



ORIGINAL ARTICLE

Item-wise validity study of the Impact of Event Scale-Revised: self-reported reactions of hospital personnel exposed to a stressful workplace

Vitor A. Petrilli-Mazon,¹  Marcos O. Carvalho-Alves,^{1,2}  Eurípedes Constantino Miguel,¹ 
Felipe Corchs,^{1,2}  Mariana Curi,³  Yuan-Pang Wang¹ 

¹Departamento de Psiquiatria (LIM-23), Instituto de Psiquiatria, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo (USP), São Paulo, SP, Brazil. ²Programa de Neurociência e Comportamento, Departamento de Psicologia Experimental, Instituto de Psicologia, USP, São Paulo, SP, Brazil. ³Departamento de Matemática Aplicada e Estatística, Instituto de Matemática e Ciências da Computação, USP, São Carlos, SP, Brazil.

Objective: Recent studies have revealed a heterogeneous prevalence and presentation of post-traumatic stress disorder across countries. It is crucial to assess the methodological and item-level measurement factors that contribute to variations in mental disorder rates across cultures. This study aimed to investigate the traumatic experiences of hospital workers by employing item analysis of the Impact of Event Scale-Revised.

Methods: Data were collected from 1,000 employees of a large hospital during the coronavirus disease 2019 pandemic and were analyzed according to item response theory analysis regarding item pool discrimination (*a*) and difficulty (*b*) parameters.

Results: The Impact of Event Scale-Revised items had good discriminative ability (*a*) and covered a range of distress severity (*b*) associated with traumatic experiences. According to our results, the Impact of Event Scale-Revised is a reliable and informative instrument for assessing individuals with moderate to severe trauma-related distress across the spectrum of trauma symptoms. In particular, items 10 ("jumpy or easily startled") and 6 ("I thought about it when I didn't mean to") reflected the post-traumatic stress disorder domains of hyperarousal and intrusion and excelled in discriminating between different levels of post-traumatic distress. Conversely, items related to avoidance and sleep disturbance showed lower discriminative ability.

Conclusions: Item analysis of the Impact of Event Scale-Revised may be used to assess trauma symptoms in the context of a trauma exposure, identifying the most discriminative and informative items for measuring post-traumatic stress disorder in our context. Our findings may help refine the Impact of Event Scale-Revised and facilitate the development of a more effective scale with optimized item parameters.

Keywords: Psychometrics; item response theory; impact of event scale; psychological trauma; hospital personnel

Introduction

Since the inclusion of post-traumatic stress disorder (PTSD) as an official diagnostic category in 1980,¹ research has shifted toward assessing traumatic experiences.² This construct was subsequently expanded to civilians based on trauma-related events, such as civil conflict, natural disasters, epidemics, and urban violence.³ Direct or indirect exposure to trauma can be considered a stressor event,¹ with indirect exposure being particularly relevant in professional contexts, for example among health care workers who treat emergency victims.

However, the nature of human response to traumatic stress has been the subject of debate⁴ due to the fact that many trauma-exposed individuals report partial PTSD associated with significant impairment and help-seeking behaviors. There also seems to be considerable cross-cultural variation in PTSD prevalence. For instance, it has been estimated that 0% of the population of Nigeria⁵ and 6.8% of the population of the United States⁶ meet the diagnostic criteria for PTSD. This has led to concerns about the suitability of measurement tools for post-traumatic distress. It is important to note that not all manifestations of trauma are universally perceived and

Correspondence: Yuan-Pang Wang, Departamento de Psiquiatria (LIM-23), Instituto de Psiquiatria, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, Rua Dr. Ovídio Pires de Campos, 785, CEP 05403-010, São Paulo SP, Brazil.
E-mail: gnap_inbox@hotmail.com
Submitted Aug 11 2024, accepted Nov 04 2024.

How to cite this article: Petrilli-Mazon VA, Carvalho-Alves MO, Miguel EC, Corchs F, Curi M, Wang YP. Item-wise validity study of the Impact of Event Scale-Revised: self-reported reactions of hospital personnel exposed to a stressful workplace. Braz J Psychiatry. 2025;47:e20243882. <http://doi.org/10.47626/1516-4446-2024-3882>

reported in the same manner in different cultures.⁷ Consequently, methodological and item-level measurement factors that influence the observed variation in mental disorder rates across cultures and racial groups must be assessed.⁸

A number of psychometric tools have been proposed to describe the results of experiencing or witnessing a traumatic event.⁹⁻¹⁴ The Impact of Event Scale-Revised (IES-R)¹⁵ is a self-report measure that has gained considerable use for post-traumatic distress symptom assessment. Its items are derived from the reports of individuals who have experienced life changes and symptoms of distress in the past week. These symptoms are classified in three dimensions: intrusion, avoidance, and hyperarousal.

The majority of evidence supporting the reliability, validity, and clinical utility of the IES-R¹⁶⁻¹⁹ is derived from the true score and measurement error framework of classical test theory. This traditional framework assumes that any observed test score consists of true scores and measurement errors.²⁰ However, person parameters (true scores) and item parameters (item difficulty and discrimination) are reliant on the specific test and the examinee's ability, creating internal dependencies that penalize practical application of person-centered and item-based statistics.²¹ Other studies have used the Rasch model with the IES-R^{22,23} and IES-6 (a reduced version of the IES-R),²⁴ as well as a bifactor item response theory (IRT) model for the IES.²⁵ Nevertheless, to the best of our knowledge, no study has analyzed IES-R item performance using 2-parameter logistic IRT models, such as that of Samejima.²⁶

IRT encompasses a family of nuanced and flexible models that describe the relationship between one or more latent traits (θ) and an individual's response to multiple-choice items. Latent traits represent unobserved quantitative attributes or constructs evaluated through observable variables. The item response function posits a non-linear association between the respondents' level of θ and their likelihood of endorsing an item that measures the latent construct. Two common types of parameters are those associated with the person (latent trait θ) and those associated with the item (usually, discrimination and difficulty/threshold parameters). Further details are provided in the specific literature.^{27,28}

In the case of trauma, as exemplified by the IES-R, the underlying construct – trauma-related distress severity – does not directly correspond to specific PTSD symptoms. Rather, it serves as a continuous measure of distress across a spectrum. Therefore, while the IES-R focuses on individual symptoms, the latent trait provides a broader, more comprehensive assessment of distress severity, which enables clinicians to evaluate the full impact of trauma. The clinical significance of identifying discriminative items lies in their capacity to differentiate between individuals who are experiencing disparate levels of trauma-related distress. Highly discriminating items allow more precise identification of symptom severity and clinical significance, thereby facilitating more tailored interventions. A more challenging item (one that requires a higher level of distress to be endorsed) can help

clinicians identify more severe cases that may require urgent attention.

The advantage of IRT over classical test theory^{21,29,30} is most attributable to the latent trait, which should be constant regardless of specific items. Item parameters (a and b) remain steady across respondent groups, ensuring test-independent measurement. Nevertheless, IRT requires a large sample size to obtain stable item parameters and it involves complex computational models. We hypothesized that the overall performance of the IES-R may be enhanced through IRT, possibly extending its application to diverse populations and contexts.

Recent extensive use of mental health measurement tools has led to much interest in item analysis. Based on data from hospital workers during the COVID-19 pandemic, this study analyzed IES-R item performance as an illustration of the IRT approach in psychiatry. Items with poor discrimination were identified and the test's optimal performance range in population sub-groups was determined; implications of this analysis will then be discussed.

Methods

Study design and sampling

This is a validation study of the IES-R using an item analysis framework. Cross-sectional data were collected online between July 1 and August 28, 2020, the first peak period of the COVID-19 pandemic in Brazil. To ensure exposure to a stressful environment, eligible participants had to be active employees at the Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo at the time of data collection. When the baseline survey was conducted, the hospital complex employed 22,056 individuals. An online invitation was distributed to all hospital workers via institutional e-mail, and the study was advertised through social media and posters displayed throughout the hospital. Participants were also encouraged to share the online survey with eligible colleagues. Through non-probabilistic sampling, data were gathered from 1,377 respondents. After eliminating respondents who incompletely filled out the surveys, 1,000 participants were included in the analysis.³¹

Table 1 shows the descriptive information on the sample ($n = 1000$). The participants were 83.9% women, 34.3% were aged between 30 and 40 years, 72.9% had a university-level education or higher, and 74.1% were health care workers. Further details can be found in the primary study.³¹

Assessment tool

Along with complementary scales, the IES-R was self-administered during a period in which the respondents were exposed to infected patients in the hospital, aiming to assess 22 symptoms of post-traumatic distress in the past 7 days. This revised scale is divided into three subscales: eight items on intrusion, eight items on avoidance, and six items on hyperarousal. The items are rated on an ordinal scale, with responses ranging from 0 (not at all) to 4 (extremely). Total scores range

Table 1 Sociodemographic and clinical characteristics of the participants (n=1,000)

Characteristics	
Age group (years)	
18-30	211 (21.1)
30-40	343 (34.3)
40-50	252 (25.2)
> 50	194 (19.4)
Gender	
Female	839 (83.9)
Male	159 (15.9)
Other	2 (0.2)
Marital status	
Unmarried	426 (42.6)
Married	574 (57.4)
Education level	
< University graduate	271 (27.1)
≥ University graduate	729 (72.9)
Living with older adults (> 60 years)	
Yes	233 (23.3)
No	767 (76.7)
Living with children	
Yes	450 (45.0)
No	550 (55.0)
Occupation	
Medical doctor	140 (14.0)
Nurse and nursing assistants	348 (34.8)
Other healthcare professionals [†]	253 (25.3)
Administrative workers [‡]	259 (25.9)
Direct contact with COVID-19 patients (hours per week)	
0	204 (20.4)
1-20	311 (31.1)
21-40	285 (28.5)
> 40	200 (20.0)
Infected with COVID-19 (self-reported)	
Yes	328 (32.8)
No	672 (67.2)
Close family or friend who was hospitalized/died due to COVID-19	
Yes	386 (38.6)
No	614 (61.4)
Changes in daily routine due to the pandemic	
Financial failure	387 (38.7)
Lack of public safety	199 (19.9)
Lack of public transport	297 (29.7)
Lack of medical care	292 (29.2)
Distancing from family and friends	620 (62.0)
Previous psychiatric or psychological treatment	
Yes	280 (28.0)
No	720 (72.0)
Previous disorder diagnosis (self-reported)	
Anxiety	91 (9.1)
Depression	78 (7.8)
PTSD	6 (0.6)
Previous psychotherapy treatment	199 (19.9)
Previous pharmacological treatment	177 (17.7)

Table 1 (continued)

Characteristics	
Psychological or psychiatric treatment after pandemic beginning	138 (13.8)
Protective health actions	
Physical activities	274 (27.4)
Meditative practices	182 (18.2)
Leisure activities/hobbies	320 (32.0)
Religious practices	310 (31.0)
I'm not doing anything in this sense	354 (35.4)
Ethical conflict	119 (11.9)

Data presented as n (%).

PTSD = post-traumatic stress disorder.

[†] Other healthcare professionals: dentists, speech therapists, psychologists, occupational therapists, dieticians, physical therapists, social workers, pharmacists, clinical laboratory technicians, and radiological technologists.

[‡] Administrative workers: receptionists, information technicians, secretaries, and security guards.

from 0 to 88, with classical test theory-based calculated cut-off scores of 9, 26, and 44 indicating mild, moderate, and severe post-traumatic distress symptoms, respectively.³² We considered a cut-off score of 26 as the PTSD threshold,³³ which resulted in 46.8% probable cases of PTSD or a greater than moderate-to-severe level of post-traumatic distress.

We used the Brazilian Portuguese version of the IES-R,³⁴ although evidence of its validity and clinical utility in different samples is still incomplete. The internal consistency of the IES-R was assessed using Cronbach's alpha coefficient, which was found to be $\alpha = 0.96$.

Statistical analysis

We initially examined endorsement patterns for the 22 IES-R items by calculating the proportion of respondents who selected each score option, which ranged from 0 to 4. We calculated the conditional reliability of the IES-R, which considers both scale information and errors. After calculating marginal reliability as an overall measure of the scale's reliability, we plotted the conditional reliability curve along the θ dimension to visualize the variation in scale reliability according to the underlying trait.

Prior to running the 2-parameter logistic graded response model,²⁶ we examined the scale's dimensionality, local independence, and monotonicity. Dimensionality was determined through exploratory factor analysis in R using the psych package. The Bartlett test of sphericity yielded a χ^2 value of 17,527.55 ($p < 0.001$), and the Kaiser-Meyer-Olkin measure of sampling adequacy reached 0.98. The results of the Kaiser eigenvalues > 1 criterion and Cattell's scree plot inspection indicated a one-factor solution. The assumption of local independence was verified by examining between-item correlations, which were controlled for a single factor.

We then estimated the graded response model via the full-information maximum likelihood fitting function of the mirt package in R. Four indexes were used to assess the model fit: the standardized root mean square residual

(SRMSR), the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). The recommended cut-off criteria for each index are $SRMSR \leq 0.08$, $CFI/TLI \geq 0.95$,³⁵ and $RMSEA \leq 0.10$.³⁶ The estimated values in the present study were $SRMSR = 0.067$, $CFI = 0.96$, $TLI = 0.96$, all of which favored using the graded response model, and $RMSEA = 0.1$, which is marginally acceptable. We assessed item fit by using the $S-\chi^2$ index to calculate the RMSEA value. All IES-R items presented an $RMSEA \leq 0.06$, indicating adequate fitness.

The 2-parameter logistic graded response model was used to estimate the relationship between IES-R responses and the underlying latent trait θ (the severity of post-traumatic distress). The graded response model is suitable for items with more than two ordered response categories and extends the 2-parameter logistic model commonly used for dichotomous items. It allows grading responses across multiple categories, such as the 0-4 point range in the IES-R. The model is based on cumulative logits and captures the ordered nature of response categories, representing the degree of achievement, level of agreement, or frequency of a statement.

In this model, the discrimination parameter (a) differentiates individuals with varying θ levels. Graphically, the slope of the item response function at the inflection point indicates the item's discrimination, with a steeper slope indicating higher discriminatory power. The values can be interpreted as very low (0.01 to 0.24), low (0.25 to 0.64), moderate (0.65 to 1.34), high (1.35 to 1.69), or very high (> 1.7).³⁷ However, category threshold parameters (b) indicate the trait level where a respondent has a 0.50

probability of responding above the threshold. Higher threshold levels require more severe trait levels to endorse an item.

In addition to examining the item discrimination (a) and item difficulty/threshold parameters (b), we plotted category response curves, item information curves, and the test information curve. All analyses and graphics were made in R 4.0.4 (<https://www.r-project.org>). The R codification can be provided upon request.

Ethics statement

The survey included an informed consent form that provided participants with information about the study design, objectives, and procedures. The study was approved by the institutional research ethics committee (decision 30710620.2.0000.0068).

Results

Table 2 presents the distribution of responses for individual IES-R items. It provides an overview of the proportion of participants who endorsed each category for each item, indicating the frequency of each response type across the scale. In general, the proportion of each response category decreases as the frequency increases, from 0 (not at all) to 4 (extremely). The items with the most responses at the lower end of the continuum (0 – not at all) were 19 “Reminders of it caused me to have physical reactions, such as sweating, trouble breathing, nausea, or a pounding heart” and 20 “I had dreams about it,” with frequencies of 52.5 and 51.2%, respectively.

Table 2 Category responses for the Impact of Event Scale-Revised items (n=1,000)

Items	Category responses				
	0	1	2	3	4
1. Any reminder brought back feelings about it	297 (29.7)	337 (33.7)	166 (16.6)	148 (14.8)	52 (5.2)
2. I had trouble staying asleep	304 (30.4)	264 (26.4)	144 (14.4)	193 (19.3)	95 (9.5)
3. Other things kept making me think about it	215 (21.5)	382 (38.2)	159 (15.9)	185 (18.5)	59 (5.9)
4. I felt irritable and angry	277 (27.7)	304 (30.4)	157 (15.7)	169 (16.9)	93 (9.3)
5. I avoided letting myself get upset when I thought about it or was reminded of it	243 (24.3)	380 (38.0)	185 (18.5)	150 (15.0)	42 (4.2)
6. I thought about it when I didn't mean to	276 (27.6)	323 (32.3)	148 (14.8)	172 (17.2)	81 (8.1)
7. I felt as if it hadn't happened or wasn't real	480 (48.0)	260 (26.0)	116 (11.6)	106 (10.6)	38 (3.8)
8. I stayed away from reminders of it	390 (39.0)	304 (30.4)	163 (16.3)	103 (10.3)	40 (4.0)
9. Pictures about it popped into my mind	391 (39.1)	285 (28.5)	128 (12.8)	146 (14.6)	50 (5.0)
10. I was jumpy and easily startled	449 (44.9)	241 (24.1)	121 (12.1)	143 (14.3)	46 (4.6)
11. I tried not to think about it	230 (23.0)	326 (32.6)	170 (17.0)	208 (20.8)	66 (6.6)
12. I was aware that I still had a lot of feeling about it, but I didn't deal with them	298 (29.8)	360 (36.0)	169 (16.9)	136 (13.6)	37 (3.7)
13. My feelings about it were kind of numb	471 (47.1)	287 (28.7)	139 (13.9)	77 (7.7)	26 (2.6)
14. I found myself acting or feeling like I was back at that time	484 (48.4)	271 (27.1)	112 (11.2)	104 (10.4)	29 (2.9)
15. I had trouble falling asleep	304 (30.4)	259 (25.9)	123 (12.3)	184 (18.4)	130 (13.0)
16. I had waves of strong feelings about it	383 (38.3)	239 (23.9)	136 (13.6)	152 (15.2)	90 (9.0)
17. I tried to remove it from my memory	322 (32.2)	294 (29.4)	126 (12.6)	193 (19.3)	65 (6.5)
18. I had trouble concentrating	276 (27.6)	319 (31.9)	161 (16.1)	160 (16.0)	84 (8.4)
19. Reminders of it caused me to have physical reactions, such as sweating, trouble breathing, nausea, or a pounding heart	525 (52.5)	189 (18.9)	102 (10.2)	115 (11.5)	69 (6.9)
20. I had dreams about it	512 (51.2)	229 (22.9)	104 (10.4)	107 (10.7)	48 (4.8)
21. I felt watchful and on-guard	348 (34.8)	319 (31.9)	140 (14.0)	137 (13.7)	56 (5.6)
22. I tried not to talk about it	372 (37.2)	280 (28.0)	148 (14.8)	138 (13.8)	62 (6.2)

Data presented as n (%).

0 = not at all; 1 = a little bit; 2 = moderately; 3 = quite a bit; 4 = extremely.

In contrast, items 15 “I had trouble falling asleep” and 2 “I had trouble staying asleep” had the most responses at the upper end, with frequencies of 13% and 9.5%, respectively. This variation suggests heterogeneous performance across IES-R items.

The IES-R's conditional reliability varied slightly along the θ dimension. The best range estimate for optimal reliability was broad, being set from -1 to +2 (Figure 1). The plot suggests good reliability (> 0.9) along the latent trait θ for post-traumatic distress. The marginal reliability was 0.95, indicating that the scale could assess a homogenous construct across θ .

Table 3 shows the discrimination (a) and difficulty/threshold (b) parameters of the IES-R items. Item

analysis generated a parameters ranging from 1.72 to 4.07, which are considered very high.³⁷ The most discriminating items were 10 “I was jumpy or easily startled” ($a = 4.07$) and 6 “I thought about it when I didn't mean to” ($a = 4.01$). They reflected two core DSM criteria for PTSD diagnosis: the hyperarousal (item 10) and intrusion (item 6) dimensions. Item 14 “I found myself acting or feeling like I was back at that time” (related to flashbacks) was also discriminative, reflecting the intrusion dimension ($a = 3.47$). The least discriminating items were 8 “I stayed away from reminders of it” ($a = 1.72$), 7 “I felt as if it hadn't happened or wasn't real” ($a = 1.87$), and 20 “I had dreams about it” ($a = 1.89$). Two of these items (8, 7) reflected the avoidance domain, while item

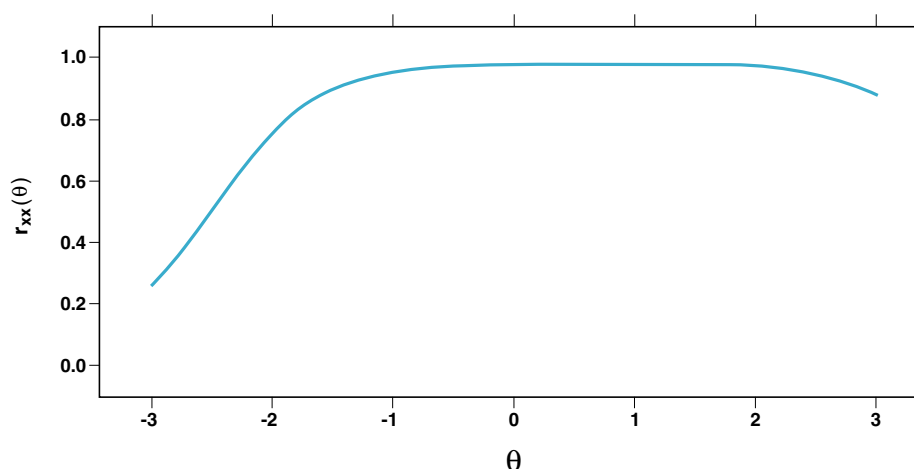


Figure 1 Conditional reliability curve of the Impact of Event Scale-Revised (n=1,000). Marginal reliability = 0.95.

Table 3 Impact of Event Scale-Revised item parameters estimated through the graded response model

IES-R items	Subscale	Discrimination	Threshold			
		a	$b1$	$b2$	$b3$	$b4$
1. Any reminder brought back feelings about it	I	2.81	-0.57	0.42	0.98	1.84
2. I had trouble staying asleep	I	1.90	-0.66	0.26	0.77	1.74
3. Other things kept making me think about it	I	3.45	-0.83	0.3	0.78	1.68
4. I felt irritable and angry	H	2.34	-0.69	0.27	0.8	1.61
5. I avoided letting myself get upset when I thought about it or was reminded of it	A	2.15	-0.85	0.42	1.08	2.15
6. I thought about it when I didn't mean to	I	4.01	-0.57	0.31	0.74	1.46
7. I felt as if it hadn't happened or wasn't real	A	1.87	-0.06	0.8	1.36	2.36
8. I stayed away from reminders of it	A	1.72	-0.35	0.72	1.47	2.45
9. Pictures about it popped into my mind	I	3.81	-0.23	0.51	0.94	1.71
10. I was jumpy and easily startled	H	4.07	-0.08	0.55	0.96	1.73
11. I tried not to think about it	A	2.17	-0.88	0.26	0.8	1.88
12. I was aware that I still had a lot of feeling about it, but I didn't deal with them	A	2.24	-0.61	0.55	1.17	2.17
13. My feelings about it were kind of numb	A	2.21	-0.06	0.85	1.54	2.39
14. I found myself acting or feeling like I was back at that time	I	3.47	0	0.75	1.21	2.01
15. I had trouble falling asleep	H	2.07	-0.64	0.22	0.64	1.44
16. I had waves of strong feelings about it	I	3.81	-0.26	0.37	0.78	1.43
17. I tried to remove it from my memory	A	2.58	-0.48	0.39	0.78	1.77
18. I had trouble concentrating	H	2.42	-0.69	0.31	0.85	1.66
19. Reminders of it caused me to have physical reactions, such as sweating, trouble breathing, nausea, or a pounding heart	H	2.86	0.11	0.65	1.03	1.68
20. I had dreams about it	I	1.89	0.05	0.83	1.32	2.20
21. I felt watchful and on-guard	H	2.85	-0.4	0.51	1.01	1.81
22. I tried not to talk about it	A	1.98	-0.38	0.54	1.11	2.00

A = avoidance; H = hyperarousal; I = intrusion.

20, which was related to nightmares, reflected the intrusion domain. Similarly, two items on sleep disturbance (2 and 15) also showed lower discriminative ability.

The difficulty/threshold (b) parameters ranged along the latent trait θ – the severity of post-traumatic distress – as follows: between -0.88 and 0.11 for the first parameter (b_1), 0.22 and 0.85 for the second parameter (b_2), 0.64 and 1.54 for the third parameter (b_3), and 1.43 and 2.45 for the fourth parameter (b_4). Figure 2 shows category response curves derived from the IRT model, which represent the endorsement probabilities for each IES-R response category. The curves correspond to “not at all,” “a little bit,” “moderately,” “quite a bit,” and “extremely,” illustrating the relationship between θ and the response likelihood. As θ increases, the likelihood of endorsing a category initially rises, then decreases as responses shift to the next higher category. These estimates indicated that the IES-R could appropriately measure people with a moderate-to-severe level of trauma-related distress along the θ dimension.

Scale performance is conveyed through the test information curve and standard error estimates (Figure 3). This curve illustrates the level of post-traumatic distress at which the IES-R provides the highest amount of information, which is observed within θ values of approximately -0.5 to 2. Individuals whose distress levels are average or high benefit the most from the IES-R in terms of reliability and meaningful information. Conversely, for θ values below -1 and above 2, the IES-R provides significantly less information. Consequently, the tool may not be suitable for assessing very low or very severe levels of post-traumatic distress.

Discussion

The present study illustrates the item analysis paradigm in trauma research. Notably, the cutting-edge IRT has yielded a comprehensive perspective of self-reported psychopathology resulting from traumatic experiences. Overall, the scale demonstrated good reliability, and all

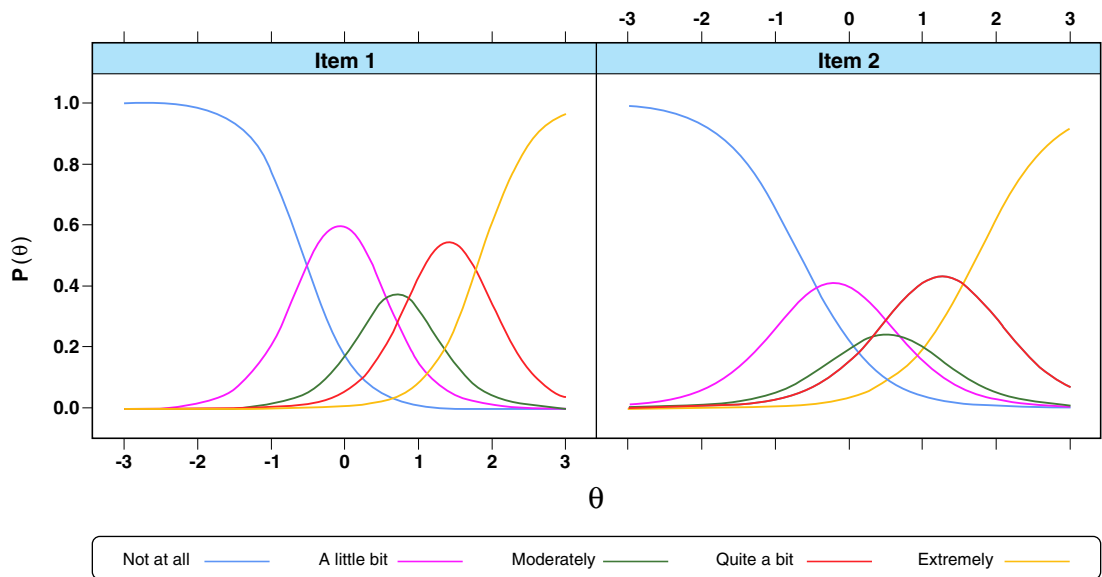


Figure 2 Category response curves of items 1 and 2.

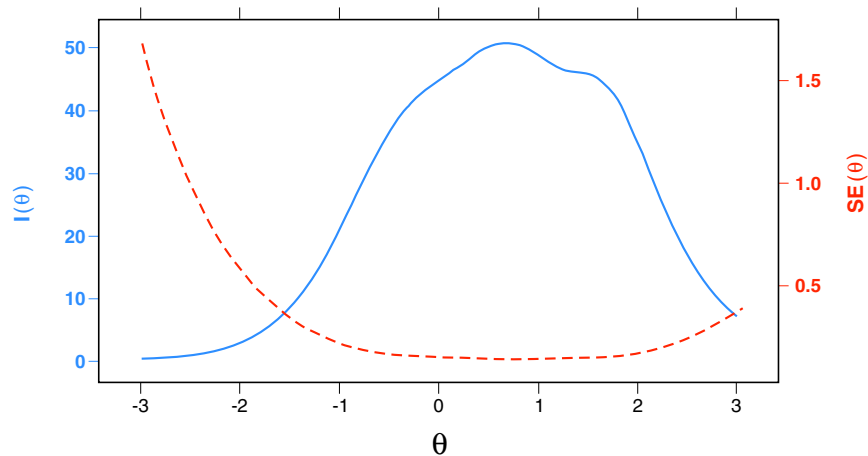


Figure 3 Test information (I) curve and standard error (SE) estimates.

items had high discrimination parameters, with thresholds indicating a right-skewed distribution of construct severity (θ). Hyperarousal and intrusion were identified as central dimensions of traumatic experiences. The IES-R proved more informative for individuals with average or high levels of distress (around $\theta = 1$), which raises concerns about its applicability in non-clinical populations. In the future, the scalable application of our item-based approach in psychiatry could improve measurement precision, scale reliability, and construct validity. This represents a significant advancement in the field. Moreover, our findings could help refine the widely-used IES-R by identifying the most relevant items.

All items in the IES-R questionnaire had high discrimination, indicating their ability to distinguish individuals with heterogeneous degrees of distress. Symptoms from the intrusion and hyperarousal domains emerged as core dimensions of traumatic experience, and should thus be considered in the next revision of PTSD criteria. Our findings are in line with those of Natesan Batley et al.,³⁸ who employed IRT when applying the PTSD Checklist for DSM-5 to Indian and American samples. Except for traumatic amnesia, all items were highly discriminative for PTSD.

Unlike previous research with English-speaking populations,³⁹ we found that items assessing the hyperarousal dimension were among the most discriminative. Items related to the avoidance dimension were the least discriminative but the most difficult to endorse, unlike the results of previous studies.³⁹⁻⁴² In line with the literature, however, the item on the intrusive experience of flashbacks had high discrimination.^{8,41,43} Notably, the item assessing nightmares had low discrimination, which challenges the prevailing theoretical and empirical perspective on the role of nightmares as a core feature of PTSD.^{40,41,43,44} Our findings among Portuguese-speaking respondents indicate that the IES-R item pool may not have captured the same subjective experience of post-traumatic distress as in English-speaking respondents regarding discrimination and severity. Consequently, cross-cultural studies should compare individuals from different linguistic and cultural backgrounds.

Of note, the IES-R is more effective at providing information about individuals with moderate to severe distress levels, since the threshold parameters (b) for the majority of items ranged from medium to high. In general, the items were difficult to endorse, which implies that individuals with milder symptoms were less likely to endorse the items on the scale. For example, item 14, which pertains to flashbacks ("I found myself acting or feeling like I was back at that time"), was highly discriminative and indicative of severity.^{41,43} Conversely, sleep disturbances (intrusion and hyperarousal) were readily endorsed, which is consistent with their pervasive association with psychiatric disorders in general.⁴⁵⁻⁴⁷ It is important to recognize that sleep-related symptoms may serve as non-specific indicators of distress severity, since they are often present at an early stage in psychiatric disorders. In some cases, they may even precede the full disorder.

Thus, our IRT analysis provided numerous indicators regarding the boundaries at which the scale obtains

maximum information about post-traumatic distress symptoms. The test information curve showed the maximized threshold range for scale performance. In the present study, the most informative θ range fell between 0 and 1.5, which suggests that the IES-R is suitable for identifying individuals with moderate-to-severe levels of post-traumatic distress. However, the IES-R's high marginal reliability suggested the possibility of redundant questions. Consequently, it may not be cost-effective to use this tool for epidemiological case-finding of subclinical or milder PTSD in non-clinical populations.³⁸

Our results also corroborate a multi-symptomatic presentation of post-traumatic reactions, in which hyperarousal and intrusion dominate the underlying construct. Regarding the IES-R's suitability for assessing post-traumatic distress, we identified the severity level and threshold range at which the tool is most effective. The analysis also revealed variations in discrimination and threshold parameters among Brazilian respondents, suggesting potential cross-cultural differences. Therefore, IRT is a valuable approach for post-traumatic distress assessment in hospital workers, which could facilitate the development of tailored interventions and support strategies.

Certain limitations should be considered when interpreting our findings. First, although the sample size was sufficiently large, it was not representative of the entire population of workers at our institution. Although there could have been participation bias, IRT requires a large sample size, rather than a representative sample, for data modeling. Second, the possibility of response bias cannot be ruled out. The data were collected from hospital workers during a period of high workload and stress, which may have influenced the accuracy and completeness of their responses. Third, the sample's lack of representativeness (i.e., including only hospital workers during the pandemic period) may limit the generalizability of the findings to other populations. Although trauma research often relies on non-probability samples,⁴⁸⁻⁵⁰ future studies with diverse groups could increase the scale's applicability. Fourth, the cross-sectional design precluded longitudinal data on the effects of trauma, which limits the ability to track symptom progression over time. Although we could not address this in the present study, our research group has recently published data on longitudinal responses to the IES-R (among other scales), providing additional insight into the progression of trauma symptoms.⁵¹ Moreover, the computational complexity of IRT and its requirement for large sample sizes may restrict more widespread clinical use. However, the benefits of IRT are considerable, and we hope to encourage broader exploration of this methodology in diverse settings, perhaps through more user-friendly applications, to narrow the gap between research and clinical practice. Finally, the prevalence of preexisting or persistent mental health symptoms may have been underestimated among the workers due to overlap with new-onset symptoms resulting from traumatic experiences with patients. Symptom onset, chronicity, and persistence were not fully investigated, which may have resulted in incomplete data. Although 28% of the

participants reported a history of mental disorders and treatment, the symptom rates in our study were notably higher than those found during the pre-pandemic period. Our study has demonstrated scale validity at an item level or the scale's latent trait. Future studies should report additional types of validity, including structural, criterion-based, and construct validity.

Item analysis of the IES-R among hospital workers during the pandemic revealed how symptoms manifested along the latent construct of post-traumatic distress. Our analysis suggested that the IES-R can identify individuals with higher levels of trauma symptoms. Understanding how people react after exposure to a traumatic event could enhance the applicability of scales and contribute to the ongoing debate about the current conceptualization of PTSD. Because IRT is already a thriving practice in different health sciences, researchers should be aware of its applicability in psychiatry.

Acknowledgements

The authors would like to thank the following colleagues for their involvement during the survey planning stage: Prof. André R. Brunoni, Dr. André Malbergier, Dr. Pedro Fukuti, Prof. Guilherme V. Polanczyk.

Disclosure

The authors report no conflicts of interest.

Author contributions

VAPM: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing.

MOCA: Data Curation, Writing – review & editing.

ECM: Writing – review & editing.

FC: Writing – review & editing.

MC: Conceptualization, Formal analysis, Methodology, Writing – review & editing.

YPW: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing.

All authors have read and approved the final version to be published.

Handling Editor: Raffael Massuda.

References

- 1 American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III). Washington: American Psychiatric Publishing; 1980.
- 2 Gelkopf M, Mazor Y, Roe D. A systematic review of patient-reported outcome measurement (PROM) and provider assessment in mental health: Goals, implementation, setting, measurement characteristics and barriers. *Int J Qual Health Care*. 2021;34:ii13-27.
- 3 Coêlho BM, Santana GL, Dantas HS, Viana MC, Andrade LH, Wang YP. Correlates and prevalence of post-traumatic stress disorders in the São Paulo metropolitan area, Brazil. *J Psychiatr Res*. 2022;156:168-76.
- 4 Ruscio AM, Ruscio J, Keane TM. The latent structure of posttraumatic stress disorder: a taxometric investigation of reactions to extreme stress. *J Abnorm Psychol*. 2002;111:290-301.
- 5 Gureje O, Lasebikan VO, Kola L, Makanjuola VA. Lifetime and 12-month prevalence of mental disorders in the Nigerian Survey of Mental Health and Well-Being. *Br J Psychiatry*. 2006;188:465-71.
- 6 Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62:593-602.
- 7 Hinton DE, Lewis-Fernández R. The cross-cultural validity of post-traumatic stress disorder: implications for DSM-5. *Depress Anxiety*. 2011;28:783-801.
- 8 Mekawi Y, Silverstein MW, Walker A, Ishiekwe M, Carter S, Michopoulos V, et al. Examining the psychometric properties of the PCL-5 in a black community sample using item response theory. *J Anxiety Disord*. 2022;87:102555.
- 9 Davidson JR, Book SW, Colket JT, Tupler LA, Roth S, David D, et al. Assessment of a new self-rating scale for post-traumatic stress disorder. *Psychol Med*. 1997;27:153-60.
- 10 Weathers F, Litz B, Herman D, Huska JA, Keane TM. The PTSD checklist: reliability, validity, and diagnostic utility. San Antonio: 9th Annual Meeting of the International Society of Traumatic Stress Studies; 1993.
- 11 Foa E, Cashman L, Jaycox L, Perry K. The validation of a self-report measure of posttraumatic stress disorder: the Posttraumatic Diagnostic Scale. *Psychol Assess*. 1997;9:445-51.
- 12 Keane TM, Caddell JM, Taylor KL. Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: Three studies in reliability and validity. *J Consult Clin Psychol*. 1988;56:85-90.
- 13 Hammarberg M. Penn Inventory for Posttraumatic Stress Disorder: psychometric properties. *Psychol Assess*. 1992;4:67-76.
- 14 Keane TM, Malloy PF, Fairbank JA. Empirical development of an MMPI subscale for the assessment of combat-related posttraumatic stress disorder. *J Consult Clin Psychol*. 1984;52:888-91.
- 15 Weiss DS, Marmar CR. The Impact of Event Scale – Revised. In: Wilson JP, Keane TM, editors. *Assessing psychological trauma and PTSD*. New York: Guilford Press; 1997. p. 399-411.
- 16 Adkins JW, Weathers FW, McDevitt-Murphy M, Daniels JB. Psychometric properties of seven self-report measures of posttraumatic stress disorder in college students with mixed civilian trauma exposure. *J Anxiety Disord*. 2008;22:1393-402.
- 17 Beck JG, Grant DM, Read JP, Clapp JD, Coffey SF, Miller LM, et al. The Impact of Event Scale-Revised: psychometric properties in a sample of motor vehicle accident survivors. *J Anxiety Disord*. 2008;22:187-98.
- 18 Rash CJ, Coffey SF, Baschnagel JS, Drobos DJ, Saladin ME. Psychometric properties of the IES-R in traumatized substance dependent individuals with and without PTSD. *Addict Behav*. 2008;33:1039-47.
- 19 Mystakidou K, Tsilika E, Parpa E, Galanos A, Vlahos L. Psychometric properties of the Impact of Event Scale in Greek cancer patients. *J Pain Symptom Manage*. 2007;33:454-61.
- 20 Guilleks H. Theory of mental tests. New York: Wiley; 1950.
- 21 Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. *Educ Meas*. 1993;12:38-47.
- 22 Aljaberi MA, Lee KH, Alareqe NA, Qasem MA, Alsalahi A, Abdallah AM, et al. Rasch modeling and multilevel confirmatory factor analysis for the usability of the Impact of Event Scale-Revised (IES-R) during the COVID-19 pandemic. *Healthcare (Basel)*. 2022;10:1858.
- 23 Lee KW, Lee H. Validation of the Korean version of the Impact of Event Scale-Revised in disaster victims using the Rasch model. *Asia Pac J Soc Work Dev*. 2024;1:15.
- 24 Jeong J, Kim AR, Hilton C, Hong I. Impact of Event Scale-6 (IES-6) for U.S. adults who experienced the COVID-19 pandemic. *BMC Psychiatry*. 2022;22:490.
- 25 Salsman JM, Schalet BD, Andrykowski MA, Cella D. The impact of events scale: a comparison of frequency versus severity approaches to measuring cancer-specific distress. *Psychooncology*. 2015;24:1738-45.
- 26 Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychom Monogr*. 1969;34:386-415.
- 27 Reise SP, Ainsworth AT, Haviland MG. Item response theory: Fundamentals, applications, and promise in psychological research. *Curr Dir Psychol Sci*. 2005;14:95-101.
- 28 Reeve BB. Item response theory modeling in health outcomes measurement. *Expert Rev Pharmacoecon Outcomes Res*. 2003;3:131-45.

- 29 Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of Item Response Theory. Thousand Oaks: Sage Publications; 1991.
- 30 Embretson SE, Reise SP. Item response theory. New York: Psychology Press; 2013.
- 31 Carvalho-Alves MO, Petrilli-Mazon VA, Brunoni AR, Malbergier A, Fukuti P, Polanczyk GV, et al. Dimensions of emotional distress among Brazilian workers in a COVID-19 reference hospital: A factor analytical study. *World J Psychiatry*. 2022;12:843-59.
- 32 Wu KK, Chan KS. The development of the Chinese version of Impact of Event Scale--Revised (CIES-R). *Soc Psychiatry Psychiatr Epidemiol*. 2003;38:94-8.
- 33 Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. *JAMA Netw Open*. 2020;3:e203976.
- 34 Caiuby AV, Lacerda SS, Quintana MI, Torii TS, Andreoli SB. [Cross-cultural adaptation of the Brazilian version of the Impact of Events Scale-Revised (IES-R)]. *Cad Saude Publica*. 2012;28:597-603.
- 35 Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1-55.
- 36 Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociol Methods Res*. 1992;21:230-58.
- 37 Baker F. The basis of item response theory. College Park: ERIC Clearinghouse on Assessment and Evaluation; 2001.
- 38 Batley PN, Contractor AA, Compton SE, Sharma R, Dranger P. Evaluation of the Posttraumatic Stress Disorder Checklist for DSM-5 across United States and India samples: An item response theory analysis using the graded response model. *Psychol Trauma*. 2024;16:676-85.
- 39 Silverstein MW, Petri JM, Kramer LB, Weathers FW. An item response theory analysis of the PTSD checklist for DSM-5: Implications for DSM-5 and ICD-11. *J Anxiety Disord*. 2020;70:102190.
- 40 King DW, King LA, Leskin GA, Weathers FW. Confirmatory factor analysis of the clinician-administered PTSD scale: Evidence for the dimensionality of posttraumatic stress disorder *Psychological Assessment*. 1998;10:90-6.
- 41 Fissette CL, Snyder DK, Balderrama-Durbin C, Balsis S, Cigrang J, Talcott GW, et al. Assessing posttraumatic stress in military service members: improving efficiency and accuracy. *Psychol Assess*. 2014;26:1-7.
- 42 Orlando M, Marshall GN. Differential item functioning in a Spanish translation of the PTSD checklist: Detection and evaluation of impact. *Psychol Assess*. 2002;14:50-9.
- 43 Palm KM, Strong DR, MacPherson L. Evaluating symptom expression as a function of a posttraumatic stress disorder severity. *J Anxiety Disord*. 2009;23:27-37.
- 44 Brewin CR, Lanius RA, Novac A, Schnyder U, Galea S. Reformulating PTSD for DSM-V: Life after Criterion A. *J Trauma Stress*. 2009;22:366-73.
- 45 Breslau N, Roth T, Rosenthal L, Andreski P. Sleep disturbance and psychiatric disorders: a longitudinal epidemiological study of young adults. *Biol Psychiatry*. 39:411-8.
- 46 Krystal AD. Psychiatric disorders and sleep. *Neurol Clin*. 2012;30:1389-1413.
- 47 Freeman D, Sheaves B, Waite F, Harvey AG, Harrison PJ. Sleep disturbance and psychiatric disorders. *Lancet Psychiatry*. 2020;7:628-37.
- 48 Andrews B, Brewin CR, Rose S, Kirk M. Predicting PTSD symptoms in victims of violent crime: the role of shame, anger, and childhood abuse. *J Abnorm Psychol*. 2000;109:69-73.
- 49 Knipscheer JW, Sleijpen M, Mooren T, Ter Heide FJ, van der Aa N. Trauma exposure and refugee status as predictors of mental health outcomes in treatment-seeking refugees. *BJPsych Bull*. 2015;39:178-82.
- 50 Weathers FW, Bovin MJ, Lee DJ, Sloan DM, Schnurr PP, Kaloupek DG, et al. The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychol Assess*. 2018;30:383-95.
- 51 Carvalho-Alves MO, Petrilli-Mazon VA, Zuccolo PF, Fatori D, Rocha FMM, Brunoni AR, et al. Trajectories of anxiety, depression, and posttraumatic stress among healthcare workers during the COVID-19 pandemic: one-year monthly follow-up. *Braz J Psychiatry*. 2024 Sep 22. doi: 10.47626/1516-4446-2024-3720. Epub ahead of print.