REVIEW

Item Response Theory Analysis of the Fear of COVID-19 Scale (FCV-19S): A Systematic Review

Zainab Alimoradi [1],*, Chung-Ying Lin [2–5],*, Irfan Ullah [6], Mark D Griffiths [7], Amir H Pakpour [8]

[1]Social Determinants of Health Research Center, Research Institute for Prevention of Non-Communicable Diseases, Qazvin University of Medical Sciences, Qazvin, Iran; [2]Institute of Allied Health Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [3]Biostatistics Consulting Center, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [4]Department of Public Health, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [5]Department of Occupational Therapy, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [6]Kabir Medical College, Gandhara University, Peshawar, Pakistan; [7]International Gaming Research Unit, Psychology Department, Nottingham Trent University, Nottingham, UK; [8]Department of Nursing, School of Health and Welfare, Jönköping University, Jönköping, Sweden

*These authors contributed equally to this work

Correspondence: Irfan Ullah, Kabir Medical College, Gandhara University, Peshawar, Pakistan, Email irfanullahecp2@gmail.com; Amir H Pakpour, Department of Nursing, School of Health and Welfare, Jönköping University, Barnarpsgatan 39, Jönköping, 55111, Sweden, Email amir.pakpour@ju.se

Background: The COVID-19 pandemic is still ongoing and is not yet under control. Evidence regarding the impacts of COVID-19 on psychological distress has been widely reported worldwide, and one of the primary concerns regarding psychological distress is fear (ie, fear of COVID-19). Therefore, having a robust instrument for assessing fear of COVID-19 is important. The present systematic review aimed to synthesize the psychometric evidence evaluated using item response theory (IRT) on the Fear of COVID-19 Scale (FCV-19S).

Methods: Utilizing the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, four academic databases (Scopus, PubMed Central, ProQuest, and ISI Web of Knowledge) were used to search target papers. Keywords used for search were "Fear of COVID-19 Scale" and its abbreviation (ie, "FCV-19S") and IRT-related terms. The COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist was then applied to evaluate the methodological quality of the reviewed papers. Moreover, psychometric properties using IRT methods were synthesized using a qualitative method.

Results: The initial search resulted in 552 papers (73 duplicates) and 479 were screened based on their titles and abstracts. Finally, 16 papers were included for review regarding their methodological quality (via COSMIN) to synthesize the psychometric evidence for FCV-19S. The 16 papers included 21 countries with 16 language versions of FCV-19S.

Conclusion: All the psychometric evidence indicated that the seven items in the FCV-19S fit with the concept of fear. The FCV-19S is a strong and valid instrument for assessing fear across different languages. The seven items in the FCV-19S appear to be unidimensional in assessing fear, which indicates that all items are necessary in the FCV-19S.

Keywords: COVID-19, fear, item response theory, psychometrics, Rasch, review

Introduction

The coronavirus disease 2019 (COVID-19) epidemic has affected (and continues to affect) all nations, continents, races, and socioeconomic groups. Early responses to the pandemic included quarantining entire communities, closing schools and colleges, implementing social isolation, and staying at home, which suddenly changed people's daily lives.[1,2] The sudden onset of COVID-19 (ie, its epidemic nature with very high rates of infection and relatively high mortality) caused individuals to naturally become anxious and afraid of becoming infected with COVID-19.[3,4]

One of the characteristics of infectious diseases is fear, which is directly related to the rate of transmission and its medium (fast and invisible) as well as its complications and mortality. This leads to other psychosocial challenges such as stigma, discrimination, and loss.[5] The threats of new COVID-19 variants are also known to have effects on psychological responses.[6–8] Such psychological reactions, especially anxiety and fear, are important factors for individuals' consequent

behaviors.[9,10] Low levels of anxiety may lead to rational adherence to preventative behaviors (eg, hand washing) and high levels of anxiety may lead to socially destructive behaviors (eg, fear of shopping).[11] High levels of fear can cause individuals not thinking clearly and rationally when reacting to COVID-19, resulting in increased damage from the disease.[12]

The psychological effects of the COVID-19 epidemic have demonstrated the need for healthcare providers having validated tools to assess individuals' psychological responses specifically to COVID-19 during the epidemic.[11] However, healthcare providers should not only care for COVID-19-related symptoms (including psychological reactions). Instead, they still need to pay attention to the "old diseases".[13,14] Given the importance of the issue of COVID-19 fear, the Fear of COVID-19 Scale (FCV-19S) has been developed to help healthcare providers to design and implement appropriate interventions to reduce fear.[12] The FCV-19S was developed by a panel of experts after an extensive literature review. Moreover, the FCV-19S items assessed using classical test theory and the Rasch model were shown to be satisfactory among the general public and other populations.[12,15–17]

The total score of the FCV-19S can be used to help diagnose, evaluate, and monitor the severity of COVID-19 fear-related anxiety. Using total score is a fast and simple application that quickly shows the severity of individuals' COVID-19 fears.[12] However, the simple summation of raw scores ignores the difference between items and the information that the response pattern can provide. Therefore, it may lead to incorrect estimation.[18] Items vary in one measurement in different ways. First, some may be more difficult than others. For example, for most individuals, repeating a noun will be more difficult than remembering a phrase or a list of words. Second, some cases may be more sensitive to the early stages of the disease and others to the later stages of the disease. Third, items vary in sensitivity to clinical changes. Finally, some items may offer no significant variation in measurement. These can be eliminated to reduce the temporal burden on assessments.[19]

Additionally, the same total score can be obtained through different response patterns. For example, two individuals who received a score of 20 on the same scale may have answered in completely different patterns. Similarly, an individual who achieves the same total score before and after treatment will be considered as not changing status. However, the intervention may have changed the pattern of answering questions.[19] Therefore, there is a need to examine the pattern of response to items, of which can be overcome using the statistical method "Item Response Theory" (IRT).[20]

IRT is a statistical method based on the probability of an individual achieving a specific score in a test that the result of that individual's status lies in the structure.[21] As this situation changes, so does the probability of achieving a specific score from the individual, and the accuracy of the measurement varies with the level of the situation.[22] IRT can show two useful features of each test, namely difficulty and discrimination (or differentiation).[23] Discrimination is an indicator that shows how an item can differentiate between individuals with different intensities. The more distinct the item, the steeper the slope, the better they can differentiate between individuals within the range of latent features.[24] The next feature is the difficulty of the items which vary according to the diagnosis or by country/region or different translations. Awareness of the order of difficulty helps physicians to design and implement their assessments according to the severity of the situation.[24]

IRT can check the sensitivity of items on a scale via the item characteristic curve (ICC). Items with a higher ICC slope are where the item is most differentiated, distinguishing well between different degrees of disorder and making them more sensitive to change. Determining the differentiation of items can indicate which items are most likely to best reflect changes in status.[25] Examining the case curves in relation to each other provides useful information about the measurement scope of a tool. The IRT can also identify key items that provide valuable information or whether the items are on an additional scale (ie, items with similar ICCs).[22]

Using IRT techniques to assess COVID-19 fear can have far-reaching implications for physicians and researchers, leading to advances in screening evaluation and diagnosis, charting disease, and measuring changes in disease progression and response to treatment. IRT has been used to analyze clinical practice in several different contexts including schizophrenia,[26] depression,[27] attachment,[28] and quality of life.[29] Given that many studies have translated and examined the psychometric properties of the FCV-19S, IRT methods have also been applied. However, no synthesized evidence regarding the IRT-assessed properties of the FCV-19S has been published. Therefore, the aim of the present systematic review was to synthesize the psychometric evidence of the FCV-19S on its IRT findings, including these studies'

methodological quality, FCV-19S item difficulties, FCV-19S item discriminations, and differential item functioning (DIF) across age and gender.

Methods
Design and Protocol Registration
The present systematic review was designed to assess the measurement properties evaluated using IRT for the FCV-19S. The project was registered (ID number CRD42020188890) in the International Prospective Register of Systematic Reviews (PROSPERO) website.[30] The study's findings are reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[31] The present systematic literature search was done using four academic databases (Scopus, PubMed Central, ProQuest, and ISI Web of Knowledge). Relevant studies were abstracted and their methodological quality was assessed using the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist.[31,32] Findings were synthesized using a narrative approach.

Search Strategy
Scopus, PubMed Central, ProQuest, and ISI Web of Knowledge were searched systematically from December 2019 to end of August 2021. "Fear of COVID-19 scale" and its abbreviation of "FCV-19S" together with IRT related terms (eg, item response theory, modern test theory, Rasch) were used to customized search syntax for the aforementioned academic databases.

Study Screening and Selection
The titles and abstracts of all retrieved papers were screened based on eligibility criteria. All peer-reviewed observational studies were considered eligible if psychometric properties of FCV-19 Scale were assessed using an IRT method. No limitation was exerted regarding participants' characteristics. After screening for potentially eligible papers, their full texts were downloaded and reviewed for final selection.

Quality Assessment
The methodological quality of the included papers was assessed using the COSMIN checklist. Generally, the COSMIN methodology is used to assess measurement properties of a single outcome measurement instrument or make a systematic evaluation of the most suitable tool in assessing a specific construct.[32,33] This checklist can be used to evaluate the methodological quality based on 10 measurement properties including outcome measure tool development, content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness.[33] Each of the aforementioned items have a clear guideline with several items for assessment. An overall score for the methodological quality of each measurement property for each study is determined by taking the lowest rating of any of the items according to the guideline.[34] The present systematic review specifically focused on the properties of internal consistency, structural validity, cross-cultural validity, criterion validity, and IRT properties. Therefore, these properties were selected and assessed through the included studies. Each item was rated using a Yes/No/Not applicable (NA) rating scale. No studies in the present review were excluded on the basis of poor methodological quality.

Data Extraction
Data extracted included the first author's name, publication date, title of the study, research area as country, occupation of participants, sample size, study design, IRT approach and their measures, COSMIN score (ie, methodological quality), and main findings of the study. Disagreements were resolved via discussion between research team members. To synthesize the data, a qualitative synthesis of findings was performed due to small number of included studies and the heterogeneity of the assessed outcomes.

Data Synthesis

To synthesize the data, a qualitative synthesis of findings was performed to report findings of included studies based on analytical method.

Results

Study Screening and Selection Process

The initial search in the four databases resulted in 552 papers. After removing duplicates, 479 papers were screened based on titles and abstracts. After removing irrelevant studies using the information from titles and abstracts, 118 papers were further evaluated using their full texts. Finally, 16 papers were deemed to be eligible and their full-texts were reviewed. Figure 1 shows the search process based on the PRISMA flowchart.

Study Description

The 16 studies comprised 32,624 participants from 21 different countries (Argentina, Bangladesh, Brazil, China, Cuba, Ethiopia, France, India, Iran, Italy, Japan, Jordan, Korea, Malaysia, Mozambique, New Zealand, Pakistan, Romania, Taiwan, Turkey, United Kingdom) and were included for final synthesis. The mean age of participants varied from 18 years to 52 years. Approximately half of overall participants were females (52.83%). Almost all studies recruited general population as the target participants, except for one study surveying adult patients with type 2 diabetes. Two main IRT methods were used: Graded Response Model (n=8 studies) and Rasch analysis (n=8 studies). The summary results of measurement properties are reported based on the IRT method used in the included studies. Table 1 provides the summary characteristics of all included studies. Table 1 also reports the link functions used in each study given that there are two types of functions (ie, logit or probit) in the IRT.[35]
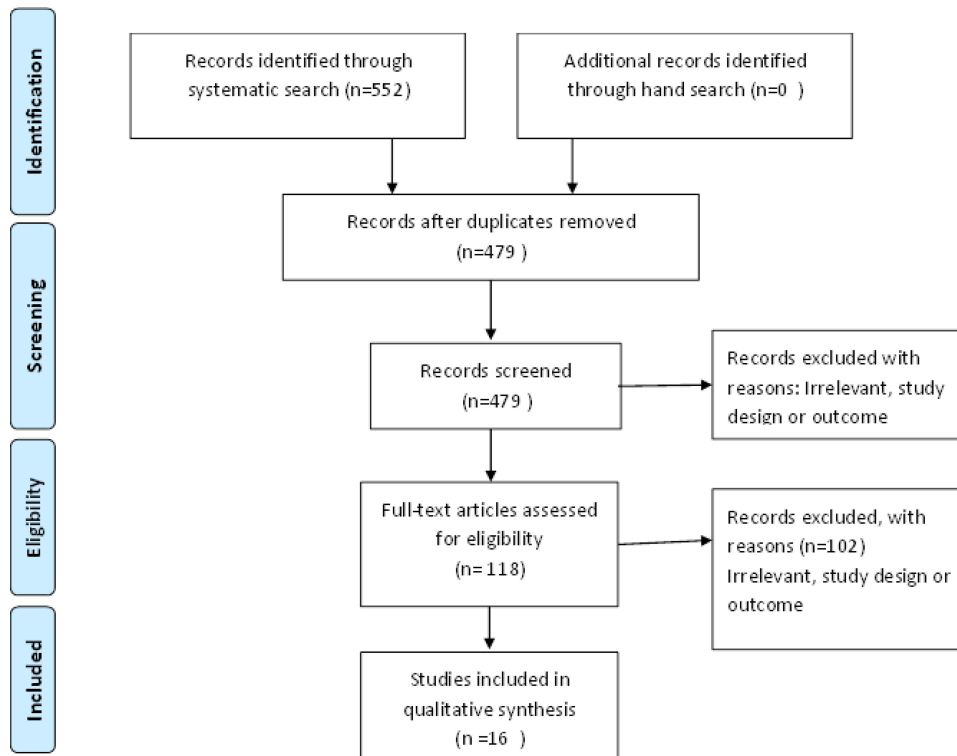

Figure 1 PRISMA flowchart.
Notes: Adapted from Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*. 2009;151(4):264–269.[31] Creative Commons.

585

Table 1 Summary Characteristics of Included Studies

First Author	Year	Country	N	Age Group (Years)	Age (Mean)	Female (%)	Language	Type of Participant	IRT Method; Link Function Type	Assumptions Tested	Software
Al-Shannaq[36]	2021	Jordan	725	18–65	33.7	56.4	Jordanian	Adults	Graded Response Model; not report	No	IRTPRO™
de Medeiros[37]	2021	Brazil	230	18–71	35.33	76.1	Portuguese	General population	Graded Response Model; not report	No	Package mirt in R
Basit[64]	2021	Pakistan	380	NR	51.93	46.05	NR	Adult type 2 diabetes patients	Graded Response Model; not report	No	Test Analysis Modules package for R
Bellamkonda[40]	2021	India	572	NR	22.70	58.7	English	College students	Graded Response Model; not report	No	R software
Elemo[42]	2020	Ethiopia Amharic	307	NR	NR	81.1	Amharic	General population	Graded Response Model; not report	No	JASP 0.11.1
Caycho-Rodrıguez[43]	2020	Argentina	1291	18–80	38.47	79.0	Spanish	General population	Graded Response Model; logit	No	"ltm" package R
Satici[44]	2020	Turkey	1304	18–64	29.47	70.3	Turkish	General population	Graded Response Model; not report	No	JASP 0.11.1
Han[45]	2021	Korea	300	19–65	NR	67.3	Korean	Adults	Rasch analysis; logit	No	jMETRIK 4.1.1
Giordani[38]	2021	Mozambique	387	18–70	34.5	51.7	Portuguese	General population	Rasch partial credit model; logit	No	R software
Chen[46]	2021	China	2445	NR	18.55	50.2	Chinese	General student population	Rasch analysis; logit	Yes	WINPEPS 3.74.0
Lin[39]	2021	Bangladesh, United Kingdom, Brazil, Taiwan, Italy, New Zealand, Iran, Cuba, Pakistan, Japan and France	15,663	10–92	29.64	53.4	Multi-language comparison	General student population	Rasch analysis; logit	No	WINPEPS 3.74.0
Winter[41]	2020	New Zealand	1397	18–88	47.5	40	English	General student population	Rasch analysis; logit	Yes	WINPEPS 3.74.0

(Continued)

Table 1 (Continued).

First Author	Year	Country	N	Age Group (Years)	Age (Mean)	Female (%)	Language	Type of Participant	IRT Method; Link Function Type	Assumptions Tested	Software
Winter[41]	2020	New Zealand	1023	18–85	42	69.7	English	General student population	Rasch analysis; logit	Yes	WINPEPS 3.74.0
Pang[47]	2020	Malaysia	228	NR	26	71.1	Malay	General student population	Rasch analysis; logit	No	jMetrik 4.1.1
Sakib[48,49]	2020	Bangladesh	8550	NR	26.5	44	Bangla	General student population	Rasch partial credit model; logit	Yes	WINSTEPS 4.3.0.
Ahorsu[12]	2020	Iran	717	<18	31.25	42	Persian	General student population	Rasch analysis; logit	No	WINSTEPS 3.75.0
Stănculescu[50]	2021	Romania	809	18 to 68	32.61	65.4	Romanian	General population	Graded Response Model; not report	No	ADANCO 2.2

Methodological Quality of Included Studies

Each item regarding the methodological quality of assessed psychometric properties (ie, internal consistency, structural validity, cross-cultural validity, criterion validity, and IRT properties) was rated on a Yes/No/Not applicable (NA) rating scale. The results of methodological assessment are presented in Figures 2–Figure 6. There were no serious flaws regarding the design or methods of the included studies in all aspects. However, most studies did not describe the percentage of missing items and how missing items were handled. IRT tests were not performed with the aspect of factor structure (ie, whether the FCV-19S is unidimensional or multidimensional) in most of studies. Moreover, most studies that evaluated the psychometric properties of translated FCV-19S were not reviewed by the original developers, the sample used in the pre-test among the translated FCV-19S was not adequately described, and no adequate descriptions were provided for the translated FCV-19S about how differences between the original and translated versions were resolved.

Summary Results of Rasch Analysis

Eight studies used Rasch analysis, a type of IRT analyses that require the discrimination parameters to be the same across all elements, to assess the measurement properties of FCV-19S. Table 2 provides the results, including information-weighted fit (infit) mean square (MnSq), outlier-sensitive (outfit) MnSq, and difficulty for each item. Infit MnSq varied from 0.73 to 1.38 (Both for FCV 3); Outfit MnSq varies from 0.68 to 1.05 (for FCV 3 and FCV 6 respectively); and difficulty of tem varied from −1.5 to 1.94 (for FCV 2 and FCV 6 respectively).

Summary Results of Graded Response Method Analysis

Eight studies used Graded Response Method IRT, a type of IRT analyses that allow element discrimination to be varied, to assess the measurement properties of FCV-19S. Table S1 provides the results for each item. In graded response analysis, a coefficient varied from 0.03 to 5.11 (for FCV5 and FCV6 respectively); b1 coefficient varied from− 5.21 to 1.19 (for FCV1 and FCV3 respectively); b2 coefficient varied from 0 to 1.86 (for FCV2 and FCV 3 respectively); b3
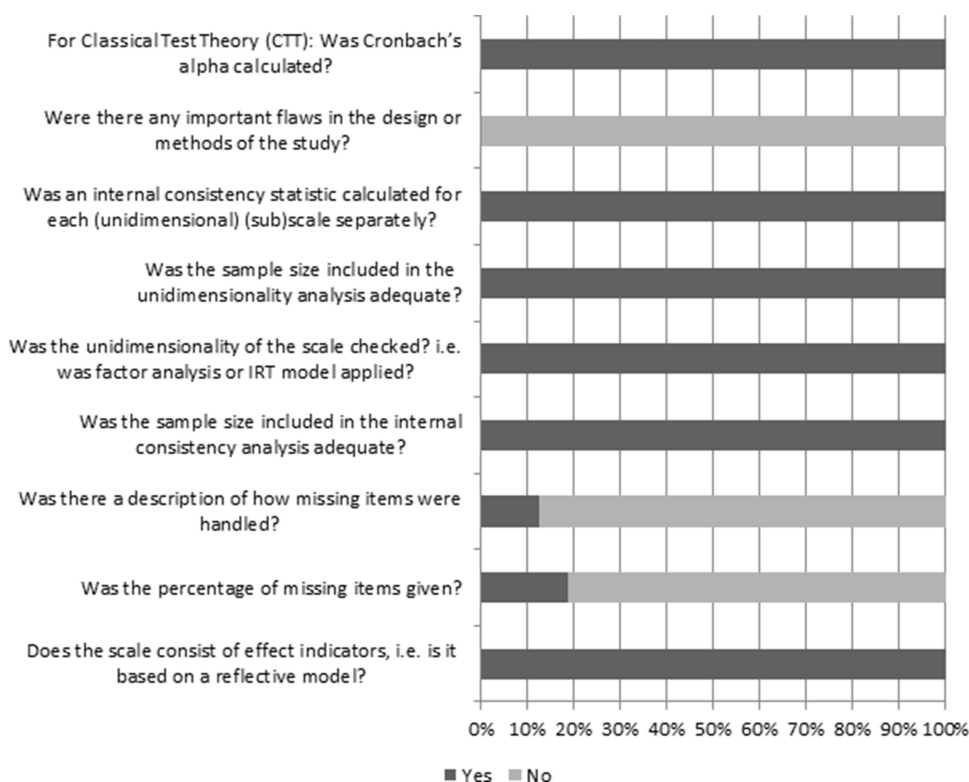

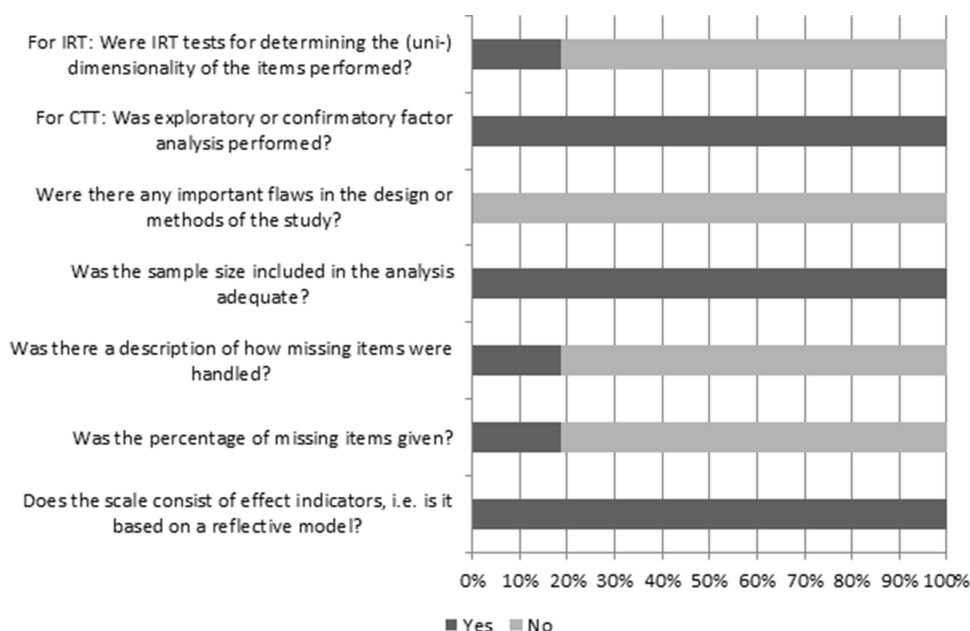**Figure 2** Results of methodological quality assessment regarding Box A. Internal consistency.

For IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?

For CTT: Was exploratory or confirmatory factor analysis performed?

Were there any important flaws in the design or methods of the study?

Was the sample size included in the analysis adequate?

Was there a description of how missing items were handled?

Was the percentage of missing items given?

Does the scale consist of effect indicators, i.e. is it based on a reflective model?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

■ Yes   ■ No

Figure 3 Results of methodological quality assessment regarding Box E – structural validity.

coefficient varied from 0.1 to 2.51 (for FCV5 and FCV3 respectively); b4 coefficient varied from 0.11 to 5.1 (for FCV5 and FCV6 respectively).

Summary Results of Differential Item Functioning (DIF) Contrast Across Age and Gender Group

Seven studies reported DIF contrast of FCV-19S items across age and gender groups. Table 3 provides the results for each item. DIF contrasts across age were ranged between 0.05 and 0.67 (FCV1); between 0 and 0.49 (FCV2); between 0 and 0.35 (FCV3); between 0 and 1.05 (FCV4); between 0 and 0.49 (FCV5); between 0.08 and 0.73 (FCV6); and between 0 and 0.76 (FCV7). DIF contrasts across gender were ranged between 0 and 0.2 (FCV1); between 0 and 0.33 (FCV2); between 0 and 0.29 (FCV3); between 0.06 and 0.33 (FCV4); between 0.04 and 0.48 (FCV5); between 0 and 0.24 (FCV6); and between 0.02 and 0.22 (FCV7). Therefore, FCV2, FCV3, FCV5 can be concluded as no substantial DIF across age; all FCV items as no substantial DIF across gender.

Discussion

The present systematic review evaluated the methodological quality of 16 studies regarding their IRT-based findings in relation to the Fear of COVID-19 Scale. Moreover, the IRT findings of the 16 studies are summarized (Tables 2,3) regarding the item properties, item discrimination, item difficulty, and DIF. Almost all the IRT properties assessed in the 16 reviewed papers were satisfactory. In addition, the IRT evidence was reported across different language versions, including Arabic (Jordan),[36] Portuguese,[37–39] English,[39–41] Amharic,[42] Spanish,[39,43] Turkish,[44] Korean,[45] Chinese,[39,46] Japanese,[39] Italian,[39] Urdu,[39] French,[39] Malay,[47] Bangla,[48,49] Persian,[12] and Romanian.[50] Therefore, the FCV-19S is verified by the findings of the present systematic review regarding its psychometric properties across different language versions.

In addition to the good properties shown in the CTT findings (eg, good internal consistency, test-retest reliability, and concurrent validity),[12,36–50] the present systematic review demonstrated the good properties were also shown on the IRT testing. From the findings of reviewed papers,[12,36–50] the seven items in the FCV-19S had satisfactory infit and outfit MnSq (ie, between 0.5 and 1.5),[51,52] indicating that every item in the FCV-19S is essential and fits in the concept of fear of COVID-19. Moreover, the item difficulties were acceptable (Table S1), which indicates that the seven items in the FCV-19S together explain a wide rage levels of fear of COVID-19.[23] Additionally, the item discriminations were good

Figure 4 Results of methodological quality assessment regarding Box G. Cross-cultural validity.

for the seven items. Therefore, the FCV-19S can efficiently and sharply differentiate low fear and high fear.[23] Another important feature in the IRT findings for the FCV-19S is that the instrument does not display substantial DIF across age and gender groups (Table 3). This indicates that the participants in different ages and different genders interpret the FCV-19S in a similar way.[53]

Although the present systematic review showed that the seven items of the FCV-19S are essential for assessing the concept of fear of COVID-19, there are some studies suggesting shortening the FCV-19S.[37,54,55] For example, Mercado-Lara et al[54] and Campo-Arias et al[55] recommended shortening the FCV-19S to a five-item scale, and de Medeiros et al[37] suggested a four-item version of the FCV-19S. However, given the synthesized findings from the present systematic review, it is apparent that shortening the FCV-19S may jeopardize the psychometric properties of the FCV-19S. Indeed, the seven items of the FCV-19S were all fully evaluated regarding their content and appropriateness for the fear concept when the FCV-19S developers generated this instrument.[12,56] Therefore, together with the synthesized findings reported here, it is strongly recommended that researchers and practitioners use the seven-item FCV-19S for assessing fear of COVID-19 rather than the briefer versions.
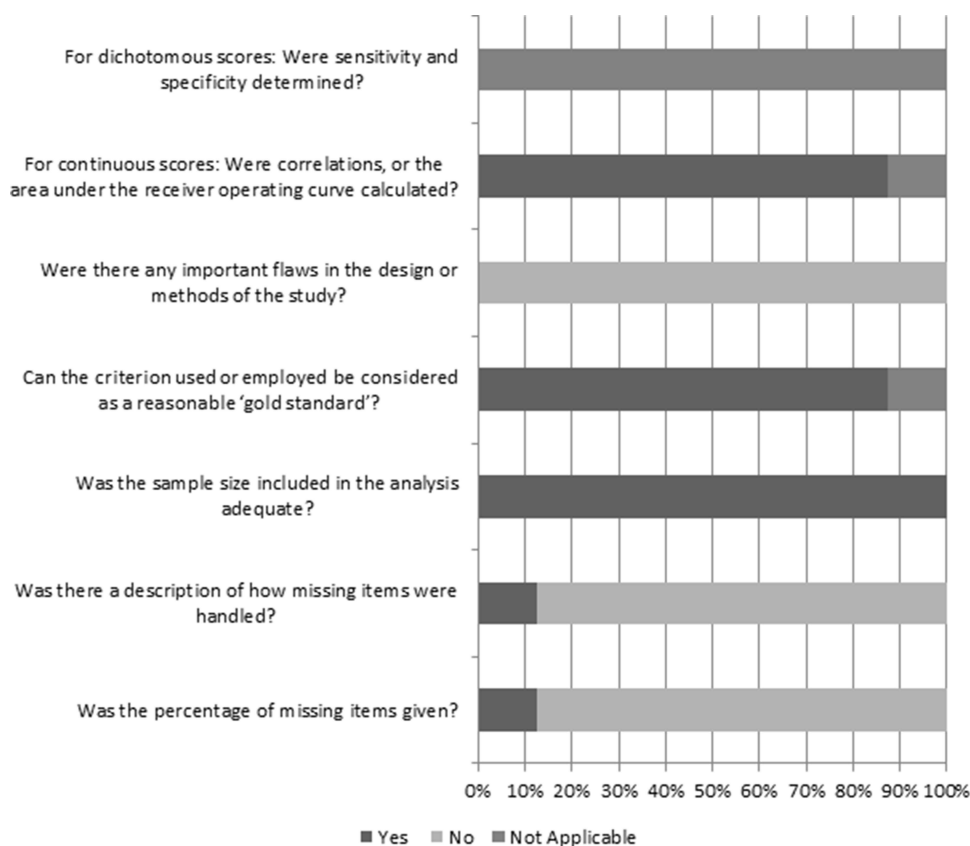
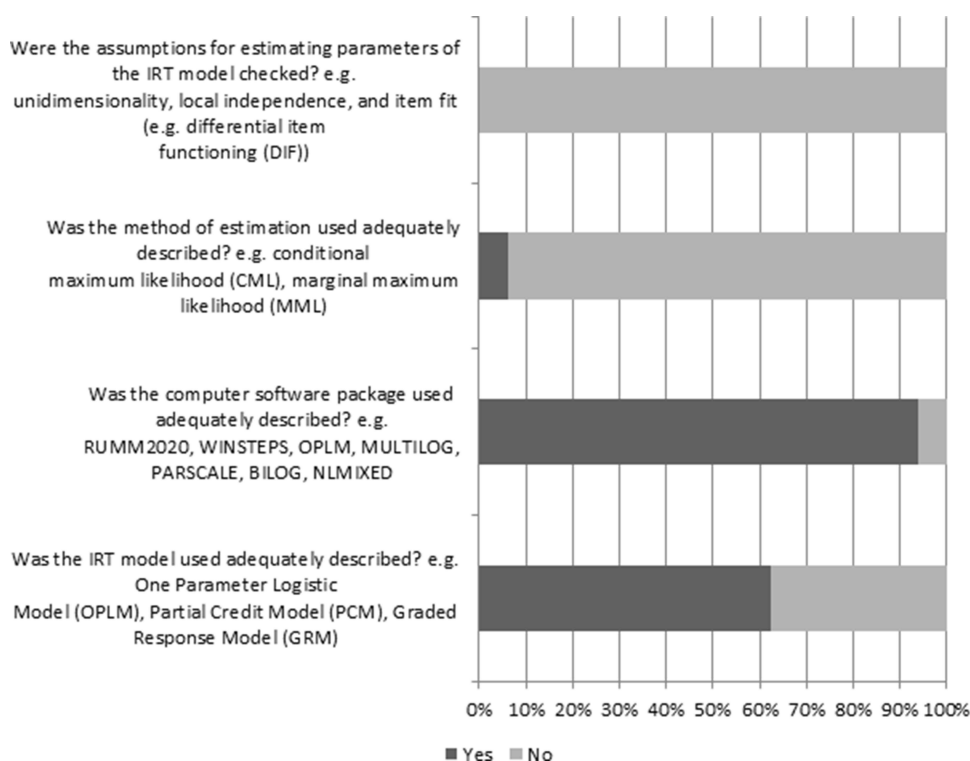Figure 5 Results of methodological quality assessment regarding Box H. Criterion validity.

Figure 6 Results of methodological quality assessment regarding Box IRT.

Table 2 Summary Results of Rasch Analysis Properties

First Author	Year	FCV1			FCV2			FCV3			FCV4			FCV5			FCV6			FCV7		
		Infit MnSq	Outfit MnSq	Difficulty	Infit MnSq	Outfit MnSq	Difficulty	Infit MnSq	Outfit MnSq	Difficulty	Infit MnSq	Outfit MnSq	Difficulty	Infit MnSq	Outfit MnSq	Difficulty	Infit MnSq	Outfit MnSq	Difficulty	Infit MnSq	Outfit MnSq	Difficulty
Han[45]	2021	1.08	1.14	−1.05	1.07	1.03	−1.50	0.76	0.68	0.63	0.88	0.89	0.37	0.91	0.90	−0.70	0.97	0.81	1.94	1.32	1.49	0.32
Giordani[38]	2021	0.85	0.90	−0.417	0.91	0.87	−0.003	0.91	0.85	1.718	0.90	0.87	−0.031	0.78	0.84	0.84	0.85	0.83	1.903	0.83	0.82	1.166
Chen[46]	2021	1.25	1.23		0.87	0.89		0.73	0.69		1.31	1.24		1.12	1.07		0.87	0.88		0.82	0.81	
Lin[39]	2021	1.07	1.13	−0.91	1.01	1.03	−0.74	0.91	0.9	1.01	1.13	1.1	−0.10	0.94	0.94	−0.63	0.88	0.86	0.98	1.02	1.01	0.39
Winter - Sample 1[41]	2020	1.11	1.17	− 1.04	0.94	0.98	− 0.86	0.87	0.72	0.96	1.36	1.25	− 0.06	1.13	1.12	− 0.64	1.07	0.92	0.78	1	0.8	0.87
Winter - Sample 2[41]	2020	1.11	1.12	− 1.11	0.93	0.91	− 0.95	0.84	0.69	1.11	1.3	1.29	0.1	1.16	1.19	− 0.81	0.97	0.87	0.82	1.06	0.89	0.85
Pang[47]	2020	0.87	0.89	−0.62	0.84	0.81	−0.44	1.38	1.3	0.99	0.83	0.76	−0.68	0.9	0.85	−0.16	1.1	1.05	0.55	1.07	1.02	0.36
Sakib[48,49]	2020	0.99	0.99	− 1.04	1.12	1.16	− 0.84	0.82	0.84	1	0.94	0.95	0.25	0.93	0.87	− 0.86	0.91	0.93	1.16	1.2	1.26	0.33
Ahorsu[12]	2020	1.26	1.25	0.98	0.8	0.84	− 0.17	0.81	0.85	0.39	1.11	1	−0.77	1.01	1	0.85	0.9	0.94	− 0.43	0.81	0.91	− 0.83
Minimum		0.85	0.89	−1.05	0.8	0.81	−1.5	0.73	0.68	0.39	0.83	0.76	0.1	0.78	0.84	−0.7	0.85	0.81	0.55	0.81	0.8	0.32
Maximum		1.26	1.25	0.98	1.12	1.16	−0.003	1.38	1.3	1.718	1.36	1.29	0.37	1.16	1.19	0.85	1.1	1.05	1.94	1.32	1.49	1.166

Table 3 Summary of Items DIF Based on Age and Gender

First Author	Year	DIF Age							DIF Gender						
		FCV 1	FCV 2	FCV 3	FCV 4	FCV 5	FCV 6	FCV 7	FCV 1	FCV 2	FCV 3	FCV 4	FCV 5	FCV 6	FCV 7
Giordani[38]	2021	−0.38	−0.01	0.14	0.31	0.24	0.14	0.31	−0.01	0.06	−0.05	−0.08	−0.04	−0.20	−0.03
Chen[46]	2021	0.1	0	0.15	0	−0.39	0.08	0	0.2	−0.23	−0.11	0.33	0.13	−0.21	−0.09
Lin[39]	2021	−0.67	−0.18	0.35	−1.05	0.28	0.66	0.76	0	0	−0.24	0.11	0.16	−0.06	−0.02
		0.30	−0.49	0.11	−0.40	0.00	0.73	0.50							
		0.37	−0.32	−0.24	0.65	−0.28	0.08	−0.26							
Winter - Sample 1[41]	2020	0.32	0.1	−0.11	0.43	−0.49	−0.19	−0.36	−0.06	0.06	−0.12	−0.25	0.18	−0.02	0.22
Winter - Sample 2[41]	2020	0.32	0	−0.08	0.42	−0.43	−0.08	−0.24	−0.04	0.05	0	−0.19	0.1	0	0.04
Pang[47]	2020								−0.07	−0.20	−0.11	−0.19	0.13	−0.22	−0.14
Sakib[48,49]	2020	−0.12	0	0	0.16	0.1	−0.28	0.15	0.1	0.17	−0.08	−0.06	−0.08	−0.06	0.05
Ahorsu[12]	2020	−0.05	−0.22	0.25	0.21	0.3	−0.31	−0.23	−0.10	−0.33	−0.29	0.29	0.48	−0.24	0.21
Absolute Minimum		0.05	0	0	0	0	0.08	0	0	0	0	0.06	0.04	0	0.02
Absolute Maximum		0.67	0.49	0.35	1.05	0.49	0.73	0.76	0.2	0.33	0.29	0.33	0.48	0.24	0.22

Notes: Giordani: DIF contrast across gender=difficulty for males - difficulty for females; DIF contrast across age=difficulty for older (> 34.5 years) - difficulty for younger (≤ 34.5 years). Chen: DIF contrast across gender, Differential Item Functioning for females and males; DIF contrast across age groups, Differential Item Functioning for younger (ie, 18 years) and older (ie, >18 years) students. Lin: DIF gender (M versus F); Age: C=children aged below 18 years; A=adults aged between 18 and 60 years; E=elderly aged over 60 years. Rows respectively are related to C versus A, C versus E, A versus E. Winter: DIF contrast across gender: difficulty for males-difficulty for females; eDIF contrast across age: difficulty for participants with younger age -difficulty for participants with older ages. Pang: DIF contrast across gender=difficulty for males (reference group) − difficulty for females (focal group). Sakib: DIF contrast across gender=Difficulty for females − Difficulty for males; DIF contrast across age categories =Difficulty for participants with older age (ie, ≥ 26.53 years) − Difficulty for participants with younger age (ie, < 26.53 years). Ahorsu: DIF contrast across gender=difficulty for males-difficulty for females eDIF contrast across age groups=Difficulty for younger (ie, ≤ 31.25 years) - Difficulty for older (ie, > 31.25 years) patients.

Strengths and Limitations

The present systematic review has a number of strengths. First, with the use of four important databases and following the PRISMA guidelines,[31] the present systematic review conducted a comprehensive search strategy. Consequently, the papers on FCV-19S properties were found and evaluated. Second, the systematic review used standardized international guidelines (ie, COSMIN checklist)[32,33] to evaluate the methodological quality for each included study and ensured that the included studies achieved an accepted set of quality criteria. This indicates that the evidence reviewed in the present systematic review is trustworthy and reliable. In particular, all the studies reviewed in the present systematic review had very good sample sizes (ranging between 228 and 15,653), which are far beyond the suggestion recommended by the COSMIN risk of bias checklist (ie, N > 100).[32,33] With the large sample sizes, the present systematic review is confident in the psychometric evidence synthesized from the studies reviewed because large sample sizes can somewhat decrease the biases in the summarized psychometric properties of FCV-19S.

Despite the aforementioned strengths, the present systematic review has a few limitations. First, the present systematic review did not use meta-analysis to quantify the findings of psychometric properties for the FCV-19S. Given that the reviewed papers used different methods in IRT (eight in Rasch and eight in graded response method), such psychometric information is hard to quantify using the meta-analysis method. Moreover, to the best of the present authors' knowledge, no proper meta-analysis methods can be used to integrate the IRT findings for the present systematic review. Therefore, the present systematic review can only provide qualitatively synthesized results for the psychometric properties of the FCV-19S. Second, responsiveness, an important psychometric property for healthcare providers to know if an instrument can be used to evaluate treatment effects,[57] was not examined in all the papers reviewed in the present systematic review. Without the information of "responsiveness", researchers cannot be confident that the FCV-19S can capture the treatment effect on fear of COVID-19 reduction. Therefore, further studies should be based on the good properties of FCV-19S found in the present systematic review to understand how much improvement shown in the FCV-19S indicates treatment effects on fear reduction. Third, the present systematic review primarily relied on academic databases to search the papers. Therefore, nonacademic databases, such as *Google Scholar* database, was not used in the search. Given that such a nonacademic database may provide more comprehensive findings (though the nonacademic database may be hard to screen because it contains a large portion of nonacademic articles in the search output) for scholars to supplement the literature search,[58] the present systematic review may have missed some IRT psychometric studies concerning the FCV-19S. However, the present authors believe that there will not be many because the four academic databases are well known for their coverage in health-related scientific papers. Fourth, as compared to classical test theory, IRT has the disadvantages that (i) it is more complex and most healthcare providers were unfamiliar with, including the software analyzing IRT models,[59,60] (ii) it requires large sample sizes to estimate accurate parameters[61] and some studies analyzed in the present systematic review seemed not to have sufficient sample size for IRT models, (iii) it needs strict assumptions and most studies analyzed in the present systematic review did not check these assumptions. Fifth, the FCV-19S, as with other psychological assessment tests,[62,63] has been modulated by cultural aspects; for example, Spanish adaptation in the FCV-19S is the Latin American adaptation from Argentina and Cuba; there are important cultural differences with the Spanish of Spain.

Conclusion

Based on the present review's findings, the seven FCV-19S items seem to be unidimensional in the assessment of fear, indicating that all items are necessary in the FCV-19S. More specifically, the psychometric properties of the FCV-19S, especially in relation to IRT properties, were systematically reviewed, summarized, and synthesized in the present paper. Apart from the IRT properties, some commonly used psychometric properties such as internal consistency and criterion validity were also reviewed. Almost all evidence reviewed in the present paper indicates that the FCV-19S is a promising instrument for assessing the fear of COVID-19. The seven items in the FCV-19S appear to be unidimensional in assessing COVID-19-related fear. Age and gender were not significant factors that affect individuals in interpreting any item in the FCV-19S, and diverse difficulties (eg, the item difficulties ranging between −5.21 and 5.11 *logit*) have been reported across the seven items in the FCV-19S that ensure different levels of fear can be captured by the FCV-19S items. However, current evidence does not indicate whether the FCV-19S has good responsiveness, ie, whether a treatment

program regarding fear of COVID-19 reduction can be effectively evaluated using the FCV-19S. Future studies are therefore needed to address this knowledge gap.

Acknowledgement

Disclosure

The authors report no conflicts of interest for this work.

References

1. Shanafelt T, Ripp J, Trockel M. Understanding and addressing sources of anxiety among health care professionals during the COVID-19 pandemic. *JAMA*. 2020;323(21):2133–2134. doi:10.1001/jama.2020.5893
2. Pakpour AH, Liu C-H, Hou W-L, et al. Comparing fear of COVID-19 and preventive COVID-19 infection behaviors between Iranian and Taiwanese older people: early reaction may be a key. *Front Public Health*. 2021;9. doi:10.3389/fpubh.2021.740333
3. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395 (10223):497–506. doi:10.1016/S0140-6736(20)30183-5
4. Lin C-Y. Social reaction toward the 2019 novel coronavirus (COVID-19). *Soc Health Behav*. 2020;3(1):1. doi:10.4103/SHB.SHB_11_20
5. Pappas G, Kiriaze I, Giannakis P, Falagas M. Psychosocial consequences of infectious diseases. *Clin Microbiol Infect*. 2009;15(8):743–747. doi:10.1111/j.1469-0691.2009.02947.x
6. Bhuiyan AI, Sakib N, Pakpour AH, Griffiths MD, Mamun MA. COVID-19-related suicides in Bangladesh due to lockdown and economic factors: case study evidence from media reports. *Int J Ment Health Addict*. 2020;19:1–6.
7. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. *Gen Psychiatry*. 2020;33(2):e100213. doi:10.1136/gpsych-2020-100213
8. Wang C, Pan R, Wan X, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *Int J Environ Res Public Health*. 2020;17(5):1729. doi:10.3390/ijerph17051729
9. Pakpour AH, Griffiths MD, Lin C-Y. Assessing psychological response to the COVID-19: the fear of COVID-19 scale and the COVID stress scales. *Int J Ment Health Addict*. 2020;19:1–4.
10. Chung G-K-K, Strong C, Chan YH, et al. Psychological distress and protective behaviors during the COVID-19 pandemic among different populations: Hong Kong general population, Taiwan healthcare workers, and Taiwan outpatients. *Front Med*. 2022;14. doi:10.3389/fmed.2022.800962
11. Taylor S, Landry CA, Paluszek MM, Fergus TA, McKay D, Asmundson GJ. Development and initial validation of the COVID stress scales. *J Anxiety Disord*. 2020;72:102232. doi:10.1016/j.janxdis.2020.102232
12. Ahorsu DK, Lin C-Y, Imani V, Saffari M, Griffiths MD, Pakpour AH. The fear of COVID-19 scale: development and initial validation. *Int J Ment Health Addict*. 2020:1–9. doi:10.1007/s11469-020-00270-8
13. Colizzi M, Bortoletto R, Silvestri M, et al. Medically unexplained symptoms in the times of Covid-19 pandemic: a case-report. *Brain Behav Immun*. 2020;5:100073. doi:10.1016/j.bbih.2020.100073
14. Montemurro N. Intracranial hemorrhage and COVID-19, but please do not forget "old diseases" and elective surgery. *Brain Behav Immun*. 2021;92:207–208. doi:10.1016/j.bbi.2020.11.034
15. Rajabimajd N, Alimoradi Z, Griffiths MD. Impact of COVID-19-related fear and anxiety on job attributes: a systematic review. *Asian J Soc Health Behav*. 2021;4(2):51–55. doi:10.4103/shb.shb_24_21
16. Chang K-C, Hou W-L, Pakpour AH, Lin C-Y, Griffiths MD. Psychometric testing of three COVID-19-related scales among people with mental illness. *Int J Ment Health Addict*. 2020;1–13. doi:10.1007/s11469-020-00361-6
17. Chen I-H, Chen C-Y, Zhao K-Y, et al. Psychometric evaluation of Fear of COVID-19 Scale (FCV-19S) among Chinese primary and middle school teachers, and their students. *Current Psychol*. 2022. doi:10.1007/s12144-021-02471-3
18. Wouters H, van Gool WA, Schmand B, Lindeboom R. Revising the ADAS-Cog for a more accurate assessment of cognitive impairment. *Alzheimer Dis Assoc Disord*. 2008;22(3):236–244. doi:10.1097/WAD.0b013e318174f8b9
19. McGrory S, Doherty JM, Austin EJ, Starr JM, Shenkin SD. Item response theory analysis of cognitive tests in people with dementia: a systematic review. *BMC Psychiatry*. 2014;14(1):1–15. doi:10.1186/1471-244X-14-47
20. Lord FM. *Applications of Item Response Theory to Practical Testing Problems*. New York, USA: Routledge Taylor and Francis Group; 2012.
21. Reise SP, Haviland MG. Item response theory and the measurement of clinical change. *J Pers Assess*. 2005;84(3):228–238. doi:10.1207/s15327752jpa8403_02
22. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of Item Response Theory*. USA: SAGE Publication; 1991.
23. Mungas D, Reed BR, Kramer JH. Psychometrically matched measures of global cognition, memory, and executive function for assessment of cognitive decline in older persons. *Neuropsychology*. 2003;17(3):380. doi:10.1037/0894-4105.17.3.380
24. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000;38:II28. doi:10.1097/00005650-200009002-00007
25. Sijtsma K, Emons WH, Bouwmeester S, Nyklíček I, Roorda LD. Nonparametric IRT analysis of quality-of-life scales and its application to the world health organization quality-of-life scale (WHOQOL-Bref). *Qual Life Res*. 2008;17(2):275–290. doi:10.1007/s11136-007-9281-6
26. Santor DA, Ascher-Svanum H, Lindenmayer J-P, Obenchain RL. Item response analysis of the positive and negative syndrome scale. *BMC Psychiatry*. 2007;7(1):1–10. doi:10.1186/1471-244X-7-66

27. Aggen SH, Neale MC, Kendler KS. DSM criteria for major depression: evaluating symptom patterns using latent-trait item response models. *Psychol Med*. 2005;35(4):475–487. doi:10.1017/S0033291704003563

28. Fraley RC, Waller NG, Brennan KA. An item response theory analysis of self-report measures of adult attachment. *J Pers Soc Psychol*. 2000;78 (2):350. doi:10.1037/0022-3514.78.2.350

29. Hill CD, Edwards MC, Thissen D, et al. Practical issues in the application of item response theory: a demonstration using items from the pediatric quality of life inventory (PedsQL) 4.0 generic core scales. *Med Care*. 2007;45(5):S39–S47. doi:10.1097/01.mlr.0000259879.05499.eb

30. Alimoradi Z, Pakpour A Fear of COVID-19: measurement properties of the scale and its relation with psychological distress. PROSPERO: international prospective register of systematic reviews; 2020.

31. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–269. doi:10.7326/0003-4819-151-4-200908180-00135

32. Mokkink LB, De Vet HC, Prinsen CA, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171–1179. doi:10.1007/s11136-017-1765-4

33. Mokkink LB, Prinsen C, Patrick DL, et al. *COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User Manual*; 2018:78.

34. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651–657. doi:10.1007/s11136-011-9960-1

35. Madsen H, Thyregod P. *Introduction to General and Generalized Linear Models*. CRC Press; 2010.

36. Al-Shannaq Y, Mohammad AA, Khader Y. Psychometric properties of the Arabic version of the Fear of COVID-19 Scale (FCV-19S) among Jordanian adults. *Int J Ment Health Addict*. 2021;1–14. doi:10.1007/s11469-021-00574-3

37. de Medeiros ED, Reis LM, Guimarães CLC, et al. Psychometric properties of the Brazilian version of the Fear of COVID-19 Scale (FCV-19S). *Current Psychol*. 2021;1–10. doi:10.1007/s12144-021-01476-2

38. Giordani RCF, Giolo SR, Muhl C, Estavela AJ, Mabuie Gove JI. Validation of the FCV-19 Scale and assessment of Fear of COVID-19 in the population of Mozambique, East Africa. *Psychol Res Behav Manag*. 2021;14:345. doi:10.2147/PRBM.S298948

39. Lin CY, Hou WL, Mamun MA, et al. Fear of COVID-19 Scale (FCV-19S) across countries: measurement invariance issues. *Nurs Open*. 2021;8 (4):1892–1908. doi:10.1002/nop2.855

40. Bellamkonda N, Pattusamy M. Validation of Fear of COVID-19 Scale in India: classical test theory and item response theory approach. *Int J Ment Health Addict*. 2021;1–8. doi:10.1007/s11469-021-00521-2

41. Winter T, Riordan BC, Pakpour AH, et al. Evaluation of the English version of the Fear of COVID-19 Scale and its relationship with behavior change and political beliefs. *Int J Ment Health Addict*. 2020:1–11. doi:10.1007/s11469-020-00342-9

42. Elemo AS, Satici SA, Griffiths MD. The Fear of COVID-19 Scale: psychometric properties of the Ethiopian Amharic version. *Int J Ment Health Addict*. 2020;1–12. doi:10.1007/s11469-020-00448-0

43. Caycho-Rodríguez T, Vilca LW, Cervigni M, et al. Fear of COVID-19 Scale: validity, reliability and factorial invariance in Argentina's general population. *Death Stud*. 2020:1–10. doi:10.1080/07481187.2020.1836071

44. Satici B, Gocet-Tekin E, Deniz ME, Satici SA. Adaptation of the Fear of COVID-19 Scale: its association with psychological distress and life satisfaction in Turkey. *Int J Ment Health Addict*. 2020;19:1–9.

45. Han J-W, Park J, Lee H. Validity and reliability of the Korean version of the Fear of COVID-19 scale. *Int J Environ Res Public Health*. 2021;18 (14):7402. doi:10.3390/ijerph18147402

46. Chen W, Liang Y, Yin X, Zhou X, Gao R. The factor structure and rasch analysis of the Fear of COVID-19 Scale (FCV-19S) among Chinese students. *Front Psychol*. 2021;12:678979.

47. Pang NTP, Kamu A, Hambali NLB, et al. Malay version of the Fear of COVID-19 Scale: validity and reliability. *Int J Ment Health Addict*. 2020:1–10. doi:10.1007/s11469-020-00355-4

48. Griffiths MD, Pakpour AH, Mamun MA. Psychometric validation of the Bangla Fear of COVID-19 Scale: confirmatory factor analysis and rasch analysis. *Int J Ment Health Addict*. 2020. doi:10.1007/s11469-020-00399-6

49. Sakib N, Bhuiyan AI, Hossain S, et al. Psychometric validation of the Bangla Fear of COVID-19 Scale: confirmatory factor analysis and Rasch analysis. *Int J Ment Health Addict*. 2020:1–12. doi:10.1007/s11469-020-00289-x

50. Stănculescu E. Fear of COVID-19 in Romania: validation of the Romanian version of the fear of COVID-19 scale using graded response model analysis. *Int J Ment Health Addict*. 2021:1–16. doi:10.1007/s11469-020-00428-4

51. Nejati B, Fan C-W, Boone WJ, Griffiths MD, Lin C-Y, Pakpour AH. Validating the Persian Intuitive Eating Scale-2 among breast cancer survivors who are overweight/obese. *Eval Health Prof*. 2020:0163278720965688. doi:10.1177/0163278720965688

52. Lin C-Y, Broström A, Griffiths MD, Pakpour AH. Psychometric evaluation of the Persian eHealth Literacy Scale (eHEALS) among elder Iranians with heart failure. *Eval Health Prof*. 2020;43(4):222–229. doi:10.1177/0163278719827997

53. Lin C-Y, Imani V, Griffiths MD, Pakpour AH. Psychometric properties of the Persian Generalized Trust Scale: confirmatory factor analysis and rasch models and relationship with quality of life, happiness, and depression. *Int J Ment Health Addict*. 2020;19:1–12.

54. Mercado-Lara MF, Campo-Arias A, Monterrosa-Castro Á. Validity and reliability of the Spanish version of Fear of COVID-19 Scale in Colombian physicians. *Int J Ment Health Addict*. 2021:1–8. doi:10.1007/s11469-020-00430-w

55. Campo-Arias A, Tirado-Otálvaro AF, Álvarez-solorza I, Cassiani-Miranda CA. Confirmatory factor analysis, internal consistency, gender differential item functioning and discriminant validity of the Fear of COVID-5 Scale amidst emerging adult university students in Mexico. *Omega J Death Dying*. 2021:1–10. doi:10.1177/00302228211016216

56. Lin C-Y, Griffiths MD, Pakpour AH. Psychometric properties of the Fear of COVID-19 Scale: a Response to Mercado-Lara et al."Validity and reliability of the Spanish version of Fear of COVID-19 Scale in Colombian physicians". *Int J Ment Health Addict*. 2021:1–3. doi:10.1007/s11469-020-00430-w

57. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol*. 2000;53(5):459–468. doi:10.1016/S0895-4356(99)00206-1

58. Poon LY, Tsang HW, Chan TY, et al. Psychometric properties of the Internet Gaming Disorder Scale–Short-Form (IGDS9-SF): systematic review. *J Med Internet Res*. 2021;23(10):e26821. doi:10.2196/26821

59. Chang -C-C, Su J-A, Tsai C-S, Yen C-F, Liu J-H, Lin C-Y. Rasch analysis suggested three unidimensional domains for Affiliate Stigma Scale: additional psychometric evaluation. *J Clin Epidemiol*. 2015;68(6):674–683. doi:10.1016/j.jclinepi.2015.01.018

60. Chang K-C, Wang J-D, Tang H-P, Cheng C-M, Lin C-Y. Psychometric evaluation, using Rasch analysis, of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment: further item validation. *Health Qual Life Outcomes*. 2014;12(1):1–9. doi:10.1186/s12955-014-0148-6

61. Chang -C-C, Lin C-Y, Gronholm PC, Wu T-H. Cross-validation of two commonly used self-stigma measures, Taiwan versions of the Internalized Stigma Mental Illness Scale and Self-Stigma Scale–Short, for people with mental illness. *Assessment*. 2018;25(6):777–792. doi:10.1177/1073191116658547

62. Hays RD, Spritzer KL, Reise SP. Using item response theory to identify responders to treatment: examples with the Patient-Reported Outcomes Measurement Information System (PROMIS®) physical function scale and emotional distress composite. *Psychometrika*. 2021;86(1):1–12. doi:10.1007/s11336-021-09748-3

63. Sischka PE, Costa AP, Steffgen G, Schmidt AF. The WHO-5 well-being index–validation based on item response theory and the analysis of measurement invariance across 35 countries. *J Affect Disord Rep*. 2020;1:100020. doi:10.1016/j.jadr.2020.100020

64. Basit KA, Zafar AB, Fawwad A, Waris N, Shaheen F, Basit A. Psychometric analysis for Fear of COVID-19 Scale (FCV-19S) and its association with depression in patients with diabetes: a cross sectional study from a tertiary care centre in Karachi, Pakistan. *Diabetes Metab Syndr*. 2021;15 (3):733–737. doi:10.1016/j.dsx.2021.03.008