

## Uso do ChatGPT para estimação da desejabilidade social de itens de autorelato

Felipe Valentini<sup>1</sup>, Vithor Rosa Franco<sup>2</sup>, Nelson Hauck<sup>3</sup>, Leonardo Botinhon de Campos<sup>4</sup>,  
Letícia da Silva de Sousa<sup>5</sup>

<sup>1</sup> Universidade São Francisco, valentini.felipe@gmail.com

<sup>2</sup> Universidade São Francisco

<sup>3</sup> Universidade São Francisco

<sup>4</sup> Universidade São Francisco

<sup>5</sup> Universidade São Francisco

### Resumo

Este estudo explora o uso do ChatGPT para estimar cargas fatoriais de desejabilidade social (DS) em avaliações de competências socioemocionais, um viés que pode distorcer escores de autorelato. Foram utilizados 12 itens balanceados entre características desejáveis e indesejáveis. No primeiro passo, o ChatGPT avaliou a desejabilidade social dos itens, resultando em uma média próxima de zero. No segundo passo, os valores de DS estimados no passo anterior fixados como cargas fatoriais em um modelo de equações estruturais, que apresentou bom ajuste. A variância de DS estimada em 0,19 indica que o viés de desejabilidade social influencia significativamente os escores.

**Palavras-chave:** ChatGPT, desejabilidade social, competências socioemocionais, viés de resposta, autorelato

### 1. Introdução

Nos últimos anos, diversas avaliações educacionais em larga escala passaram a incluir dados sobre competências socioemocionais dos estudantes. Na maioria dessas avaliações, utilizam-se instrumentos de autorelato, que são amplamente aplicados por sua simplicidade, facilidade de implementação e baixo custo. No entanto, os escores gerados por esses instrumentos podem ser influenciados por vieses de resposta, que são erros sistemáticos não relacionados ao traço psicológico que se deseja mensurar (Bäckström et al., 2012; Primi et al., 2020).

Um desses vieses é a desejabilidade social (DS), que se refere à tendência de os respondentes ajustarem suas respostas para causar uma impressão positiva (Ziegler et al., 2012). Por exemplo, uma pessoa introvertida, que evita interações sociais com novas pessoas e prefere ouvir e refletir a falar, pode endossar o item “sou uma pessoa comunicativa” durante um processo seletivo para um cargo de vendedor, acreditando que essa característica seja essencial para a vaga.

A desejabilidade social envolve pelo menos três aspectos: características pessoais (como a importância dada pelo indivíduo à boa impressão), a situação de avaliação (se o resultado é relevante para o respondente) e o conteúdo do item. Este último ponto considera que os itens são multidimensionais, compostos por um conteúdo descritivo (ligado ao fator psicológico) e por um conteúdo valorativo (a valência da desejabilidade social) (Peabody, 1967).

Modelar a desejabilidade social em avaliações psicométricas é desafiador. Embora seja possível estimar modelos multidimensionais ou bifatoriais, o fator de desejabilidade social frequentemente se correlaciona com competências socioemocionais como amabilidade, dificultando a separação entre conteúdo descritivo e valorativo. Além disso, a inclusão de uma dimensão adicional, sem controle ou restrições, pode superestimar o viés de resposta.

Uma forma de mitigar esse problema é impor restrições nas cargas fatoriais dos itens. Pettersson et al. (2012), por exemplo, utilizaram avaliações de juízes especialistas para determinar o quão desejável cada item era. A média dessas avaliações foi usada para fixar as cargas fatoriais da desejabilidade social. No entanto, esse método apresenta limitações, como a subjetividade das avaliações e a baixa concordância entre os juízes. Este estudo propõe o uso da inteligência artificial, especificamente o ChatGPT, para estimar as cargas fatoriais utilizadas na modelagem da desejabilidade social.

## 2. Objetivos

O objetivo do nosso estudo é estimar as cargas fatoriais da desejabilidade social utilizando o ChatGPT e testá-lo em uma modelagem com dados reais.

## 3. Metodologia

Para estimar a desejabilidade social, solicitamos ao ChatGPT que avaliasse cada item com base em uma definição detalhada do conceito e no conteúdo dos itens. O ChatGPT utilizou uma escala de -1 a 1 para classificar a desejabilidade social dos itens. Essas avaliações foram então incorporadas à modelagem dos dados reais.

Os dados foram coletados de uma amostra de 430 estudantes universitários, dos quais 341 eram mulheres (79,3%), com idades entre 18 e 45 anos ( $M = 25,3$ ,  $DP = 7,8$ ). Os participantes responderam a 12 itens de competências socioemocionais, balanceados quanto ao conteúdo descritivo (positivo e negativo) e valorativo (desejável e indesejável), conforme a seguinte distribuição: três itens positivos e desejáveis, três positivos e indesejáveis, três negativos e desejáveis, e três negativos e indesejáveis. O instrumento utilizou itens do tipo Likert, com cinco pontos de ancoragem.

Utilizamos modelagem por equações estruturais para analisar os dados, estimando cinco dimensões específicas para as competências socioemocionais e uma dimensão geral de desejabilidade social. As cargas fatoriais da desejabilidade social foram fixadas nos valores estimados pelo ChatGPT (versão 4).

### 3. Resultados e Discussão

O primeiro passo foi estimar o nível de desejabilidade social (DS) dos itens por meio do ChatGPT. A média das avaliações foi de  $-0,05$  ( $DP = 0,63$ ), refletindo o balanceamento entre itens desejáveis e indesejáveis. Alguns itens foram classificados como altamente desejáveis (por exemplo, “Consigno ficar calmo mesmo diante de situações difíceis”,  $DS = +1$ ), outros como extremamente indesejáveis (“Sou indiferente, frio e distante dos outros”,  $DS = -1$ ), e alguns como neutros (“Consigno expressar minha irritação”,  $DS = 0$ ).

Os itens com polos congruentes (positivo e desejável ou negativo e indesejável) apresentaram médias de DS mais elevadas ( $-0,66$  e  $+0,66$ ), enquanto os itens com polos incongruentes (negativo e desejável ou positivo e indesejável) mostraram médias mais baixas (negativo e desejável =  $0,3$ ; positivo e indesejável =  $-0,48$ ).

No segundo passo, estimamos o modelo fatorial com três dimensões específicas e uma dimensão geral. A dimensão geral, que captou a desejabilidade social, teve suas cargas fixadas nos valores estimados pelo ChatGPT. O modelo apresentou um bom ajuste aos dados ( $\chi^2/gf = 123,64/50$ ;  $CFI = 0,98$ ;  $TLI = 0,98$ ;  $RMSEA = 0,05$ ). Todas as cargas fatoriais foram consistentes com o polo descritivo dos itens (por exemplo, itens positivos apresentaram cargas positivas, mesmo os indesejáveis), e os valores absolutos das cargas foram superiores a  $0,40$ . Isso sugere que os fatores captaram o conteúdo descritivo dos itens, e não a desejabilidade social. Além disso, a variância atribuída à desejabilidade social foi estimada em  $0,19$ , indicando que uma parcela significativa da variância dos itens pode ser explicada por esse viés.

## Referências

- 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). Oxford University Press, Inc.
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology*, 7(4), Pt.2), 1–18.  
<https://doi.org/10.1037/h0025230>
- Pettersson, E., Turkheimer, E., Horn, E. E., & Menatti, A. R. (2012). The General Factor of Personality and Evaluation. *European Journal of Personality*, 26(3), 292–302.  
<https://doi.org/10.1002/per.839>
- Primi, R., Hauck-Filho, N., Valentini, F., & Santos, D. (2020). Classical Perspectives of Controlling Acquiescence with Balanced Scales. In: M. Wiberg, D. Molenaar, J. González, U. Böckenholt, J. S. Kim (Eds.), *Quantitative Psychology. IMPS 2019. Springer Proceedings in Mathematics & Statistics*, vol 322. Springer, Cham.  
[https://doi.org/10.1007/978-3-030-43469-4\\_25](https://doi.org/10.1007/978-3-030-43469-4_25)
- Ziegler, M., Maccann, C., & Roberts, R. D. (2012). Faking: Knows, Unknowns, and Points of Contention. In M. Ziegler, C. Maccann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3–16). Oxford.