality inventories are attentive in investigations of tools that might help reduce test vulnerability to response distortion, such as social desirability. However, assessment of troublesome traits, like DT traits, still imposes a challenge for researchers. This article proposed two new measures to assess DT traits with a set neutralized items and a version of forced-choice response format. The models emerged from participants of the general community, with a high level of education, and the forced-choice method

involved Thurstonian modeling. The aim was to examine the psychometric quality of the scales concerning the internal structure and latent intercorrelation measures.

Both formats were unable to recover a three-factor structure hypothesized by DT literature. They suggested a two-factor model with a better theoretical interpretation. In both scenarios, items referring to Machiavellianism and Psychopathy collapsed in the same factor, and items relating to Narcissism composed the second factor. In line with this finding, results from different research brought into question the “real” structure of dark traits, especially what concerns Machiavellianism and Psychopathy, that some author describes them as virtually identical (e.g., Jones & Paulhus, 2009). In a study that investigated the structure of the SD3, Persson, Kajonius, and Garcia (2017), using exploratory and confirmatory analysis conclude that Machiavellianism and Psychopathy are “very similar constructs that are possible to subsume under one factor” (p.12). Hence, to be able to recover a three-factor structure is necessary a clear differentiation between Machiavellianism and Psychopathy that goes beyond the idea of Machiavellianism as “psychopathy without impulsivity” (Persson et al., 2017)

The factor measuring similar attributes and different response formats had a low correlation. The FC-MP had a significant correlation with L-MP, but the Narcissism factors from both versions had a non-significant correlation. One possible reason for this occurrence is that the versions have different items, and on the FC format, each block was controlled by their level of social desirability. Also, the FC format simulates a real-life decision process, where the individual is presented with a different choice, not always they are all good or all bad, but they demand a decision (Müller, Welter, & Capitão, 2007). While the Likert format has two disadvantages aspects, first is typical on researchers and job selection, so it is strongly associated with social judgment, even if under a scientific perspective, making them more likely to be manipulated. Also, it allows each item to be analyzed individually, which facilitates that the examinee becomes involved in a self-protective behavior, then hiding her/his right level

on the assessed trait. Thus a response distortion can occur consciously or unconsciously, in either case, the actual level is masked what influences in the correlation with other measures (Watts et al., 2016). An additional aspect is that instruments with heterogeneous types of response format, tend to present lower correlation than instruments with homogenous types of response format, this occurs because of the effect of the common method variance, that states that if two measures use the same method, the magnitude of relationships (e.g., correlations) among variables can inflate (Spector, Rosen, Richardson, Williams, & Johnson, 2019).

The relation with the constructed instruments of DT traits and a common measure, the SD3, had a different pattern of correlation with each version. The forced-choice version presented a clear pattern of correlation with the SD3 factors, being that FC-MP had a significant relationship with Machiavellianism and Psychopathy factors from SD3, and FC-N had a significant relationship with Narcissism factor of SD3, being better discriminate when considering multivariate analysis. The Likert version presented a mixed correlation with all factors of SD3. Based on this, two aspects might be considered, those mentioned above, common variance effect and the debate about the ‘real’ structure of the triad, and the successive suggestion of a dark dyad. Of course, these two aspects are not entirely responsible for the low correlations; the content of the items of the final Likert version is more mixed than those selected for the final forced-choice version, suggesting that each scale is measuring different aspects of the traits.

Among the factors of forced choice, only Narcissism had a positive and moderate relationship with egoistic scale, indicating that this type of format is less susceptible to the bias of social desirability. A probable cause of the relation of egoistic scale to Narcissism can be explained by the fact that measures of social desirability are not pure measures of bias, but that they also assess substance, that is, it may occur that people actually behave like reported on these scales. Thus, this positive and moderate relationship can occur due to a common content



variance, since the content of the items on the egoistic scale are like items used to assess Narcissism (Paulhus, 2002). On the other hand, a pattern of significant relationships between factors of social desirability and factors of the Likert scale indicate that these scales share both a common method variance and substance. Thus, even in the face of the obvious weaknesses of instruments of social desirability, the forced choice response format is less influenced by the bias, capturing in a prevalent way the common variance of personality trait.

Considering the results, some limitations might be highlighted. First, the high correlation between the dark traits imposed a natural challenge on the moment of elaborate the items. Thus, we suggest the creation of a review version, combining items that assess more discriminative content of these traits with items referring to different attributes that have correlation near to zero with the trait of interest. Also, it would be informative to apply the items from the forced-choice format as single-stimulus to verify if this pattern of correlation persists. Nonetheless, this does not invalidate the use of both scales in research context and call for further investigation of the use of neutralized items and forced-choice format to assess Dark Triad traits. In general, this study shows that forced-choice scales can be used to measures dark traits, but they require attention from the researcher.

## References

- Ackerman, R. A., Donnellan, M. B., Roberts, B. W., & Fraley, R. C. (2016). The effect of response format on the psychometric Properties of the Narcissistic Personality Inventory: Consequences for item meaning and factor structure. *Assessment, 23*(2), 203–220. <https://doi.org/10.1177/1073191114568113>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and*

- psychological testing*. Washington, DC, US: American Psychological Association.
- Anglim, J., Bozic, S., Little, J., & Lievens, F. (2018). Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings. *Advances in Health Sciences Education, 23*(2), 311–321. <https://doi.org/10.1007/s10459-017-9796-8>
- Artino, A. R. J., Gehlbach, H., & Durning, S. J. (2011). AM Last Page: Avoiding Five Common Pitfalls of Survey Design. *Academic Medicine, 86*(10), 1327. <https://doi.org/10.1097/ACM.0b013e31822f77cc>
- Bäckström, M., & Björklund, F. (2013). Social desirability in personality inventories: symptoms, diagnosis and prescribed cure. *Scandinavian Journal of Psychology, 54*(2), 152–159. <https://doi.org/10.1111/sjop.12015>
- Bäckström, M., & Björklund, F. (2014). *Social Desirability in Personality Inventories The Nature of the Evaluative Factor*. <https://doi.org/10.1027/1614-0001/a000138>
- Bäckström, M., Björklund, F., & Larsson, M. R. (2012). Social desirability in personality assessment. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 201–213). New York, NY: Oxford University Press, Inc.
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation, 19*(5). <https://doi.org/https://doi.org/10.1146/annurev.psych.53.100901.135239>.
- Baka, A., & Figgou, L. (2012). “Neither agree, nor disagree”: A critical analysis of the middle answer category in Voting Advice Applications. *International Journal of Electronic Governance, 5*(3/4), 244–263. <https://doi.org/10.1504/IJEG.2012.051306>
- Bandalos, D. L. (2018a). The test development process. In D. L. Bandalos (Ed.), *Measurement Theory and Applications for the Social Sciences* (pp. 41–84). New York, NY: The Guilford Press.

- Bandalos, D. L. (2018b). Writing noncognitive items. In D. L. Bandalos (Ed.), *Measurement Theory and Applications for the Social Sciences* (pp. 85–119). New York, NY: The Guilford Press.
- Baughman, H. M., Jonason, P. K., Lyons, M., & Vernon, P. A. (2014). Liar liar pants on fire: Cheater strategies linked to the Dark Triad. *Personality and Individual Differences*, *71*(71), 35–38. <https://doi.org/10.1016/j.paid.2014.07.019>
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*(4), 317–335. <https://doi.org/10.1111/j.1468-2389.2006.00354.x>
- Blasberg, S. A., Rogers, K. H., & Paulhus, D. L. (2014). The bidimensional impression management index (BIMI): Measuring agentic and communal forms of impression management. *Journal of Personality Assessment*, *96*(5), 523–531. <https://doi.org/10.1080/00223891.2013.862252>
- Brown, A., & Maydeu-olivares, A. (2016). Modeling forced-choice response formats. In P. Irwing, T. Booth, & D. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing* (pp. 1–64). London: John Wiley & Sons, Inc.
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, *71*(3), 460–502. <https://doi.org/10.1177/0013164410375112>
- Brown, Anna. (2016). Item response models for forced-choice questionnaires: A common framework. *Psychometrika*, Vol. 81, pp. 135–160. <https://doi.org/10.1007/s11336-014-9434-9>
- Brown, Anna, & Maydeu-Olivares, A. (2012). Fitting a thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods*, *44*(4), 1135–1147.

<https://doi.org/10.3758/s13428-012-0217-x>

- Brown, Anna, & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, Vol. 18, pp. 36–52. <https://doi.org/10.1037/a0030641>
- Cabooter, E., Weijters, B., Geuens, M., & Vermeir, I. (2016). Scale format effects on response option interpretation and use. *Journal of Business Research*, 69(7), 2574–2584. <https://doi.org/10.1016/j.jbusres.2015.10.138>
- Cattell, R. B. (1958). What is “objective” in “objective personality tests?” *Journal of Counseling Psychology*, 5(4), 285–289. <https://doi.org/10.1037/h0046268>
- Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309–336). New York, NY: Routledge/Taylor & Francis Group.
- Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, 30(1), 99–121. <https://doi.org/10.2333/bhmk.30.99>
- Christiansen, N. D., Burns, G. N., & George, E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, 18(3), 267–307. [https://doi.org/10.1207/s15327043hup1803\\_4](https://doi.org/10.1207/s15327043hup1803_4)
- Christie, R., & Geis, F. L. (1970). *Studies in machiavellianism*. <https://doi.org/10.1016/B978-0-12-174450-2.50008-7>
- Collison, K. L., Vize, C. E., Miller, J. D., & Lynam, D. R. (2018). Development and preliminary validation of a Five Factor Model measure of Machiavellianism. *Psychological Assessment*, 30(10).
- Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal Of Consulting Psychology*, 24(4), 349–354.

<https://doi.org/10.1037/h0047358>

- Cunha, C. M., Almeida Neto, O. P., & Stackfleth, R. (2016). Principais métodos de avaliação psicométrica da confiabilidade de instrumentos de medida. *Revista de Atenção à Saúde*, *14*(49), 98–103. <https://doi.org/10.13037/rbcs.vol14n49.3671>
- Dolnicar, S., & Grün, B. (2013). “Translating” between survey answer formats. *Journal of Business Research*, *66*(9), 1298–1306. <https://doi.org/10.1016/j.jbusres.2012.02.029>
- Donovan, J., & Dwright, S. A. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology*, *29*(September 2013), 479–493. <https://doi.org/10.1007/s10869-013-9318-5>
- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology*, *21*(3), 465–476. <https://doi.org/http://dx.doi.org/10.5380/psi.v21i3.53054>
- Dubas, J. S., Baams, L., Doornwaard, S. M., & Aken, M. A. G. Van. (2017). Dark Personality Traits and Impulsivity Among Adolescents: Differential Links to Problem Behaviors and Family Relations. *Journal of Abnormal Child Psychology*, *126*(7), 877–889.
- Ellingson, J. E. (2012). People fake only when they need to fake. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 19–33). New York, NY: Oxford University Press, Inc.
- Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social and Personality Psychology Compass*, Vol. 7, pp. 199–216. <https://doi.org/10.1111/spc3.12018>
- Geel, M. Van, Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences*, *106*, 231–235. <https://doi.org/10.1016/j.paid.2016.10.063>

- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96.
- Hauenstein, N. M. A., Bradley, K. M., O’Shea, P. G., Shah, Y. J., & Magill, D. P. (2017). Interactions between motivation to fake and personality item characteristics: Clarifying the process. *Organizational Behavior and Human Decision Processes, 138*, 74–92. <https://doi.org/10.1016/j.obhdp.2016.11.002>
- He, J., van de Vijver, F. J. R., Dominguez Espinosa, A., Abubakar, A., Dimitrova, R., Adams, B. G., ... Villieux, A. (2015). Socially Desirable Responding. *Cross-Cultural Research, 49*(3), 227–249. <https://doi.org/10.1177/1069397114552781>
- Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, Vol. 91*, pp. 9–24. <https://doi.org/10.1037/0021-9010.91.1.9>
- Hervé, H. (2017). Psychopathy across the ages: A history of the Hare psychopath. In H. Hugues & J. C. Yuille (Eds.), *The Psychopath: Theory, Research, and Practice* (pp. 31–55). Routledge.
- Hogan, R. (1991). Personality and personality measurement. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 873–919). Palo Alto, CA: Consulting Psychologists Press.
- Hontangas, P. M., Leenen, I., de la Torre, J., Ponsoda, V., Morillo, D., & Abad, F. J. (2016). Traditional scores versus IRT estimates on forced-choice tests based on a dominance model. *Psicothema, Vol. 28*, pp. 76–82. Hontangas, Pedro M.: Facultad de Psicología, Universidad de Valencia, Valencia, Spain, 46011, [hontanga@uv.es](mailto:hontanga@uv.es): Colegio Oficial de Psicólogos del Principado de Asturias.

- Hontangas, P. M., Leenen, I., Torre, J. De, Ponsoda, V., & Morillo, D. (2016). Traditional scores versus IRT estimates on forced-choice tests based. *Psicothema*, *28*(1), 76–82. <https://doi.org/10.7334/psicothema2015.204>
- Hontangas, P. M., Torre, J. De, Ponsoda, V., Leenen, I., Morillo, D., & Abad, F. J. (2015). Comparing traditional and IRT scoring of forced-choice tests. *Applied Psychological Measurement*, *39*(8), 598–612. <https://doi.org/10.1177/0146621615585851>
- Johnson, J. A. (2004). The impact of item characteristics on item and scale validity. *Multivariate Behavioral Research*, *39*(2), 273–302. <https://doi.org/10.1207/s15327906mbr3902>
- Jonason, P. K., & Fletcher, S. A. (2018). Agentic and communal behavioral biases in the Dark Triad traits. *Personality and Individual Differences*, *130*(August), 76–82. <https://doi.org/10.1016/j.paid.2018.03.044>
- Jonason, P. K., Lyons, M., Baughman, H. M., & Vernon, P. A. (2014). What a tangled web we weave: The dark triad traits and deception. *Personality and Individual Differences*, *70*, 117–119. <https://doi.org/10.1016/j.paid.2014.06.038>
- Jonason, P. K., Strosser, G. L., Kroll, C. H., Duineveld, J. J., & Baruffi, S. A. (2015). Valuing myself over others: The Dark Triad traits and moral and social values. *Personality and Individual Differences*, *81*, 102–106. <https://doi.org/10.1016/j.paid.2014.10.045>
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, Vol. 22, pp. 420–432. <https://doi.org/10.1037/a0019265>
- Jones, D. N., & Paulhus, D. (2009). Machiavellianism. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 93–108). The Guilford Press.
- Jones, Daniel N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3). *Assessment*, *21*(1), 28–41. <https://doi.org/10.1177/1073191113514105>
- Jones, Daniel N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief

- measure of dark personality traits. *Assessment*, 2(1), 28–41.  
<https://doi.org/10.1177/1073191113514105>
- Kline, P. (2015). Making test reliable II: Personality inventories. Item writing. In P. Kline (Ed.), *A handbook of test construction: Introduction to psychometric design* (pp. 24–76). Routledge/Taylor & Francis Group.
- Kowalski, C. M., Vernon, P. A., & Schermer, J. A. (2019). The Dark Triad and facets of personality. *Current Psychology*, (November). <https://doi.org/10.1007/s12144-019-00518-0>
- Krammer, G., Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment*, 99(5), 510–523.  
<https://doi.org/https://doi.org/10.10>
- Krammer, Georg, Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment*, 99(5), 510–523.  
<https://doi.org/10.1080/00223891.2017.1285781>
- Krizan, Z., & Herlache, A. D. (2017). The narcissism spectrum model: A synthetic view of narcissistic personality. *Personality and Social Psychology Review*, 1–29.  
<https://doi.org/10.1177/1088868316685018>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22  
140, 55. <https://doi.org/2731047>
- Luft, J., & Ingham, H. (1961). The johari window. *Human Relations Training News*, 5(1), 6–7.
- MacCann, C., Ziegler, M., & Roberts, R. D. (2012). Faking in personality assessment: reflections and recommendations. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.),



*New perspectives on faking in personality assessment* (pp. 309–329). New York, NY: Oxford University Press, Inc.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, *136*(3), 450–470. <https://doi.org/10.1037/a0019216>

Morey, L. C. (2012). Detection of Response Bias in Applied Assessment: Comment on McGrath et al. (2010). *Psychological Injury and Law*, *5*(3–4), 153–161. <https://doi.org/10.1007/s12207-012-9131-x>

Müller, G., Welter, R., & Capitão, C. G. (2007). Medidas ipsativas na avaliação psicológica. *Avaliação Psicológica*, *6*(2), 157–165.

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology*, *142*(2), 71–89. <https://doi.org/10.1080/00221309.2014.994590>

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*(6), 660–679. <https://doi.org/10.1037/0021-9010.81.6.660>

Parker, R. I., Vannest, K. J., & Davis, J. L. (2013). Reliability of multi-category rating scales. *Journal of School Psychology*, *51*(2), 217–229. <https://doi.org/10.1016/j.jsp.2012.12.003>

Paulhus, D. L. (2002). Socially desirable responding: the evolution of a construct. In H. I. Brown, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.

Paulhus, D. L. (2006). *Comprehensive Inventory of Desirable Responding (CIDR)*. New Orleans.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*,

66(6), 1025–1060. <https://doi.org/10.1111/1467-6494.00041>

- Paulhus, D. L., Neumann, C. S., Hare, R. D., Williams, K. M., & Hemphill, J. F. (2016). *Self-report Psychopathy Scale - 4* (4th ed.). Toronto, ON, Canada: Multi-Healthy Systems.
- Paulhus, D. L. (1984). Two-Component Models of Socially Desirable Responding. *Personality Processes and Individual Differences*, 46(3), 598–609.
- Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)
- Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences*, 35(2), 263–275. [https://doi.org/10.1016/S0191-8869\(02\)00187-3](https://doi.org/10.1016/S0191-8869(02)00187-3)
- Persson, B. N., Kajonius, P. J., & Garcia, D. (2017). Revisiting the structure of the Short Dark Triad. *Assessment*, 1–14. <https://doi.org/10.1177/1073191117701192>
- Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K. N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychological Assessment*, 21(3), 365–379. <https://doi.org/10.1037/a0016530>
- Raskin, R. N., & Hall, C. S. (1979). A Narcissistic Personality Inventory. *Psychological Reports*, 45(2), 590–590. <https://doi.org/10.2466/pr0.1979.45.2.590>
- Shoss, M. K., & Strube, M. J. (2011). How do you fake a personality test? An investigation of cognitive models of impression-managed responding. *Organizational Behavior and Human Decision Processes*, 116(1), 163–171. <https://doi.org/10.1016/j.obhdp.2011.05.003>
- Simonet, D. V., Tett, R. P., Foster, J., Angelback, A. I., & Bartlett, J. M. (2017). Dark-Side Personality Trait Interactions: Amplifying Negative Predictions of Leadership Performance. *Journal of Leadership & Organizational Studies*, (September), 1–18.

<https://doi.org/10.1177/1548051817727703>

- Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Propriedades psicométricas na avaliação de instrumentos: Avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*, 26(3), 649–659. <https://doi.org/10.5123/S1679-49742017000300022>
- Spector, P. E., Rosen, C. C., Richardson, H. A., Williams, L. J., & Johnson, R. E. (2019). A new perspective on method variance: A measure-Centric approach. *Journal of Management*, 45(3), 855–880.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions : The multidimensional pairwise-preference model. *Applied*, 29(3), 184–203. <https://doi.org/10.1177/0146621604273988>
- Sutton, G. (2019). *The Dark Triad in personnel selection: An exploration of narcissism, psychopathy, and Machiavellianism in candidates involved in an organisational selection process*. University of Canterbury.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 101(2), 266–270.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33(4), 529–554.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). An introduction and a point of view. In R. Tourangeau, L. J. Rips, & K. Rasinski (Eds.), *The psychology of survey response* (pp. 1–22). Cambridge, UK: Cambridge University Press.
- Tutz, G., & Schauberger, G. (2018). *Uncertainty as Response Style in Latent Trait Models*.
- Usami, S., Sakamoto, A., Naito, J., & Abe, Y. (2016). Developing pairwise preference-based personality test and experimental investigation of its resistance to faking effect by item response model. *International Journal of Testing*, 16(4), 288–309.

<https://doi.org/10.1080/15305058.2016.1145123>

- van Hooft, E. A. J., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology, 97*(2), 301–316. <https://doi.org/10.1037/a0025711>
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research, 25*(2), 195–217. <https://doi.org/10.1093/ijpor/eds021>
- Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance, 19*(3), 175–199. [https://doi.org/10.1207/s15327043hup1903\\_1](https://doi.org/10.1207/s15327043hup1903_1)
- Vecchione, M., Alessandri, G., & Barbaranelli, C. (2013). Measurement and application of egoistic and moralistic self-enhancements. *International Journal of Selection and Assessment, 21*(2), 170–182. <https://doi.org/10.1111/ijsa.12027>
- Viswesvaran, C., & Ones, D. S. (1999). Meta-Analyses of Fakability Estimates: Implications for Personality Measurement. *Educational and Psychological Measurement, 59*(2), 197–210. <https://doi.org/10.1177/00131649921969802>
- Vize, C. E., Lynam, D. R., Collison, K. L., & Miller, J. D. (2018). Differences among dark triad components: A meta-analytic investigation. *Personality Disorders: Theory, Research, and Treatment, 9*(2), 101.
- Wang, W., Qiu, X., Chen, C., Ro, S., & Jin, K. (2017). Item response theory models for ipsative tests with Multidimensional Pairwise Comparison Items. *Applied Psychological Measurement, 41*(8), 1–14. <https://doi.org/10.1177/0146621617703183>
- Watts, A. L., Lilienfeld, S. O., Edens, J. F., Douglas, K. S., Skeem, J. L., Verschuere, B., & LoPilato, A. C. (2016). Does response distortion statistically affect the relations between self-report psychopathy measures and external criteria? *Psychological Assessment, 28*(3),

294–306. <https://doi.org/10.1037/pas0000168>

Weijters, B., Cabooter, E., & Schillewaert, N. (2010a). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247. <https://doi.org/10.1016/j.ijresmar.2010.02.004>

Weijters, B., Cabooter, E., & Schillewaert, N. (2010b). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247. <https://doi.org/https://doi.org/10.1016/j.ijresmar.2010.02.004>

Weijters, B., Geuens, M., & Baumgartner, H. (2013). The Effect of Familiarity with the Response Category Labels on Item Response to Likert Scales. *Journal of Consumer Research*, 40(2), 368–381. <https://doi.org/10.1086/670394>

Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956–972. <https://doi.org/10.1177/0013164404268674>

Wetzel, E., & Greiff, S. (2018). The World Beyond Rating Scales. *European Journal of Psychological Assessment*, 34(1), 1–5. <https://doi.org/10.1027/1015-5759/a000469>

Willits, F. K., Theodori, G. L., & Luloff, A. E. (2016). Another look at Likert scales. *Journal of Rural Social Sciences*, 31(3), 126–139. Retrieved from [http://journalofruralsocialsciences.org/pages/Articles/JRSS\\_2016\\_31/3/JRSS\\_2016\\_31\\_3\\_126-139.pdf](http://journalofruralsocialsciences.org/pages/Articles/JRSS_2016_31/3/JRSS_2016_31_3_126-139.pdf)

Ziegler, Matthias. (2015). “F\*\*\* you, I won’t do what you told me!”—Response biases as threats to psychological assessment. *European Journal of Psychological Assessment*, 31(3), 153–158. <https://doi.org/10.1027/1015-5759/a000292>

Ziegler, Matthias, & Buehner, M. (2009). Modeling Socially Desirable Responding and Its

Effects. *Educational and Psychological Measurement*, 69(4), 548–565.

## **Considerações finais**

Ao longo da construção do projeto, pôde-se aprender muito sobre os processos envolvidos na construção de escalas com diferentes formatos de resposta, em particular os formatos tipo Likert e escolha forçada. Por meio dos resultados ficou evidente que construir escalas vai muito além de escrever frases sobre comportamentos, sentimentos e pensamentos de pessoas com determinados traços de personalidade. Esse processo envolve decisões, que apesar de aparentemente simples, afetam a forma como o respondente percebe, interpreta e responde aos itens, e isso reflete nos resultados de análises estatísticas usadas para tratar os dados.

Os resultados desta pesquisa demonstram que o estudo de traços sombrios da personalidade é inesgotável e longe de uma conclusão definitiva. Doravante, pesquisa futuras devem considerar a influência de diferentes contextos e diferentes formatos de avaliação, por interferirem na apreensão dessas características de forma acurada. Neste sentido a mensuração dos traços sombrios da personalidade por meio de instrumento de autorrelato ainda que possível e que ofereça resultados relativamente confiáveis, ainda necessitam de mais investigações, como por exemplo, sua relação com outros vieses de resposta além de desejabilidade social. Assim, como estudos futuros propõe-se comparar as escalas em diferentes contextos de avaliação, compulsório e de pesquisa para verificar a relação entre formatos de resposta e traços indesejáveis no controle de desejabilidade social.

## Referências

- Ackerman, R. A., Donnellan, M. B., Roberts, B. W., & Fraley, R. C. (2016). The effect of response format on the psychometric Properties of the Narcissistic Personality Inventory: Consequences for item meaning and factor structure. *Assessment, 23*(2), 203–220. <https://doi.org/10.1177/1073191114568113>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC, US: American Psychological Association.
- Anglim, J., Bozic, S., Little, J., & Lievens, F. (2018). Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings. *Advances in Health Sciences Education, 23*(2), 311–321. <https://doi.org/10.1007/s10459-017-9796-8>
- Artino, A. R. J., Gehlbach, H., & Durning, S. J. (2011). AM Last Page: Avoiding Five Common Pitfalls of Survey Design. *Academic Medicine, 86*(10), 1327. <https://doi.org/10.1097/ACM.0b013e31822f77cc>
- Bäckström, M., & Björklund, F. (2013). Social desirability in personality inventories: symptoms, diagnosis and prescribed cure. *Scandinavian Journal of Psychology, 54*(2), 152–159. <https://doi.org/10.1111/sjop.12015>
- Bäckström, M., & Björklund, F. (2014). *Social Desirability in Personality Inventories The Nature of the Evaluative Factor*. <https://doi.org/10.1027/1614-0001/a000138>
- Bäckström, M., Björklund, F., & Larsson, M. R. (2012). Social desirability in personality assessment. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 201–213). New York, NY: Oxford University Press, Inc.
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration



- using FACTOR. *Practical Assessment, Research & Evaluation*, 19(5).  
<https://doi.org/https://doi.org/10.1146/annurev.psych.53.100901.135239>.
- Baka, A., & Figgou, L. (2012). “Neither agree, nor disagree”: A critical analysis of the middle answer category in Voting Advice Applications. *International Journal of Electronic Governance*, 5(3/4), 244–263. <https://doi.org/10.1504/IJEG.2012.051306>
- Bandalos, D. L. (2018a). The test development process. In D. L. Bandalos (Ed.), *Measurement Theory and Applications for the Social Sciences* (pp. 41–84). New York, NY: The Guilford Press.
- Bandalos, D. L. (2018b). Writing noncognitive items. In D. L. Bandalos (Ed.), *Measurement Theory and Applications for the Social Sciences* (pp. 85–119). New York, NY: The Guilford Press.
- Baughman, H. M., Jonason, P. K., Lyons, M., & Vernon, P. A. (2014). Liar liar pants on fire: Cheater strategies linked to the Dark Triad. *Personality and Individual Differences*, 71(71), 35–38. <https://doi.org/10.1016/j.paid.2014.07.019>
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14(4), 317–335. <https://doi.org/10.1111/j.1468-2389.2006.00354.x>
- Blasberg, S. A., Rogers, K. H., & Paulhus, D. L. (2014). The bidimensional impression management index (BIMI): Measuring agentic and communal forms of impression management. *Journal of Personality Assessment*, 96(5), 523–531. <https://doi.org/10.1080/00223891.2013.862252>
- Brown, A., & Maydeu-olivares, A. (2016). Modeling forced-choice response formats. In P. Irwing, T. Booth, & D. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing* (pp. 1–64). London: John Wiley & Sons, Inc.

- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*(3), 460–502. <https://doi.org/10.1177/0013164410375112>
- Brown, Anna. (2016). Item response models for forced-choice questionnaires: A common framework. *Psychometrika, Vol. 81*, pp. 135–160. <https://doi.org/10.1007/s11336-014-9434-9>
- Brown, Anna, & Maydeu-Olivares, A. (2012). Fitting a thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods, 44*(4), 1135–1147. <https://doi.org/10.3758/s13428-012-0217-x>
- Brown, Anna, & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods, Vol. 18*, pp. 36–52. <https://doi.org/10.1037/a0030641>
- Cabooter, E., Weijters, B., Geuens, M., & Vermeir, I. (2016). Scale format effects on response option interpretation and use. *Journal of Business Research, 69*(7), 2574–2584. <https://doi.org/10.1016/j.jbusres.2015.10.138>
- Cattell, R. B. (1958). What is “objective” in “objective personality tests?” *Journal of Counseling Psychology, 5*(4), 285–289. <https://doi.org/10.1037/h0046268>
- Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309–336). New York, NY: Routledge/Taylor & Francis Group.
- Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika, 30*(1), 99–121. <https://doi.org/10.2333/bhmk.30.99>
- Christiansen, N. D., Burns, G. N., & George, E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance, 18*(3), 267–307.

[https://doi.org/10.1207/s15327043hup1803\\_4](https://doi.org/10.1207/s15327043hup1803_4)

- Christie, R., & Geis, F. L. (1970). *Studies in machiavellianism*. <https://doi.org/10.1016/B978-0-12-174450-2.50008-7>
- Collison, K. L., Vize, C. E., Miller, J. D., & Lynam, D. R. (2018). Development and preliminary validation of a Five Factor Model measure of Machiavellianism. *Psychological Assessment, 30*(10).
- Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal Of Consulting Psychology, 24*(4), 349–354. <https://doi.org/10.1037/h0047358>
- Cunha, C. M., Almeida Neto, O. P., & Stackfleth, R. (2016). Principais métodos de avaliação psicométrica da confiabilidade de instrumentos de medida. *Revista de Atenção à Saúde, 14*(49), 98–103. <https://doi.org/10.13037/rbcs.vol14n49.3671>
- Dolnicar, S., & Grün, B. (2013). “Translating” between survey answer formats. *Journal of Business Research, 66*(9), 1298–1306. <https://doi.org/10.1016/j.jbusres.2012.02.029>
- Donovan, J., & Dwright, S. A. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal Of Business and Psychology, 29*(September 2013), 479–493. <https://doi.org/10.1007/s10869-013-9318-5>
- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology, 21*(3), 465–476. <https://doi.org/http://dx.doi.org/10.5380/psi.v21i3.53054>
- Dubas, J. S., Baams, L., Doornwaard, S. M., & Aken, M. A. G. Van. (2017). Dark Personality Traits and Impulsivity Among Adolescents: Differential Links to Problem Behaviors and Family Relations. *Journal of Abnormal Child Psychology, 126*(7), 877–889.
- Ellingson, J. E. (2012). People fake only when they need to fake. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 19–

33). New York, NY: Oxford University Press, Inc.

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social and Personality Psychology Compass*, Vol. 7, pp. 199–216.

<https://doi.org/10.1111/spc3.12018>

Geel, M. Van, Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences*, 106, 231–235.

<https://doi.org/10.1016/j.paid.2016.10.063>

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96.

Hauenstein, N. M. A., Bradley, K. M., O’Shea, P. G., Shah, Y. J., & Magill, D. P. (2017). Interactions between motivation to fake and personality item characteristics: Clarifying the process. *Organizational Behavior and Human Decision Processes*, 138, 74–92.

<https://doi.org/10.1016/j.obhdp.2016.11.002>

He, J., van de Vijver, F. J. R., Dominguez Espinosa, A., Abubakar, A., Dimitrova, R., Adams, B. G., ... Villieux, A. (2015). Socially Desirable Responding. *Cross-Cultural Research*, 49(3), 227–249. <https://doi.org/10.1177/1069397114552781>

Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, Vol. 91, pp. 9–24.

<https://doi.org/10.1037/0021-9010.91.1.9>

Hervé, H. (2017). Psychopathy across the ages: A history of the Hare psychopath. In H. Hugues & J. C. Yuille (Eds.), *The Psychopath: Theory, Research, and Practice* (pp. 31–55). Routledge.

- Hogan, R. (1991). Personality and personality measurement. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 873–919). Palo Alto, CA: Consulting Psychologists Press.
- Hontangas, P. M., Leenen, I., de la Torre, J., Ponsoda, V., Morillo, D., & Abad, F. J. (2016). Traditional scores versus IRT estimates on forced-choice tests based on a dominance model. *Psicothema*, Vol. 28, pp. 76–82. Hontangas, Pedro M.: Facultad de Psicología, Universidad de Valencia, Valencia, Spain, 46011, [hontanga@uv.es](mailto:hontanga@uv.es): Colegio Oficial de Psicólogos del Principado de Asturias.
- Hontangas, P. M., Leenen, I., Torre, J. De, Ponsoda, V., & Morillo, D. (2016). Traditional scores versus IRT estimates on forced-choice tests based. *Psicothema*, 28(1), 76–82. <https://doi.org/10.7334/psicothema2015.204>
- Hontangas, P. M., Torre, J. De, Ponsoda, V., Leenen, I., Morillo, D., & Abad, F. J. (2015). Comparing traditional and IRT scoring of forced-choice tests. *Applied Psychological Measurement*, 39(8), 598–612. <https://doi.org/10.1177/0146621615585851>
- Johnson, J. A. (2004). The impact of item characteristics on item and scale validity. *Multivariate Behavioral Research*, 39(2), 273–302. <https://doi.org/10.1207/s15327906mbr3902>
- Jonason, P. K., & Fletcher, S. A. (2018). Agentic and communal behavioral biases in the Dark Triad traits. *Personality and Individual Differences*, 130(August), 76–82. <https://doi.org/10.1016/j.paid.2018.03.044>
- Jonason, P. K., Lyons, M., Baughman, H. M., & Vernon, P. A. (2014). What a tangled web we weave: The dark triad traits and deception. *Personality and Individual Differences*, 70, 117–119. <https://doi.org/10.1016/j.paid.2014.06.038>
- Jonason, P. K., Strosser, G. L., Kroll, C. H., Duineveld, J. J., & Baruffi, S. A. (2015). Valuing myself over others: The Dark Triad traits and moral and social values. *Personality and*

- Individual Differences*, 81, 102–106. <https://doi.org/10.1016/j.paid.2014.10.045>
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, Vol. 22, pp. 420–432. <https://doi.org/10.1037/a0019265>
- Jones, D. N., & Paulhus, D. (2009). Machiavellianism. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 93–108). The Guilford Press.
- Jones, Daniel N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3). *Assessment*, 21(1), 28–41. <https://doi.org/10.1177/1073191113514105>
- Jones, Daniel N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, 2(1), 28–41. <https://doi.org/10.1177/1073191113514105>
- Kline, P. (2015). Making test reliable II: Personality inventories. Item writing. In P. Kline (Ed.), *A handbook of test construction: Introduction to psychometric design* (pp. 24–76). Routledge/Taylor & Francis Group.
- Kowalski, C. M., Vernon, P. A., & Schermer, J. A. (2019). The Dark Triad and facets of personality. *Current Psychology*, (November). <https://doi.org/10.1007/s12144-019-00518-0>
- Krammer, G., Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment*, 99(5), 510–523. <https://doi.org/https://doi.org/10.10>
- Krammer, Georg, Sommer, M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment*, 99(5), 510–523. <https://doi.org/10.1080/00223891.2017.1285781>
- Krizan, Z., & Herlache, A. D. (2017). The narcissism spectrum model: A synthetic view of

- narcissistic personality. *Personality and Social Psychology Review*, 1–29.  
<https://doi.org/10.1177/1088868316685018>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22  
140, 55. <https://doi.org/2731047>
- Luft, J., & Ingham, H. (1961). The johari window. *Human Relations Training News*, 5(1), 6–  
7.
- MacCann, C., Ziegler, M., & Roberts, R. D. (2012). Faking in personality assessment:  
reflections and recommendations. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.),  
*New perspectives on faking in personality assessment* (pp. 309–329). New York, NY:  
Oxford University Press, Inc.
- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as  
a source of error variance in applied assessment. *Psychological Bulletin*, 136(3), 450–470.  
<https://doi.org/10.1037/a0019216>
- Morey, L. C. (2012). Detection of Response Bias in Applied Assessment: Comment on  
McGrath et al. (2010). *Psychological Injury and Law*, 5(3–4), 153–161.  
<https://doi.org/10.1007/s12207-012-9131-x>
- Müller, G., Welter, R., & Capitão, C. G. (2007). Medidas ipsativas na avaliação psicológica.  
*Avaliação Psicológica*, 6(2), 157–165.
- Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and  
interpretation of mid-points in items on questionnaires. *The Journal of General  
Psychology*, 142(2), 71–89. <https://doi.org/10.1080/00221309.2014.994590>
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality  
testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81(6),  
660–679. <https://doi.org/10.1037/0021-9010.81.6.660>
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2013). Reliability of multi-category rating scales.

- Journal of School Psychology*, 51(2), 217–229. <https://doi.org/10.1016/j.jsp.2012.12.003>
- Paulhus, D. L. (2002). Socially desirable responding: the evolution of a construct. In H. I. Brown, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.
- Paulhus, D. L. (2006). *Comprehensive Inventory of Desirable Responding (CIDR)*. New Orleans.
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, 66(6), 1025–1060. <https://doi.org/10.1111/1467-6494.00041>
- Paulhus, D. L., Neumann, C. S., Hare, R. D., Williams, K. M., & Hemphill, J. F. (2016). *Self-report Psychopathy Scale - 4* (4th ed.). Toronto, ON, Canada: Multi-Healthy Systems.
- Paulhus, D. L. (1984). Two-Component Models of Socially Desirable Responding. *Personality Processes and Individual Differences*, 46(3), 598–609.
- Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)
- Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences*, 35(2), 263–275. [https://doi.org/10.1016/S0191-8869\(02\)00187-3](https://doi.org/10.1016/S0191-8869(02)00187-3)
- Persson, B. N., Kajonius, P. J., & Garcia, D. (2017). Revisiting the structure of the Short Dark Triad. *Assessment*, 1–14. <https://doi.org/10.1177/1073191117701192>
- Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K. N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychological Assessment*, 21(3), 365–379. <https://doi.org/10.1037/a0016530>
- Raskin, R. N., & Hall, C. S. (1979). A Narcissistic Personality Inventory. *Psychological*



*Reports*, 45(2), 590–590. <https://doi.org/10.2466/pr0.1979.45.2.590>

Shoss, M. K., & Strube, M. J. (2011). How do you fake a personality test? An investigation of cognitive models of impression-managed responding. *Organizational Behavior and Human Decision Processes*, 116(1), 163–171. <https://doi.org/10.1016/j.obhdp.2011.05.003>

Simonet, D. V., Tett, R. P., Foster, J., Angelback, A. I., & Bartlett, J. M. (2017). Dark-Side Personality Trait Interactions: Amplifying Negative Predictions of Leadership Performance. *Journal of Leadership & Organizational Studies*, (September), 1–18. <https://doi.org/10.1177/1548051817727703>

Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Propriedades psicométricas na avaliação de instrumentos: Avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*, 26(3), 649–659. <https://doi.org/10.5123/S1679-49742017000300022>

Spector, P. E., Rosen, C. C., Richardson, H. A., Williams, L. J., & Johnson, R. E. (2019). A new perspective on method variance: A measure-Centric approach. *Journal of Management*, 45(3), 855–880.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions : The multi-unidimensional pairwise-preference model. *Applied*, 29(3), 184–203. <https://doi.org/10.1177/0146621604273988>

Sutton, G. (2019). *The Dark Triad in personnel selection: An exploration of narcissism, psychopathy, and Machiavellianism in candidates involved in an organisational selection process*. University of Canterbury.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 101(2), 266–270.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33(4),

529–554.

- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). An introduction and a point of view. In R. Tourangeau, L. J. Rips, & K. Rasinski (Eds.), *The psychology of survey response* (pp. 1–22). Cambridge, UK: Cambridge University Press.
- Tutz, G., & Schauberger, G. (2018). *Uncertainty as Response Style in Latent Trait Models*.
- Usami, S., Sakamoto, A., Naito, J., & Abe, Y. (2016). Developing pairwise preference-based personality test and experimental investigation of its resistance to faking effect by item response model. *International Journal of Testing*, *16*(4), 288–309. <https://doi.org/10.1080/15305058.2016.1145123>
- van Hooft, E. A. J., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*(2), 301–316. <https://doi.org/10.1037/a0025711>
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217. <https://doi.org/10.1093/ijpor/eds021>
- Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance*, *19*(3), 175–199. [https://doi.org/10.1207/s15327043hup1903\\_1](https://doi.org/10.1207/s15327043hup1903_1)
- Vecchione, M., Alessandri, G., & Barbaranelli, C. (2013). Measurement and application of egoistic and moralistic self-enhancements. *International Journal of Selection and Assessment*, *21*(2), 170–182. <https://doi.org/10.1111/ijsa.12027>
- Viswesvaran, C., & Ones, D. S. (1999). Meta-Analyses of Fakability Estimates: Implications for Personality Measurement. *Educational and Psychological Measurement*, *59*(2), 197–210. <https://doi.org/10.1177/00131649921969802>
- Vize, C. E., Lynam, D. R., Collison, K. L., & Miller, J. D. (2018). Differences among dark

- triad components: A meta-analytic investigation. *Personality Disorders: Theory, Research, and Treatment*, 9(2), 101.
- Wang, W., Qiu, X., Chen, C., Ro, S., & Jin, K. (2017). Item response theory models for ipsative tests with Multidimensional Pairwise Comparison Items. *Applied Psychological Measurement*, 41(8), 1–14. <https://doi.org/10.1177/0146621617703183>
- Watts, A. L., Lilienfeld, S. O., Edens, J. F., Douglas, K. S., Skeem, J. L., Verschuere, B., & LoPilato, A. C. (2016). Does response distortion statistically affect the relations between self-report psychopathy measures and external criteria? *Psychological Assessment*, 28(3), 294–306. <https://doi.org/10.1037/pas0000168>
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010a). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247. <https://doi.org/10.1016/j.ijresmar.2010.02.004>
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010b). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247. <https://doi.org/https://doi.org/10.1016/j.ijresmar.2010.02.004>
- Weijters, B., Geuens, M., & Baumgartner, H. (2013). The Effect of Familiarity with the Response Category Labels on Item Response to Likert Scales. *Journal of Consumer Research*, 40(2), 368–381. <https://doi.org/10.1086/670394>
- Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956–972. <https://doi.org/10.1177/0013164404268674>
- Wetzel, E., & Greiff, S. (2018). The World Beyond Rating Scales. *European Journal of Psychological Assessment*, 34(1), 1–5. <https://doi.org/10.1027/1015-5759/a000469>

- Willits, F. K., Theodori, G. L., & Luloff, A. E. (2016). Another look at Likert scales. *Journal of Rural Social Sciences*, 31(3), 126–139. Retrieved from [http://journalofruralsocialsciences.org/pages/Articles/JRSS\\_2016\\_31/3/JRSS\\_2016\\_31\\_3\\_126-139.pdf](http://journalofruralsocialsciences.org/pages/Articles/JRSS_2016_31/3/JRSS_2016_31_3_126-139.pdf)
- Ziegler, Matthias. (2015). “F\*\*\* you, I won’t do what you told me!”—Response biases as threats to psychological assessment. *European Journal of Psychological Assessment*, 31(3), 153–158. <https://doi.org/10.1027/1015-5759/a000292>
- Ziegler, Matthias, & Buehner, M. (2009). Modeling Socially Desirable Responding and Its Effects. *Educational and Psychological Measurement*, 69(4), 548–565.