Research report

Editor:
Evandro Morais Peixoto

Conflict of interest:
The authors declare that there are no conflicts of interest.

Received: 13/01/2025
Final Version: 12/12/2025
Approved: 22/12/2025

https://doi.org/10.1590/14138271202631e293381

DATA AVAILABILITY STATEMENT: The research data are not available.

Psychometric properties of Generalized Anxiety Disorder-7 during COVID-19: item response theory analysis

Beatriz Oliveira-Santos[1]
Luanna dos Santos Silva[1]
Dandara Palhano[1]
André Faro[1]

[1]*Federal University of Sergipe, São Cristóvão, Sergipe, Brazil*

Abstract
This study evaluated the psychometric properties of the Generalized Anxiety Disorder Scale 7-item in a Brazilian sample during the COVID-19 pandemic using item response theory. The graded response model was applied to estimate discrimination and difficulty parameters and generate the item characteristic curve, item information curve, and test information curve. Differential Item Functioning (DIF) was analyzed by gender, and anxiety and theta levels were assessed. The sample included 10,069 participants, with a mean age of 36 years (standard deviation = 13.67). The scale exhibited satisfactory performance, with items associated with worry being particularly informative. Item 5 was highlighted for identifying severe anxiety symptoms. No differential item functioning was observed between genders. Collectively, these results reinforce the reliability of the Generalized Anxiety Disorder 7-item as an effective tool for assessing anxiety symptoms.
Keywords: Anxiety; COVID-19; Item Response Theory; Psychometrics; Health Psychology

Propriedades psicométricas da Generalized Anxiety Disorder-7 durante a COVID-19: análise da teoria de resposta ao item

Resumo
Este estudo avaliou as propriedades psicométricas da *Generalized Anxiety Disorder-7* em uma amostra brasileira durante a pandemia da Covid-19, utilizando a Teoria de Resposta ao Item. Aplicou-se o Modelo de Resposta Gradual para estimar parâmetros de discriminação e dificuldade, além de gerar as Curvas Características dos Itens, Curva de Informação dos Itens e Curva de Informação do Teste. O Funcionamento Diferencial do Item (DIF) foi analisado para gênero, e os níveis de ansiedade e theta foram verificados. A amostra incluiu 10.069 participantes, com idade média de 36 anos (desvio-padrão = 13,67). A escala apresentou desempenho satisfatório, com itens sobre preocupações sendo especialmente informativos. O item 5 destacou-se na identificação de sintomas severos de ansiedade. Não foi observado DIF entre gêneros. Em conjunto, os resultados reforçam a confiabilidade da GAD-7 como uma ferramenta eficaz para avaliação dos sintomas de ansiedade.
Palavras-chave: Ansiedade; COVID-19; Teoria de Resposta ao Item; Psicometria; Psicologia da Saúde

Propiedades psicométricas del Trastorno de Ansiedad Generalizada-7 durante la COVID-19: análisis de la teoría de respuesta al ítem

Resumen
Este estudio evaluó las propiedades psicométricas del Generalized Anxiety Disorder-7 en una muestra brasileña durante la pandemia de Covid-19, utilizando la Teoría de Respuesta al Ítem. Se aplicó el Modelo de Respuesta Gradual para estimar los parámetros de discriminación y dificultad, además de generar las Curvas Características de los Ítems, la Curva de Información de los Ítems y la Curva de Información de la Prueba. Se analizó el Funcionamiento Diferencial del Ítem (DIF) por género y se verificaron los niveles de ansiedad y theta. La muestra incluyó a 10 069 participantes, con una edad media de 36 años (desviación estándar = 13,67). La escala presentó rendimiento satisfactorio, especialmente informativos los ítems sobre preocupaciones. El ítem 5 destacó en la identificación de síntomas graves de ansiedad. No se observó DIF entre géneros. Los resultados refuerzan la fiabilidad del GAD-7 como una herramienta eficaz para la evaluación de los síntomas de ansiedad.
Palabras clave: Ansiedad; SARS-CoV-2; Teoría de respuesta al ítem; Psicometría; Psicología de la salud

Introduction

Three years following the COVID-19 pandemic's onset, the World Health Organization announced, in May 2023, the end of the Public Health Emergency of International Concern (Pan American Health Organization, 2023). During this period, the pandemic evolved through different phases — namely, pre-crisis, in-crisis, and post-crisis (Faro et al.,

2020). The emergence of multiple SARS-CoV-2 variants posed novel challenges to global health. In Brazil, the Gamma variant, detected in April 2021, resulted in the health system's collapse and markedly increased deaths. Subsequently, between July and November 2021, the Delta variant's predominance coincided with heightened vaccine coverage and reduced cases. However, between December 2021 and January 2022, the Omicron variant exponentially increased reported cases, surpassing previous waves in magnitude (Oswaldo Cruz Foundation [Fiocruz], 2022). The pandemic's consequences extended beyond physical impacts, highlighting COVID-19 as a mental health crisis (Faro et al., 2020). During this period, anxiety was particularly prominent. A meta-analysis covering studies from the first months of the pandemic, from December 2019 to August 2020, revealed that the prevalence of anxiety was three times higher in the general population (Santabárbara et al., 2021). Globally, a systematic review and meta-analysis from December 2019 to March 2023 found that the pooled prevalence of anxiety was 29% (Bin et al., 2023). In a study conducted during the pandemic's first year in Brazil, 80% of respondents reported anxiety symptoms (Goularte et al., 2021).

Considering the high incidence of anxiety symptoms during the pandemic, the need for robust and reliable research instruments to assess this condition became evident. Thus, investigating psychometric stability in this new context is important, as validity is not a fixed property of an instrument but depends on the adequacy of inferences to the context and target population (Boateng et al., 2018; Messick, 1995). In this regard, the Generalized Anxiety Disorder 7-item (GAD-7) is widely recognized for its effectiveness in measuring anxiety symptoms and is among the most commonly used instruments for this purpose (Doi et al., 2018; Hinz et al., 2017; Lee & Kim, 2019). As GAD-7 was developed and validated using classical psychometric procedures, such as item–total correlation and internal consistency, the scale aims to identify possible cases of generalized anxiety disorder (Spitzer et al., 2006). Alongside its brevity, unidimensionality, and ease of self-report, the GAD-7 exhibits satisfactory diagnostic accuracy, item clarity, and administrative efficiency (Mughal et al., 2020)—characteristics that support its use in epidemiological studies and contexts warranting rapid screening.

The GAD-7's psychometric properties have been extensively tested in studies based on Classical Test Theory (CTT), indicating satisfactory levels of validity and reliability. Studies conducted in different countries—such as South Korea (Lee & Kim, 2019), Saudi Arabia (Alghadir et al., 2020), and Finland (Tiirikainen et al., 2019)—confirm the adequacy of the scale's structure, with favorable results for a unidimensional model and consistent levels of measurement precision. Various studies in the Brazilian context also reinforce this evidence. Investigations with university and high school students found a satisfactory fit for a one-factor model and adequate internal consistency (Moreno et al., 2016). Research with adolescents reported robust indices of factorial validity, including confirmation of unidimensionality and expected correlations with other measures of anxiety and depression, indicating concurrent validity (Leite & Faro, 2022). Similar results were observed in a non-clinical sample, which demonstrated acceptable fit indices and reliability (Silva et al., 2023). Overall, this converging evidence demonstrates that the GAD-7 is a valid, reliable, and structurally stable measure across different contexts, while also reinforcing the need for ongoing validation processes to ensure its applicability in new populations. Considering that the GAD-7 was developed based on the diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders, whose editions are periodically updated, instruments derived from this framework must also be regularly reviewed and updated (DeSousa et al., 2013).

Additionally, it is important to highlight the need to examine measurement invariance in instruments over time (Putnick & Bornstein, 2016). As the prevalence of anxiety tends to be higher among women (Jalnapurkar et al., 2018) and because differences in item functioning may prevail between groups (Putnick & Bornstein, 2016), verifying whether the GAD-7 measures the construct equivalently for men and women is essential. Evidence of invariance ensures that comparisons between genders are valid and do not result from measurement bias. Alongside CTT, another important approach in test evaluation is Item Response Theory (IRT; Araujo et al., 2009), which comprises a set of mathematical models based on the premise that individuals' latent characteristics influence the probability of responding to items in a particular way (Jordan et al., 2017). The latent trait is defined as a characteristic that is not directly observable, such as anxiety. To assess these variables, using observable secondary variables, such as the items of a scale, is necessary (Pasquali & Primi, 2003).

Notably, IRT aims to complement CTT rather than replace it (Sartes & Souza-Formigoni, 2013), offering significant advantages. It does not require

the assumption of a metric scale and is suitable for categorical responses. Additionally, IRT differentiates item difficulty from discrimination capacity, unlike CTT, which may exclude items owing to their extreme difficulty. Moreover, IRT provides more precise confidence intervals for individuals' latent traits, recognizing that measurement precision varies according to the true score (Jordan et al., 2017). Another important difference in IRT is that the results are independent of the sample used, ensuring the invariance of these parameters within the model. By contrast, in CTT, item parameters are influenced by respondents' latent traits, requiring the sample to be representative of the population under study (Sartes & Souza-Formigoni, 2013). Another advantage of IRT is that, unlike CTT, it focuses on individual items rather than considering the test as a whole (Pasquali & Primi, 2003). Further, IRT uses different models that vary in the parameters employed to describe item characteristics. The one-parameter logistic model focuses only on item difficulty (b). The two-parameter logistic model, in addition to item difficulty (b), adds a discrimination parameter (a). The three-parameter logistic model includes a guessing parameter (c) in addition to (a) and (b) (Bichi & Talibi, 2018). Among the models widely utilized in IRT, the Graded Response Model (GRM) (Samejima, 1969) stands out. This model assumes that items may have more than two response categories, specifying multiple category threshold parameters, which enables a more detailed modeling of individuals' responses in relation to their latent level on the construct being assessed. Moreover, IRT is considered the gold standard for the assessment of Differential Item Functioning (DIF) (Gao & Liu, 2024), which evaluates whether items function equivalently across different population subgroups. Its identification and correction are essential to ensure instrument precision, guaranteeing that results are reliable across diverse contexts and populations (Ellis, 1989).

Considering the adaptive challenges imposed by COVID-19, the need for a robust measure of anxiety becomes evident. Investigating the psychometric properties of the GAD-7 is essential to establish sound criteria for identifying anxiety symptoms, facilitating a better understanding of the pandemic's impact on the Brazilian population's mental health. This study primarily aimed to establish a reliable parameter for assessing anxiety in a Brazilian sample within the context of the COVID-19 pandemic, through the investigation of the GAD-7's psychometric properties using IRT. The specific objectives included analyzing discrimination and difficulty parameters, the item characteristic curve (ICC), the item information curve (IIC), and the test information curve (TIC) using the IRT model, as well as evaluating the presence of DIF in the scale between genders and identifying levels of anxiety and theta in the sample.

Method

Participants

This study included 10,069 participants, comprising both genders, 88.2% of whom were women (n = 8,880), with a mean age of 36.1 years (standard deviation [*SD*] = 13.67; minimum = 18 and maximum = 87). Most of the sample self-identified as White (54.4%), followed by Brown (34.5%), Black (9.7%), Asian (0.8%), Other (0.4%), and Indigenous (0.2%). Participants were drawn from all regions of the country, with greater representation from the Northeast (n = 3,951; 39.2%), followed by the Southeast (n = 3,891; 38.6%), South (n = 1,388; 13.8%), Midwest (n = 528; 5.2%), and North (n = 311; 3.1%).

Instruments

The GAD-7 (Spitzer et al., 2006) was utilized to measure anxiety symptoms. The GAD-7 comprises seven items rated on a 4-point Likert scale ranging from "not at all" (0) to "nearly every day" (3). The total score ranges from 0 to 21 and is classified into the following four strata: 0–4 (no anxiety), 5–9 (mild symptoms), 10–14 (moderate symptoms), and 15–21 (severe symptoms). A score of 10 or higher indicates clinically significant anxiety symptoms (Spitzer et al., 2006). The Cronbach's alpha was 0.89, indicating excellent internal consistency. A sociodemographic questionnaire was administered to characterize the sample, including questions on age (in years), gender (woman, man), skin color or race (White, Black, Brown, Asian, Indigenous, other), and region (Northeast, Southeast, South, Midwest, North).

Procedures and ethical aspects

This study was approved by the National Research Ethics Commission (CONEP; CAAE: registration: 30485420.6.0000.0008). The data were collected online over four years. The first data collection phase was conducted in the second half of March 2020, while the second phase was conducted in early June. In 2021, the third data collection period occurred between March

and April. In 2022 and 2023, data were collected in the second half of March. Individuals were invited to participate through social media platforms, such as Instagram, Facebook, and WhatsApp. Participant consent was obtained through acceptance of the informed consent form, presented on the online questionnaire's initial screen. Exclusion criteria included being under 18 years of age and not residing in Brazil at the time of data collection. A cross-sectional design was adopted, using convenience sampling through the snowball method.

Data analysis

The SPSS software (version 25) was used to conduct model fitting procedures and descriptive analyses, such as frequencies, means, standard deviations, and percentages. The GAD-7 was evaluated based on IRT, with analyses conducted in RStudio using the mirt package (Chalmers, 2012). Considering the polytomous and ordinal nature of the scale items, the GRM was applied (Samejima, 1969). Item discrimination (a) and difficulty (b) parameters were estimated, and the following were generated and analyzed: the ICC, which illustrates a specific response's probability as a function of the latent trait level (theta); the item information function (IIF), which indicates the amount of information each item provides across different levels of theta; and the TIC, which presents the amount of information provided by the test as a whole across different levels of theta (Bichi & Talib, 2018; Pasquali, 2007). Theta values, which represent participants' anxiety levels, were calculated based on responses to the scale. Further, DIF was investigated to ensure that GAD-7 items functioned equivalently across genders (with a non-significant value expected). Model fit was evaluated using fit quality indices, including the root mean square error of approximation, with an expected value below 0.08, and the Tucker–Lewis index, with a desired value above 0.90.

Results

The mean total GAD-7 score was 11.4 (SD = 5.80), with a minimum score of 0 and a maximum score of 21. Regarding theta, values ranged from −2.30 to 1.45. Most participants reported severe anxiety (35.2%), followed by mild (27.5%), moderate (24.6%), and normal levels (12.8%). Model fit indices were 0.054 for root mean square error of approximation and 0.99 for the Tucker–Lewis index. Table 1 presents the discrimination (*a*) and difficulty (*b*) parameters for the seven GAD-7 items in the two-parameter IRT model. All items exhibited satisfactory values (Bichi & Talib, 2018) for the discrimination parameter (M = 2.5; SD = 0.69), ranging from 1.66 (item 5) to 3.27 (item 2). Higher discrimination was observed for item 2 (3.27), which addressed difficulty controlling worries, followed by item 3 (3.25), related to levels of excessive worry, and item 1 (3.14), which assessed feeling anxious, nervous, or tense. The lowest discrimination value was observed for item 5 (1.66), related to restlessness.

Low values for the difficulty parameter indicated that participants tended to endorse low difficulty responses across all items on the GAD-7 scale. The lowest difficulty threshold (b1–b3) was identified for item 3, with a mean of −2.03. By contrast, item 5 stood out as the most difficult, exhibiting a lower probability of endorsement in response categories 3 ("more than half the days") and 4 ("nearly every day"). These categories correspond to the b3 difficulty parameter, which presented a value of 1.54. Additionally, item 5 demonstrated the highest mean (b1–b3), with a value of 1.04.

Figure 1 presents the ICC of the participants' responses. The response categories 1, 2, and 4 demonstrated satisfactory coverage across theta levels. Category 1 ("not at all") covered theta values ranging from −6 to −4. Category 2 ("several days") stood out in terms of coverage, except for items 5 and 7. Category 4 demonstrated good coverage, encompassing theta values from 2 to 6 for items 1–4. In turn, category 3 was overlapped by the other categories across all items. Figure 2 presents the IIC. Items 1, 2, and 3 were identified as the most informative, with greater information in theta intervals between −2 and 1, whereas items 5 and 7 were identified as the least informative. The total information area was 41.25. Figure 3 presents the TIC. In sum, there was greater information between theta values of −3 and 2.5; outside this interval, higher error and a reduction in information were observed.

Table 2 presents the results of the DIF analysis by gender. Considering the large sample size, $p < 0.001$ was adopted as the criterion indicative of DIF. No statistically significant differences were observed in *p*-values across analyses, indicating that the GAD-7 measured men and women equivalently at the same level of the latent anxiety trait (theta).

Discussion

This study aimed to establish a reliable parameter for assessing anxiety in a Brazilian sample in the

Table 1.
a and b parameters for the items of the Generalized Anxiety Disorder Scale–7 (GAD-7)

Items	Discrimination a	Difficulty b1	Difficulty b2	Difficulty b3	S-X²	p	M b1-b3
1. Feeling nervous, anxious, or very tense	3,14	-1,61	-0,18	0,26	118,00	<0,01	-1,70
2. Being unable to stop or control worrying	3,27	-1,29	0,00	0,54	118,11	<0,01	-1,11
3. Worrying too much about various things	3,25	-1,79	-0,29	0,15	317,90	<0,01	-2,03
4. Difficulty relaxing	2,86	-1,45	-0,14	0,36	111,80	<0,01	-1,47
5. Becoming so restless that it is hard to sit still	1,66	-0,30	0,83	1,54	455,02	<0,01	1,04
6. Becoming easily annoyed or irritable	2,01	-1,44	0,09	0,67	195,69	<0,01	-1,13
7. Feeling afraid as if something awful might happen.	1,93	-0,90	0,27	0,78	221,89	<0,01	-0,37

Note. a = item discrimination parameter; b = item difficulty parameter; M = mean

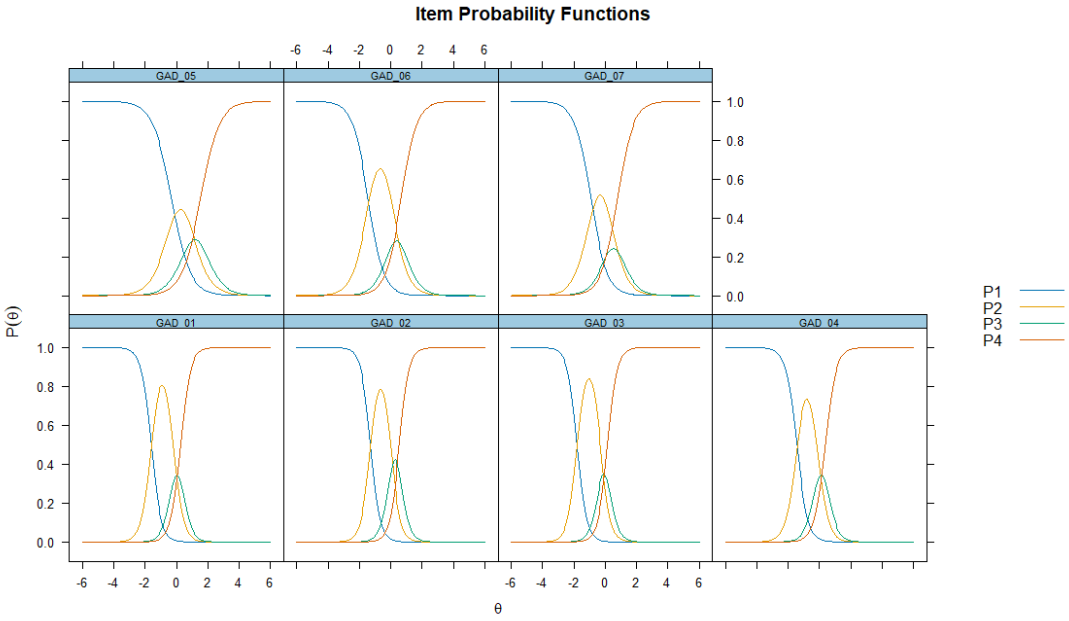

Figure 1. Item Characteristic Curves of the Generalized Anxiety Disorder Scale–7 (GAD-7)

context of the COVID-19 pandemic by investigating the GAD-7's psychometric properties using IRT. Severe anxiety symptoms were the most prevalent in the sample. Regarding item analysis, all items exhibited adequate discrimination values, with particular emphasis on those related to worries. Concerning item difficulty, items tended to show low difficulty across all response categories. The ICC indicated satisfactory coverage of anxiety levels for response categories 1 ("not at all"), 2 ("several days"), and 4 ("nearly every day"), with category 3 ("more than half the days") being overlapped by the others. Items 1, 2,

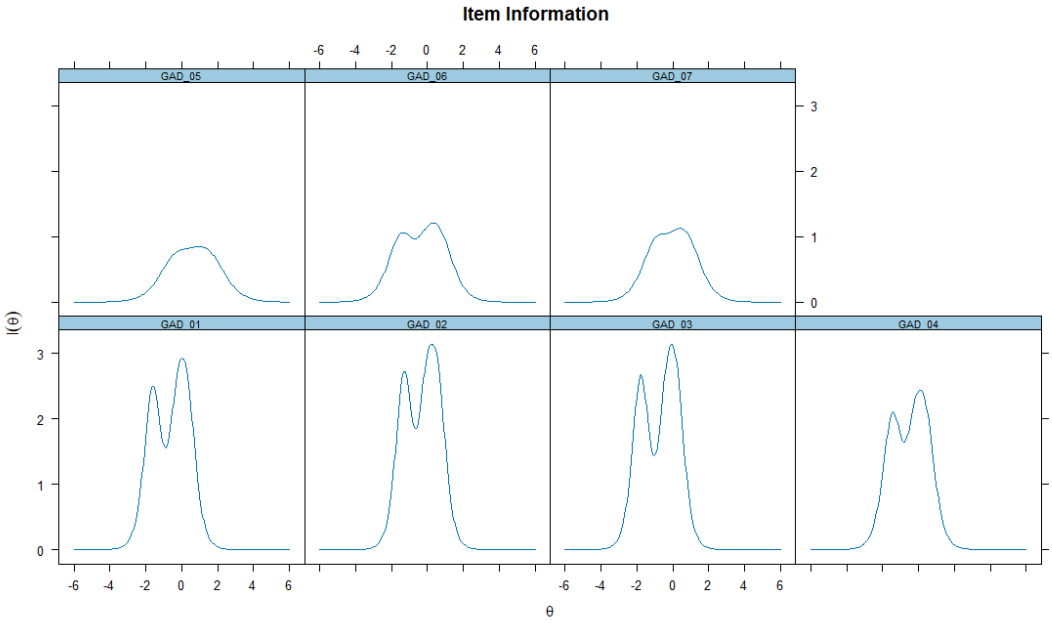

Figure 2. Item Information Curve of the Generalized Anxiety Disorder Scale–7 (GAD-7)

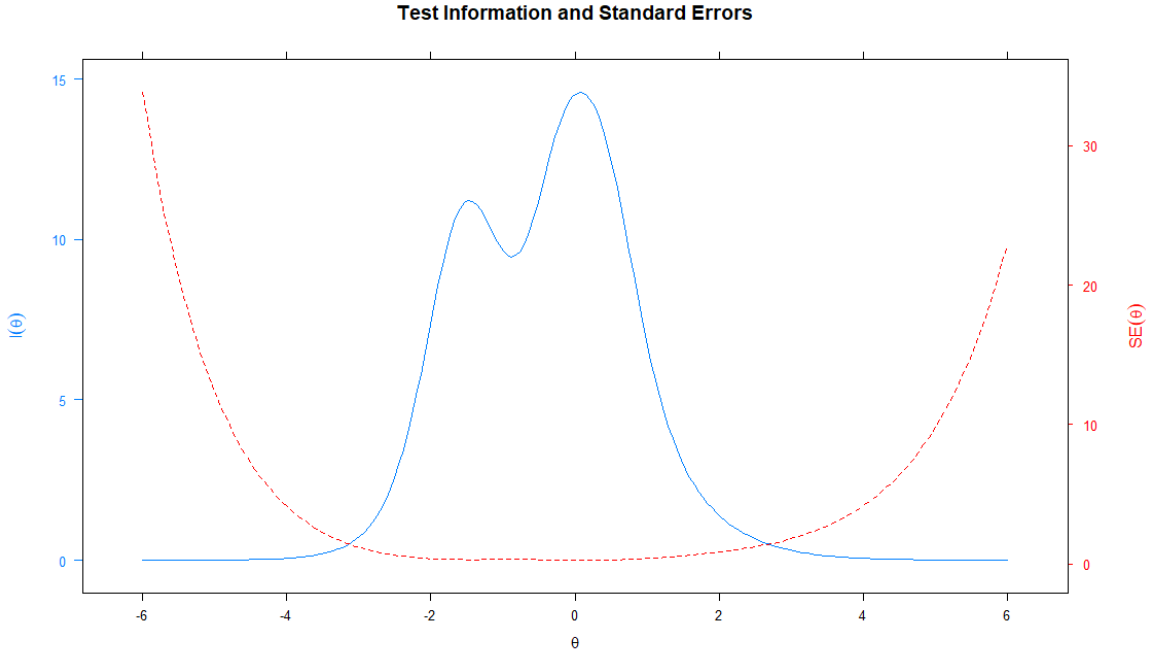

Figure 3. Test Information Curve of the Generalized Anxiety Disorder Scale–7 (GAD-7)

and 3 were the most informative, especially in theta intervals between -2 and 1, whereas items 5 and 7 provided less information. No DIF was observed in the analysis across genders.

A higher frequency of severe anxiety symptoms was observed in the sample (35.2%), suggesting that the COVID-19 pandemic's impact remained significant across its different stages, considering that this study

Table 2.
Differential Item Functioning of the Generalized Anxiety Disorder Scale–7 (GAD-7) by Gender

Items	p	X2	DF
1. Feeling nervous, anxious, or very tense.	0.011	17.051	4
2. Being unable to stop or control worrying.	0.138	7.363	4
3. Worrying too much about various things.	0.011	16.007	4
4. Difficulty relaxing.	0.377	4.219	4
5. Becoming so restless that it is hard to sit still.	0.013	14.682	4
6. Becoming easily annoyed or irritable.	0.138	7.567	4
7. Feeling afraid as if something awful might happen.	0.076	9.822	4

Note. P = adjusted p-value; χ^2 *= chi-square; DF = degrees of freedom*

was conducted from 2020 to 2023. The persistence of these symptoms is understandable, as Brazil experienced several waves of reinfections associated with multiple virus variants. During the pandemic, the Brazilian health system faced periods of collapse, alongside fluctuations in vaccination coverage and in the numbers of cases and deaths with the emergence of new variants (Fiocruz, 2022). Given the instability and uncertainty generated by the virus, it is natural that severe anxiety symptoms predominated in this sample, as challenges and stressful situations persisted over the years.

Adequate discrimination was observed in the analysis of the GAD-7 items, indicating effectiveness in distinguishing between individuals with different anxiety levels (Bichi & Talib, 2018). Items 2 ("not being able to stop or control worrying"; a = 3.27), 3 ("worrying too much about different things"; a = 3.25), and 1 ("feeling nervous, anxious, or very tense"; a = 3.14) stood out in particular. These items were not only the most discriminative but also the most informative, as evidenced by the IIC analysis. Considering the content of these items, they appear particularly related to worries, as well as to the experience of anxiety symptoms—central aspects in the diagnosis of generalized anxiety disorder in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (American Psychiatric Association [APA], 2022). These items' relevance further corroborates evidence from brief instruments, such as the Generalized Anxiety Disorder-2 (GAD-2; Kroenke et al., 2007) and Patient Health Questionnaire-4 (PHQ-4; Kroenke et al., 2009), as brief measures are important for rapid screening and initial assessment in clinical practice and research contexts. The GAD-2 comprises the first two items of the GAD-7, whereas the PHQ-4 includes the first two items of the GAD-7 and a pair of items from the PHQ-9. A study conducted with primary care patients in Hamburg, which evaluated GAD-7's psychometric properties using IRT, also found higher discrimination for items 2 (3.55) and 3 (3.10) (Jordan et al., 2017). Additionally, high information was observed for items 1 and 2, and moderate information for item 3. These findings support the current study's results, highlighting these items' clinical importance in identifying and assessing anxiety.

Compared with items 1, 2, and 3, item 5 ("being so restless that it is hard to sit still"; a = 1.66) exhibited lower discrimination, although still within the expected range. This suggests that item 5 may be less sensitive in differentiating and identifying anxiety symptoms among individuals with different anxiety levels. This lower discriminative capacity was also observed in a previous study that analyzed the GAD-7 in a primary care sample (Jordan et al., 2017), in which a value of 1.69 was obtained, highlighting item 5 as the least discriminative in the sample. In addition to having low discrimination, item 5 of the GAD-7 was the most difficult for participants to endorse in categories 3 ("more than half the days") and 4 ("nearly every day") and was the least informative, according to the IIC. By contrast, the other GAD-7 items exhibited low difficulty, suggesting ease in identifying mild symptoms. A difficulty value above 1 indicates that an item is difficult, suggesting that a higher level of the latent trait is required for an individual to endorse the item. Conversely, an item with a difficulty value below -1 is considered easy, indicating that a lower level of the latent trait is sufficient for endorsement (Bichi & Talib, 2018). That is, an item with high difficulty indicates that individuals with lower anxiety

levels are unlikely to endorse response categories that reflect higher symptom frequency. Therefore, the difficulty of item 5 in relation to categories indicating high anxiety levels is an important aspect. Endorsing these categories requires a high anxiety level, which renders item 5 relevant for differentiating individuals with severe symptoms, contributing to the identification of more intense symptoms in the evaluated sample and the sample's overall composition.

Overall, although the GAD-7's global precision is higher for lower anxiety levels, the items with greater difficulty, especially item 5, demonstrate utility in identifying individuals with higher levels of the latent trait. Thus, despite it being used more efficiently as a screening instrument, high scores should be interpreted as indicating a greater likelihood of moderate to severe symptoms, requiring more detailed complementary assessments. The GAD-7 response categories demonstrated satisfactory coverage of theta levels, except for category 3 ("more than half the days"), which was overlapped by other categories across all items. This result may be associated with the semantic similarity between categories 2, 3, and 4, rendering them insufficiently distinct to be considered separate, which makes it difficult for participants to precisely distinguish the frequency with which they experienced the symptoms.

Discussions regarding category 3 have been addressed in other studies that employed IRT to analyze the scale. One study that investigated the GAD-7's psychometric performance in military veterans after orthopedic surgery also identified overlap in category 3. This study suggested merging the response options "more than half the days" and "nearly every day" into a single category, or providing clearer differentiation of the third response option ("more than half the days") in relation to the options "several days" and "nearly every day" (Liu et al., 2022). Moreover, the authors considered the possibility that these results were influenced by the small sample size (n = 67) and lower anxiety levels among participants, leading them to select responses that reflect lower symptom frequency. However, the present study—with a significantly larger sample size and the majority of participants reporting severe anxiety symptoms—suggests that the problems associated with category 3 are not necessarily linked to these factors. Another recent study aimed at validating the GAD-7 in a sample of French adolescents also observed overlap in response category 3 (Todorović et al., 2023). This study, therefore, suggests merging categories 3 and 4, resulting in a single category labeled "more than half the days." This adjustment transforms the 4-point Likert scale into a 3-point scale, as also suggested by Barthel et al. (2014). Another suggestion presented in the literature is the administration of responses in two steps (Weobong et al., 2009). Initially, participants would respond to the items dichotomously, with "yes" or "no" responses. In the case of an affirmative response, the second step would involve a question about the frequency with which they experienced the symptoms.

According to Samejima's model (1969), the reorganization of response categories in polytomous items may be considered based on specific criteria: (a) when the discrimination parameter is below 0.70 or above 4.0; (b) when the difficulty parameter exceeds ± 4.0; and/or (c) when the estimates present very large standard errors (Sartes & Souza-Formigoni, 2013). Based on the current study's results, none of the specified criteria were satisfied. Thus, the regrouping of response category 3, despite its overlap with the others in the analyses, would not be the most optimal option as it would not significantly improve the precision of the parameter estimates. Regarding item functioning, none of the items evaluated men and women differently. The DIF analysis exhibited no significant differences in how the items were answered by both genders, indicating that the GAD-7 has equivalent psychometric properties for men and women. A similar result was identified in a study with patients after traumatic brain injury (Teymoori et al., 2019). Collectively, these results suggest that the GAD-7 presents good indicators for screening anxiety symptoms, enabling the comparison of scores between men and women.

Although this study presents relevant evidence regarding the GAD-7's psychometric properties in a Brazilian sample during the COVID-19 pandemic, some limitations should be acknowledged. Selection bias may have been introduced because individuals who were more psychologically affected were potentially more likely to participate in the convenience sampling. Further, the sample does not fully capture population diversity by adopting a binary gender approach. Future research may consider conducting DIF analyses, including other gender identities, races, and educational levels, to examine possible differences in item functioning, support cultural adaptations of the GAD-7, and explore DIF across different periods. Investigating GAD-7's performance in other crisis contexts or collective adverse events using larger samples and more robust data collection methods is important for assessing the stability and potential generalizability of the observed

psychometric properties. Finally, examining adjustments to the response scale is recommended, given the overlap identified in category 3, which may improve discrimination among reported symptom frequencies.

In conclusion, this study reinforces evidence of the GAD-7's psychometric properties via IRT in the Brazilian context. The absence of DIF between genders and the item functioning, with emphasis on the discriminative capacity of worry-related items and the usefulness of Item 5 for severe symptoms, attests to the measure's quality. The greater precision of the items at lower levels of the latent trait (theta) reinforces the instrument as ideal for initial screening, allowing differentiation between clinical cases and the general population. This characteristic is essential, as brief screening tools optimize resource allocation prior to complex diagnostic assessments (Ali et al., 2016). This applicability directly benefits mental health professionals, epidemiologists, and health managers, supporting decision-making to distinguish individuals with mild anxiety levels from those requiring greater attention and specialized diagnostic evaluations.

References

Alghadir, A., Manzar, M. D., Anwer, S., Albougami, A., & Salahuddin, M. (2020). Psychometric properties of the generalized anxiety disorder scale among saudi university male students. *Neuropsychiatric disease and treatment*, *16*, 1427–1432. https://doi.org/10.2147/NDT.S246526

Ali, G.-C., Ryan, G., & De Silva, M. J. (2016). Validated screening tools for common mental disorders in low and middle income countries: A systematic review. *PloS One*, *11*(6), e0156939. https://doi.org/10.1371/journal.pone.0156939

American Psychiatric Association [APA]. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Araujo, E. A. C. de, Andrade, D. F. de, & Bortolotti, S. L. V. (2009). Teoria da Resposta ao Item. *Revista da Escola de Enfermagem da U S P*, *43*, 1000–1008. https://doi.org/10.1590/s0080-62342009000500003

Barthel, D., Barkmann, C., Ehrhardt, S., Bindt, C., & International CDS Study Group. (2014). Psychometric properties of the 7-item Generalized Anxiety Disorder scale in antepartum women from Ghana and Côte d'Ivoire. *Journal of Affective Disorders*, *169*, 203–211. https://doi.org/10.1016/j.jad.2014.08.004

Bichi, A. A., & Talib, R. (2018). Item Response Theory: na introduction to latente trait models to test and item development. *International Journal of Evaluation and Research in Education*, *17*(2), 124-151. http://doi.org/10.11591/ijere.v7i2.12900

Bin, X., Qu, K. Y., Wang, Y. H., Chen, L., Xiong, Y.J., Wen, J. F., Wei, H. B., Bing, T., Dan, C.-Y., & Zhu, J. Q. (2023). Prevalence of depression, anxiety in China during the COVID-19 pandemic: an updated systematic review and meta-analysis. *Frontiers in Public Health*, *11*, 1267764. https://doi.org/10.3389/fpubh.2023.1267764

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health, 6,* 149. https://doi.org/10.3389/fpubh.2018.00149

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the Renvironment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

DeSousa, D. A., Moreno, A. L., Gauer, G., Manfro, G. G., & Koller, S. H. (2013). Revisão sistemática de instrumentos para avaliação de ansiedade na população brasileira. *Avaliação Psicológica*, *12*(3), 397–410. https://www.redalyc.org/articulo.oa?id=335030096015

Doi, S., Ito, M., Takebayashi, Y., Muramatsu, K., & Horikoshi, M. (2018). Factorial validity and invariance of the 7-item Generalized Anxiety Disorder Scale (GAD-7) among populations with and without self-reported psychiatric diagnostic status. *Frontiers in Psychology*, *9*, 1741. https://doi.org/10.3389/fpsyg.2018.01741

Ellis, B. B. (1989). Differential item functioning: Implications for test translations. The *Journal of Applied Psychology, 74*(6), 912–921. https://doi.org/10.1037/0021-9010.74.6.912

Faro, A., Bahiano, M. A., Nakano, T. C., Reis, C., Silva, B. F. P., & Vitti, L. S. (2020). COVID19 e saúde mental: A emergência do cuidado. *Estudos de Psicologia*, 37, e200074. https://doi.org/10.1590/1982-0275202037e200074

Fundação Oswaldo Cruz [Fiocruz]. (2022). *Boletim observatório Covid-19 – Balanço de dois anos da pandemia Covid-19*. Ministério da Saúde. https://portal.fiocruz.br/documento/boletim-covid-balanco-de-2-anos-da-pandemia

Gao, X., & Liu, Z. (2024). Analyzing the psychometric properties of the PHQ-9 using item response theory in a Chinese adolescent population. *Annals of General Psychiatry*, *23*(1). https://doi.org/10.1186/s12991-024-00492-3

Goularte, J. F., Serafim, S. D., Colombo, R., Hogg, B., Caldieraro, M. A., & Rosa, A. R. (2021). COVI-19 and mental health in Brazil: Psichiatric symptoms in the general population. *Journal of psychiatric research, 132,* 32-37. https://doi.org/10.1016/j.jpsychires.2020.09.021

Hinz, A., Klein, A. M., Brähler, E., Glaesmer, H., Luck, T., Riedel-Heller, ... Hilbert, A. (2017). Psychometric evaluation of the Generalized Anxiety Disorder Screener GAD-7, based on a large German general population sample. *Journal of affective disorders, 210*, 338-344. https://doi.org/10.1016/j.jad.2016.12.012

Jalnapurkar, I., Allen, M., & Pigott, T. (2018). Sex differences in anxiety disorders: a review. *HSOA journal of psychiatry, depression & anxiety, 4*(1), 1-9. http://dx.doi.org/10.24966/PDA-0150/100011

Jordan, P., Shedden-Mora, M. C., & Löwe, B. (2017). Psychometric analysis of the Generalized Anxiety Disorder scale (GAD-7) in primary care using modern item response theory. *PloS One*, *12*(8), e0182162. https://doi.org/10.1371/journal.pone.0182162

Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*, *50*(6), 613–621. https://doi.org/10.1176/appi.psy.50.6.613

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317. https://doi.org/10.7326/0003-4819-146-5-200703060-00004

Lee, B., & Kim, Y. E. (2019). The psychometric properties of the Generalized Anxiety Disorder scale (GAD-7) among Korean university students. *Psychiatry and Clinical Psychopharmacology, 29*(*4*), 864-871. https://doi.org/10.1080/24750573.2019.1691320

Leite, M. F., & Faro, A. (2022). Evidências de Validade da GAD-7 em Adolescentes Brasileiros. *Psico-USF*, *27*(2), 345–356. https://doi.org/10.1590/1413-82712022270211

Liu, W., Dindo, L., Hadlandsmyth, K., Unick, G. J., Zimmerman, M. B., St Marie, B., Embree, J., Tripp-Reimer, T., & Rakel, B. (2022). Item response theory analysis: PROMIS® anxiety form and generalized anxiety disorder scale. *Western Journal of Nursing Research*, *44*(8), 765–772. https://doi.org/10.1177/01939459211015985

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Moreno, A. L., DeSousa, D. A., Souza, A. M. F. L. P., Manfro, G. G., Salum, G. A., Koller, S. H., Osório, F. L., & Crippa, J. A. S. (2016). Factor structure, reliability, and item parameters of the Brazilian-Portuguese version of the GAD-7 questionnaire. *Temas em Psicologia*, *24*(1), 367–376. https://doi.org/10.9788/tp2016.1-25

Mughal, A. Y., Devadas, J., Ardman, E., Levis, B., Go, V. F., & Gaynes, B. N. (2020). A systematic review of validated screening tools for anxiety disorders and PTSD in low to middle income countries. *BMC Psychiatry*, *20*(1), 338. https://doi.org/10.1186/s12888-020-02753-3

Organização Pan-Americana da Saúde [OPAS]. (2023). *OMS declara fim da Emergência de Saúde Pública de Importância Internacional referente à COVID-19*. https://www.paho.org/pt/noticias/5-5-2023-oms-declara-fim-da-emergencia-saude-publica-importancia-internacional-referente

Pasquali, L. (2007). As funções de informação e de eficiência dos itens e do teste. *TRI: Teoria de resposta ao item – Teoria, procedimentos e aplicações.* (1th ed., pp. 53-65). LabPAM/UnB.

Pasquali, L., & Primi, R. (2003). Fundamentos da teoria da resposta ao item: TRI. *Avaliaçao Psicologica: Interamerican Journal of Psychological Assessment*, *2*(2), 99–110. https://dialnet.unirioja.es/servlet/articu

lo?codigo=5115864&info=resumen&idioma=ENG

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Revien, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*, 1–97. https://doi.org/10.1007/bf03372160

Santabárbara, J., Lasheras, I., Lipnicki, D. M., Bueno-Notivol, J., Pérez-Moreno, M., Lópes-Antón, R. ... Gracia-García, P. (2021). Prevalence of anxiety in the COVID-19 pandemic: An updated meta-analysis of community-based studies. *Progress in Neuropsychopharmacology & Biological Psychiatry, 109*, 1-15. https://doi.org/10.1016/j.pnpbp.2020.110207

Sartes, L. M. A., & Souza-Formigoni, M. L. O. (2013). Avanços na psicometria: da Teoria Clássica dos Testes à Teoria de Resposta ao Item. *Psicologia: Reflexão e Crítica*, *26*(2), 241–250. https://doi.org/10.1590/S0102-79722013000200004

Silva, L. S., Leite, M. F., Feitosa, A. L. B., & Faro, A. (2023). Propriedades psicométricas da GAD-7 no Brasil. *Psico*, *54*(1), e39902. https://doi.org/10.15448/1980-8623.2023.1.39902

Spitzer, R.L., Kroenke, K., Williams, J.B.W., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of internal medicine, 166* (10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Teymoori, A., Real, R., Gorbunova, A., Haghish, E. F., Andelic, N., Wilson, L., Asendorf, T., Menon, D., & von Steinbüchel, N. (2020). Measurement invariance of assessments of depression (PHQ-9) and anxiety (GAD-7) across sex, strata and linguistic backgrounds in a European-wide sample of patients after Traumatic Brain Injury. *Journal of Affective Disorders*, *262*, 278–285. https://doi.org/10.1016/j.jad.2019.10.035

Tiirikainen, K., Haravuori, H., Ranta, K., Kaltiala-Heino, R., & Marttunen, M. (2019). Psychometric properties of the 7-item Generalized Anxiety Disorder Scale (GAD-7) in a large representative sample of Finnish adolescents. *Psychiatry Research, 272*, 30-35. https://doi.org/1016/j.psychres.2018.12.004

Todorović, A., Baumann, C., Blanchin, M., & Bourion-Bédès, S. (2023). Validation of generalized anxiety disorder 6 (GAD-6)-A modified structure of screening for anxiety in the adolescent French population. *International Journal of Environmental Research and Public Health*, *20*(8). https://doi.org/10.3390/ijerph20085546

Weobong, B., Akpalu, B., Doku, V., Owusu-Agyei, S., Hurt, L., Kirkwood, B., & Prince, M. (2009). The comparative validity of screening scales for postnatal common mental disorder in Kintampo, Ghana. *Journal of Affective Disorders*, *113*(1–2), 109–117. https://doi.org/10.1016/j.jad.2008.05.009

Acknowledgments:

This work was supported by the Research Coordination Office (COPES) of the Federal University of Sergipe.

About the authors:

Beatriz Oliveira-Santos – MSc student in Psychology, Graduate Program in Psychology, Federal University of Sergipe, São Cristóvão, Sergipe, Brazil.
ORCID: https://orcid.org/0000-0003-0914-4943
E-mail: breatriiz.oliveira@gmail.com

Luanna dos Santos Silva – Luanna dos Santos Silva – PhD in Psychology, Graduate Program in Psychology, Federal University of Sergipe, São Cristóvão, Sergipe, Brazil.
ORCID: https://orcid.org/0000-0003-0259-1337
E-mail: luanna.psi.ufs@gmail.com

Dandara Palhano – PhD, Researcher, Department of Psychology, Federal University of Sergipe, São Cristóvão, Sergipe, Brazil.
ORCID: https://orcid.org/0000-0003-1625-6871
E-mail: palhanodandara@gmail.com

André Faro – PhD, Faculty, Department of Psychology, Federal University of Sergipe, São Cristóvão, Sergipe, Brazil.
ORCID: https://orcid.org/0000-0002-7348-6297
E-mail: andre.faro.ufs@gmail.com

Contact:

Cidade Univ. Prof. José Aloísio de Campos
Av. Marcelo Deda Chagas, s/n, Bairro Rosa Elze
São Cristóvão-SE, Brasil
CEP: 49107-230
Telefone: (+55) 79 3194-6600