Open Access Full Text Article

ORIGINAL RESEARCH

Validation of a Chinese Version of the Digital Stress Scale and Development of a Short Form Based on Item Response Theory Among Chinese College Students

Chao Zhang (iD), Buyun Dai (iD), Lingkai Lin (iD)

School of Psychology, Jiangxi Normal University, Nanchang, Jiangxi, People's Republic of China

Correspondence: Buyun Dai, Jiangxi Normal University, 99 Ziyang Ave, Nanchang, Jiangxi, 330022, People's Republic of China, Tel +86-13870055780, Email biweijianpsy@qq.com

Purpose: In 2021, Hall et al developed the Digital Stress Scale (DSS), but its psychometric characteristics were only tested using classical test theory (CTT). In this study, we use item response theory (IRT) and CTT to develop and verify a Chinese version of the DSS and its short version, which can improve the reliability and effectiveness of the digital stress measurement tool for Chinese college students.

Methods: In this study, we developed a Chinese version of the DSS (DSS-C) and recruited 1506 Chinese college students as participants to analyze its psychometric characteristics based on CTT and item response theory methods. First, we used CTT, including common method bias, construct validity, criterion-related validity, internal consistency, test-retest reliability and measurement invariance. Then, we adopted the IRT approach to examine the item parameters, item characteristics, item information, differential item function, test information, and test reliability of the DSS-C. Finally, a short form (DSS-C-S) was constructed, and the psychometric characteristics of the DSS-C-S were examined.

Results: (1) The five-factor structure of the DSS-C was verified. The DSS-C shows good internal reliability, test-retest reliability, criterion-related validity and measurement invariance between urban college students and rural college students. (2) All 24 items had reasonable discrimination parameters and location parameters and were DIF-free by gender. Except for Items 4 and 10, all the other items had high information and measurement reliability at medium θ levels. (3) Compared with the DSS-C, the 22-item short form also has good reliability and validity and maintains sufficient measurement accuracy while reducing items of poor quality.

Conclusion: Both the DSS-C and the DSS-C-S have good psychometric characteristics and are accurate and effective tools for measuring the digital stress of Chinese college students.

Keywords: digital stress scale, item response theory, measurement invariance, short form

Introduction

Digital media is an important element of contemporary social life and is becoming increasingly popular, especially among teenagers and young adults. Digital media has created opportunities for the learning, development, personal exploration and growth of adolescents and young adults, but it also raises the risk of a series of mental health problems.[1,2] Previous studies have shown an association between digital media use and psychosocial functions such as depressive symptoms, anxiety, loneliness, happiness and quality of life.[3] A series of meta-analyses of different types of digital media use and well-being have shown that interaction, self-presentation, and entertainment on digital media were associated with better well-being, whereas consuming digital media content was associated with poorer well-being.[4]

Combined with relevant concepts from the past,[1,5,6] Steele et al[7] defined digital stress as "the stress and anxiety brought about by the notification and use of information and communication technologies of mobile phones and social

2897

media". A growing body of research[1,6,8,9] has shown that digital stress is an important variable for understanding the relationship between digital media use and the mental health level of adolescents and young adults.

Previous studies have proposed two explanations for how digital stress affects the relationship between digital media use and psychosocial functioning.[7] The first explanation suggests that digital stress may play a mediating role in the association between digital media use and psychosocial functioning.[6,8,10,11] The second explanation suggests that digital stress may play a moderating role in the association between digital media use and psychosocial outcomes.[12]

At present, many qualitative and quantitative research questionnaires in various countries involve the concept of digital stress, such as the fear of missing out (FoMO) scale,[13,14] the perceived information overload scale,[15] the communication load scale,[1] the connection overload scale,[9] the accessibility stress questionnaire,[6] and the mobile entrapment scale.[8,11] However, past studies have used a variety of different terms to describe similar or the same structure of digital stress, and the lack of a unified structure to measure digital stress has hindered the development of digital stress research to some extent.[7] Therefore, Steele et al[7] synthesized the components of digital stress discussed in the multidisciplinary literature to propose a multidimensional conceptual model of digital stress in which digital stress was viewed as an overall (higher-order) structure consisting of four substructures: availability stress, approval anxiety, FoMO and connection overload.

Hall et al[16] developed the Digital Stress Scale (DSS) based on the multidimensional digital stress model proposed by Steele et al[8] and the interview results. The scale includes five dimensions: availability stress, approval anxiety, FoMO, connection overload and online vigilance. Among them, (1) availability stress is defined as the "distress (including guilt and anxiety) resulting from beliefs about others' expectations that the individual responds and be available by digital means". For example, "For my friends, it is important that I am constantly available online". (2) Approval anxiety is defined as the "uncertainty and anxiety about others' responses and reactions to one's posts or to elements of one's digital footprint". For example, "I am nervous about how people will respond to my posts and photos". (3) FoMO is defined as the "distress resulting from the real, perceived, or anticipated social consequences of others engaging in rewarding experiences from which one is absent". For example, "I get worried when I find out my friends are having fun without me". (4) Connection overload is defined as the "distress resulting from the subjective experience of receiving excessive input from digital sources, including notifications, text messages, posts, etc.". For example, "I feel overwhelmed with the flow of messages/notifications on my phone". (5) Online vigilance focuses on compulsive checking of social media accounts and a strong desire to have access to one's mobile device/phone. For example, "I have to have my phone with me to know what's going on".

Previous studies have shown that the original DSS has good psychometric characteristics, and the five-factor model has a good fitting result.[16] The total score of DSS and anxiety and other psychosocial functions measured by the Patient-Reported Outcome Measurement Information System (PROMIS;[17]) were significantly correlated. The Cronbach coefficient (α) is between 0.85 and 0.93 for the total scale and each dimension. However, the results of previous studies are from the Western cultural background. Whether the structure of DSS in China is consistent with that of Western studies and whether the items of DSS are also applicable to Chinese people under the background of Chinese culture remain to be tested. To construct a digital stress scale in line with the Chinese cultural background, it is necessary to revise the Chinese version of the DSS (DSS-C).

A notable issue is that previous studies of the DSS did not assess whether the structure of the instrument remained unchanged across family locations (urban vs rural). This is important because there may be some differences in digital stress between urban and rural areas in China. On the one hand, there is a large gap in the development of the internet and digital media between urban and rural areas in China, forming a "digital divide" phenomenon,[18] and the experience of digital stress is highly dependent on situational contexts.[19] On the other hand, there are significant differences in mobile attachment between urban and rural college students in China,[20] and higher mobile attachment is associated with more serious FoMO and online vigilance.[21,22] Therefore, it is necessary to test whether the structure of the DSS-C has measurement invariance in urban college students and rural college students in China.

All previous studies on the DSS were based on classical test theory (CTT). Although CTT is practical and easy to implement, it still has many shortcomings. For example, the estimation of test and item indicators is only effective for the current sample group. The estimation of measurement error is general, and the assumption that all

scores have the same measurement error is inconsistent with the actual situation. Compared with CTT, the estimated parameters of item response theory (IRT) have sample invariance. The measurement accuracy can be determined according to the information provided by each item and test for each level of participants, and the measurement error of each item and test can be accurately estimated for different participants.[23]

Despite the shortcomings mentioned above, CTT is still widely used because it is easy to understand and has strong operability (e.g.,[24]). Many scholars have combined CTT with IRT to analyze the psychometric characteristics of the scale (e.g.,[25,26]). Therefore, we will do the same in this study.

A fundamental question in testing psychometric tools is whether they can be used to compare groups with different characteristics in a fair and reliable manner. To make the scores of different groups comparable, the functions of items in all groups must be identical,[27] namely, there is no differential item functioning (DIF). With reference to Hall et al,[16] we will further examine whether DSS-C items have DIF in terms of gender under the framework of IRT. In addition, the fact that a scale has good psychometric characteristics does not mean that all items are of good quality. Some items may be redundant and of relatively poor quality, and deleting them will not significantly affect the overall quality of the scale. In practice, we often use more than one scale at a time, and too many items may burden the participants. If each scale is kept as short as possible, testing burden and fatigue can be reduced. Therefore, we try to delete some items whose CTT and IRT analysis results show poor psychometric characteristics and to construct a short form of the DSS-C (DSS-C-S) with fewer items but good psychometric quality.

Therefore, the present study aims to develop the DSS-C and examine its psychometric characteristics using CTT and IRT. First, CTT was used to test the reliability and validity of the DSS-C, and the results were compared with those of the original authors of the DSS. Then, we tested whether the five-factor structure of the DSS-C has measurement invariance in urban college students and rural college students in China. In addition, we used IRT to further evaluate the psychometric characteristics of the DSS-C to test whether the psychometric characteristics of the DSS-C are good and whether it can be used as an effective and reliable tool to measure digital stress in Chinese college students. Finally, we constructed a short scale with fewer items but good psychometric quality.

Methods

Participants

A total of 1506 Chinese university students aged 18–26 from 31 provincial administrative regions in China were recruited to complete the DSS-C online questionnaire. After 132 invalid answers were excluded due to attention check failures, 1374 valid questionnaires (M_{age} = 20.33, SD = 1.65) were included, with an effective response rate of 91.2%. There were 655 males (47.7%) and 719 females (52.3%); 671 (48.8%) participants lived in urban areas, and 703 (51.1%) participants lived in rural areas. There were 317 freshman (23.0%), 389 sophomore (28.3%), 381 junior (27.7%), and 287 senior (20.8%) participants. There were 503 (36.6%) participants majoring in science, 491 (35.7%) participants majoring in liberal arts, 204 (14.8%) participants majoring in engineering, 55 (4.0%) participants majoring in arts, 23 (1.6%) participants majoring in medicine, and 98 (7.1%) participants majoring in other fields. Of the participants with valid answers, 997 (M_{age} = 20.26, SD = 1.76), including 546 men (54.7%) and 451 women (45.2%), completed the PROMIS anxiety and fatigue scale (Chinese version) at the same time. A total of 451 of these participants (45.2%) lived in urban areas, and 546 (54.7%) lived in rural areas. There were 270 freshmen (27.0%), 269 sophomores (26.9%), 238 juniors (23.8%) and 220 seniors (22.0%).

After 20 days, a subgroup of 34 participants took part in the DSS-C retest, which was used to analyze the retest reliability of the scale. Among them, 16 were male and 18 were female. There were 8 freshmen, 8 sophomores, 11 juniors, and 7 seniors.

Procedure

The WJX platform (https://www.wjx.cn/) was used to administer an online questionnaire including the demographic variables, the DSS-C, and the PROMIS anxiety and fatigue scale used in this study. The online questionnaire was

distributed to university students at various schools between April and October 2022 through social media such as QQ and WeChat and the Aishiyan platform (http://aishiyan.bnu.edu.cn). Each participant was informed of the purpose and confidentiality of the questionnaire before answering, and was asked to answer in a quiet place for no less than 1 minute and no more than 5 minutes.

Measures

Previous studies have shown significant associations between digital stress and anxiety.[2,28,29] Referring to the study of Hall et al,[16] the Chinese version of the PROMIS anxiety scale was selected as the criterion scale in this study.

Demographic Information

The demographic information included age, gender, grade, family location, and major.

DSS-C

The scale consisted of 24 items, including five dimensions: availability stress, approval anxiety, FoMO, connection overload and online vigilance. Connection overload and availability stress contained 6 items each, and acceptance anxiety, FoMO, and online vigilance contained 4 items each. A 5-point Likert-type scale ranging from 1 (never) to 5 (always) was used for the item.

The items of the DSS-C were translated by a university teacher in the field of psychometrics, an undergraduate majoring in English and an undergraduate majoring in quantitative psychology. First, we translated English into Chinese and ensured that the Chinese meaning was clear and consistent with the English meaning. Then, other English professionals not involved in the English-Chinese translation process, including an English major graduate student and a university teacher with experience studying in an English-speaking country, were asked to check for discrepancies with the original meaning. The statements of the items with disagreement were modified until a final agreement was reached. Finally, we checked the readability of DSS-C items with 9 non-psychology college students who were not involved in the previous steps to ensure that there was no difficulty or ambiguity in understanding each item among ordinary college students. The items of the DSS-C can be found in Supplementary Table 1.

The Chinese Version of PROMIS Anxiety Scale

The original PROMIS is a set of scales developed by the National Institutes of Health to assess the physical, mental, and social health of adults and adolescents.[17] The Chinese version of the PROMIS anxiety scale was revised by Wu[30] and found to have good reliability and validity after CTT and IRT tests. The anxiety scale contains 8 items, each of which adopts a Likert score ranging from 1 (never) to 5 (always). The participants were asked to evaluate their anxiety experience in the past 7 days. The higher the score was, the more serious their anxiety symptoms were. The Cronbach coefficient (α) of this scale in this study was 0.951. The items of the Chinese version of PROMIS anxiety scale can be found in Supplementary Table 2.

Analysis
CTT analysis
Common Method Bias (CMB) Test

CMB was examined by controlling for the effects of an unmeasured latent methods factor (ULMC), and all items of the DSS-C, PROMIS anxiety scale were used as method factor indicators to establish a bifactor model on the basis of the original trait single factor. If the change amplitude of CFI and TLI is more than 0.1 and the change amplitude of RMSEA and SRMR is more than 0.05 between the two-factor model and the single-factor model containing only traits, there is a serious common method bias.[31,32]

Structural Validity

A previous study[1] showed that the DSS has a good degree of fitting in both the five-factor first-order model and the five-factor high-order model. Therefore, we carried out a confirmatory factor analysis (CFA) of the DSS-C based on these two models and selected the root mean square error of approximation (RMSEA), standardized root mean square residual

(SRMR), comparative fit index (CFI) and Tucker–Lewis index (TLI) as indicators to evaluate the degree of model fitting. If RMSEA and SRMR are less than 0.08 and CFI and TLI are more than 0.90, the fitted model is acceptable.[33–35] The Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used as indicators for model comparison to select better fitting models for further analysis. The smaller the values of AIC and BIC are, the better the model fits.

Criterion-Related Validity

By analyzing the correlation degree between the total score of the DSS-C and the total score of each dimension and the scores of the Chinese PROMIS anxiety scale, the criterion-related validity of the DSS-C was evaluated.

Reliability

Cronbach coefficients (α) of each dimension and total scale were obtained. A Cronbach coefficient (α) greater than 0.700 indicates good internal consistency.[36] In addition, the Pearson correlation coefficient was calculated to obtain the test-retest reliability of the DSS-C.

Measurement Invariance

The measurement invariance analysis of urban college students and rural college students was carried out by multigroup confirmatory factor analysis (MCFA). Depending on the object of inspection, measurement invariance from low to high can be divided into four levels,[37] including ① configural invariance: verify whether the composition or patterns of latent variables are the same across the groups; ② weak invariance: verify whether the factor loadings are equal across the groups; ③ strong invariance: verify whether the intercepts of the observed variable are invariable across the groups; and ④ strict invariance: verify whether the error variances are equal across the groups. Therefore, in this study, we analyze and compare the configural invariance, weak invariance, strong invariance and strict invariance of the DSS-C in urban college students and rural college students step by step. According to the standard of measuring invariance tested by the method of the difference of fitting index CFI, TLI, RMSEA proposed by Cheung and Rensvold,[38] a difference <0.01 indicates no difference, a difference between 0.01 and 0.02 indicates a moderate difference, and a difference >0.02 indicates an obvious difference.

IRT Analysis

Fit Analysis of the IRT Model

The commonly used multilevel score models for IRT are the generalized partial credit model (GPCM)[39] and the graded response model (GRM).[40] In this study, the fitting degree of the DSS-C under the GRM and the GPCM was compared according to the fitting index AIC, BIC and −2Log-Lik, and the model with the best fitting index was selected for subsequent IRT analysis.

Item Measurement Characteristics Analysis

Based on IRT, the discrimination parameters (a) and location parameters (b) of the DSS-C items were estimated, and the quality of the item parameters was analyzed. Meanwhile, the item characteristic curves (ICCs) and item information function (IIF) of each item were compared.

DIF Analysis

In this study, McFadden's[41] pseudo R^2 method was used for logistic regression to test whether there was gender DIF in items. According to the pseudo R^2 method, DIF exists for the items with a change in R^2 of 0.02 or more.

Test Measurement Characteristics Analysis

The analysis of test measurement features includes test information and test reliability. The test information is obtained by adding the item information of all items, which is an index to evaluate the test accuracy.[27] Under IRT, test reliability (r_{xx}) can be calculated by the amount of information (I). r_{xx} is calculated as $r_{xx}(\theta) = 1 - \frac{1}{1(\theta)}$.[42] After the test information is obtained, the test reliability is calculated according to the formula.

Construction of a Short Form

A short form of the DSS-C was constructed by retaining the items of better quality among the above analysis indexes, and the psychometric characteristics of the DSS-C-S were analyzed. The psychometric analysis of the DSS-C-S included CFA, internal consistency, measurement invariance, fit analysis of the IRT model, item parameter estimation, DIF analysis, test information and reliability analysis.

Data Analysis Software

SPSS 23.0 software was used for reliability analysis and criterion-related validity analysis. ULMC, CFA and measurement invariance were performed with Mplus 8.8. The remaining analyses were performed using packages in R: the mirt package was used for the fit analysis of the IRT model, for estimating IRT parameters and for plotting ICCs. The catR package was used to draw IIFs, test information curves (TIFs) and test reliability curves (TRCs). The lordif package was used to identify DIFs.

Results
CTT Analysis
CMB Test

The ULMC results are as follows: The fitting indexes of the single factor model are CFI = 0.945, TLI = 0.940, RMSEA = 0.053, SRMR = 0.042. The fitting indexes of the bifactor model were CFI = 0.961, TLI = 0.954, RMSEA = 0.046, SRMR = 0.031. CFI, TLI, RMSEA, and SRMR did not vary by more than 0.05, so there was no significant CMB in this study.

Structural Validity

With reference to the hypothesis of the original authors of the DSS, we verify the five-factor first-order model and five-factor higher-order model of the DSS-C. The figures of the CFA models are shown in Supplementary Figure 1 and 2. Table 1 shows that the CFI and TLI of the five-factor first-order model and the five-factor high-order model are both greater than 0.9, and the values of RMSEA and SRMR are both less than 0.08, indicating that the models fit well with the data, and the five-factor structure of the DSS-C has been verified in China. It can be seen from Table 2 that the 24 items all have high loads in their dimensions. The AIC value and BIC value of the five-factor first-order model are both smaller than those of the five-factor high-order model, indicating that the five-factor first-order model has a better fit. Therefore, the analysis will be conducted based on the five-factor first-order model in the following paper.

Criterion-Related Validity

The results are shown in Table 3. In the sample of adult college students, the total score of the DSS-C was significantly positively correlated with the total score of the Chinese version of the PROMIS anxiety scale. Except for the dimension of availability stress, the other four dimensions were moderately positively correlated with anxiety, which was consistent with the findings of the DSS in adult samples.[16] Greater digital stress is associated with more anxiety. Therefore, the DSS-C has good criterion-related validity.

Reliability

The results are shown in Table 3. The Cronbach coefficient (α) of the DSS-C is 0.937, and the Cronbach coefficient (α) of the five dimensions is between 0.866 and 0.912, indicating good internal consistency reliability of the DSS-C. At the same time, the DSS-C test-retest reliability is 0.768, and the test-retest reliabilities of the five dimensions are between 0.497 and 0.795. Therefore, the DSS-C has good test-retest reliability.

Table 1 Fit Index of CFA (n = 1374)

	χ^2	df	CFI	TLI	RMSEA	SRMR	AIC	BIC
Five-factor first-order model	1546.031	242	0.938	0.929	0.063	0.044	85,198.030	85,626.520
Five-factor high-order model	1673.766	247	0.932	0.924	0.065	0.051	85,315.765	85,718.128

Table 2 Item Parameters and Factor Loading of the DSS-C and the DSS-C-S (n = 1374)

Item	Dimension	DSS-C						DSS-C-S					
		a	b_1	b_2	b_3	b_4	Factor Loading	a	b_1	b_2	b_3	b_4	Factor Loading
1	Availability stress	3.36	−1.53	−0.65	0.22	1.21	0.83	3.45	−1.52	−0.65	0.22	1.21	0.84
2	Availability stress	4.09	−1.37	−0.47	0.37	1.21	0.88	4.07	−1.37	−0.47	0.37	1.22	0.88
3	Availability stress	3.39	−1.47	−0.58	0.28	1.25	0.83	3.22	−1.49	−0.58	0.28	1.28	0.82
4	Availability stress	1.75	−1.00	−0.23	0.70	1.74	0.64						
5	Approval anxiety	2.94	−1.10	−0.32	0.44	1.37	0.79	2.91	−1.11	−0.32	0.46	1.39	0.79
6	Approval anxiety	4.42	−0.80	−0.07	0.60	1.32	0.87	4.55	−0.80	−0.07	0.61	1.34	0.88
7	Approval anxiety	4.62	−0.71	−0.02	0.65	1.45	0.87	4.80	−0.71	−0.01	0.67	1.47	0.88
8	Approval anxiety	3.53	−0.75	−0.08	0.62	1.45	0.83	3.34	−0.77	−0.08	0.64	1.49	0.82
9	Approval anxiety	2.64	−0.88	−0.15	0.65	1.59	0.76	2.42	−0.91	−0.15	0.69	1.67	0.74
10	Approval anxiety	1.91	−0.99	0.03	0.83	1.88	0.66						
11	FoMO	2.97	−0.23	0.47	1.14	1.75	0.79	2.96	−0.21	0.49	1.16	1.77	0.79
12	FoMO	2.69	−0.57	0.26	1.06	1.84	0.79	2.68	−0.55	0.27	1.09	1.87	0.79
13	FoMO	2.64	−0.45	0.42	1.18	1.89	0.79	2.63	−0.44	0.43	1.21	1.92	0.79
14	FoMO	4.30	−0.29	0.37	1.03	1.64	0.86	4.26	−0.28	0.39	1.05	1.66	0.86
15	Connection overload	2.07	−1.27	−0.20	0.58	1.48	0.68	2.09	−1.26	−0.19	0.57	1.47	0.68
16	Connection overload	2.86	−0.65	0.17	0.96	1.68	0.78	2.89	−0.64	0.17	0.96	1.66	0.78
17	Connection overload	2.93	−0.65	0.12	0.80	1.72	0.77	2.98	−0.64	0.12	0.79	1.70	0.77
18	Connection overload	3.51	−0.72	0.07	0.78	1.59	0.80	3.55	−0.71	0.07	0.78	1.57	0.80
19	Connection overload	2.62	−1.00	−0.07	0.65	1.52	0.73	2.63	−1.00	−0.07	0.65	1.51	0.73
20	Connection overload	2.34	−1.21	−0.06	0.82	1.73	0.72	2.36	−1.20	−0.06	0.82	1.72	0.72
21	Online vigilance	2.69	−1.57	−0.74	−0.07	0.97	0.77	2.67	−1.59	−0.75	−0.08	0.97	0.77
22	Online vigilance	3.76	−1.24	−0.41	0.21	0.96	0.85	3.77	−1.26	−0.42	0.20	0.96	0.85
23	Online vigilance	2.49	−1.94	−0.97	−0.16	0.90	0.75	2.48	−1.97	−0.98	−0.17	0.90	0.75
24	Online vigilance	2.75	−1.15	−0.28	0.41	1.24	0.78	2.76	−1.16	−0.29	0.40	1.24	0.78

Notes: a = item discrimination parameter, b_1, b_2, b_3, b_4 = item location parameters. Factor loading is derived from the five-factor first-order model.

Table 3 Reliability Analysis and Criterion-Related Validity of the DSS-C (n = 997)

	M	SD	Cronbach's α (n = 1374)	Test-Retest Reliability (n = 34)	Total Digital Stress	Availability Stress	Approval Anxiety	FoMO	Connection Overload	Online Vigilance
Total digital stress	65.869	18.484	0.937	0.768**	1					
Availability stress	12.148	3.939	0.868	0.497**	0.552**	1				
Approval anxiety	16.485	6.073	0.912	0.722**	0.864**	0.347**	1			
FoMO	8.872	4.119	0.879	0.578**	0.808**	0.322**	0.692**	1		
Connection overload	15.713	5.659	0.882	0.690**	0.885**	0.332**	0.620**	0.652**	1	
Online vigilance	12.650	4.028	0.866	0.795**	0.718**	0.234**	0.540**	0.412**	0.592**	1
Anxiety	17.703	7.672	0.951		0.601**	0.118**	0.534**	0.563**	0.571**	0.461**

Note: **Indicates p <0.01.

Measurement Invariance

As shown in Table 4, each fitting index of the configural invariance model (M1) is good, indicating that the DSS-C has the same five-factor first-order structure in the urban college student group and the rural college student group, which meets the prerequisite for subsequent invariance analysis. Comparing M1 and M2, it can be found that ΔCFI = 0.001<0.01, ΔCFI = 0.002<0.01, and ΔRMSEA = 0.001<0.01, indicating that weak invariance holds. Then, we constructed a strong invariance model (M3) and compared M2 with M3; at this time, ΔCFI = 0.002<0.01, ΔCFI = 0.000<0.1, and ΔRMSEA = 0.000<0.01, indicating that strong invariance is established.

Table 4 Measurement Invariance Results of the DSS-C and the DSS-C-S (n = 1374)

	Model	x^2	df	CFI	TLI	RMSEA	Model Comparison	ΔCFI	ΔTLI	ΔRMSEA
DSS-C	M1	1882.463	484	0.942	0.934	0.055				
	M2	1887.646	503	0.941	0.935	0.054	M1 vs M2	0.001	0.001	0.001
	M3	1937.679	522	0.939	0.935	0.054	M2 vs M3	0.002	0.000	0.000
	M4	2049.713	546	0.936	0.935	0.054	M3 vs M4	0.003	0.000	0.000
DSS-C-S	M1	1308.873	398	0.959	0.952	0.049				
	M2	1337.479	415	0.958	0.953	0.048	M1 vs M2	0.001	0.001	0.001
	M3	1364.915	432	0.957	0.954	0.048	M2 vs M3	0.001	0.001	0.000
	M4	1467.894	454	0.954	0.953	0.048	M3 vs M4	0.003	0.001	0.000

Notes: M1, configural invariance mode; M2, weak invariance model; M3, strong invariance model; M4, strict invariance model.

Therefore, the strict invariance model (M4) is constructed, and comparing the model with M3, ΔCFI = 0.003<0.01, ΔCFI = 0.001<0.01, ΔRMSEA = 0.000<0.01, strict invariance holds. The five-factor first-order structure of the DSS-C shows cross-group invariance between urban and rural college students. This shows that differences in measurement scores between groups of urban and rural college students can be interpreted as differences in digital stress or its dimensions, rather than as differences in the understanding of the DSS-C meaning.

IRT Analysis
Fit Analysis of the IRT Model
The results are as follows: The fitting indexes of the GRM are AIC = 82,061.79, BIC = 82,688.85, −2Log-Lik = 81,821.80. The fitting indexes of the GPCM are AIC = 82,927.11, BIC = 83,554.17, and −2Log-Lik = 82,687.12. The values of AIC, BIC and −2Log-Lik of the GRM model are all smaller, so the GRM model has better data fitting. Therefore, the psychometric characteristics of the DSS-C will be further examined in the GRM framework.

Item Measurement Characteristics Analysis
The results of item parameter estimation are shown in Table 2. An IRT discrimination parameter below 0.7 indicates a poor item.[43] All items in the DSS-C have high discrimination parameters, ranging from 1.75 to 4.62, indicating that the items could distinguish individuals with similar digital stress (θ) levels.[27] The location parameters of all items are distributed reasonably and widely, which means that the items can distinguish individuals at most θ levels well.[44] However, for individuals with low or very low θ levels, their scores tend to cluster at the low end. The gap between these individuals' scores could not be widened; therefore, the items could not accurately distinguish individuals with very low θ levels from those with low θ levels.

The 24 items in the DSS-C are graded at 5 levels, meaning that each item has 5 ICCs, all of which are displayed in a separate coordinate system (see Supplementary Figure 3). The IIFs of 24 items are placed in different coordinate systems by dimension (see Figure 1) to facilitate a simple and clear comparison of each item's information.

As shown in the ICCs in the appendix, the probability distribution of answers for each option of the 24 items in the DSS-C is reasonable.

As shown in the IIFs in Figure 1, the information values provided by Items 4 and 10 are less than 1 for the participants with all θ levels. Therefore, Items 4 and 10 can be deleted, which not only reduces the length of the test but also has no great influence on the final result.

DIF Analysis
Each item of the DSS-C was tested for gender DIF, which was divided into two levels: male and female. By the pseudo R^2 method, items with a change in R^2 of more than 0.02 had DIF, and the results showed that none of the 24 items in the DSS-C had gender DIF.
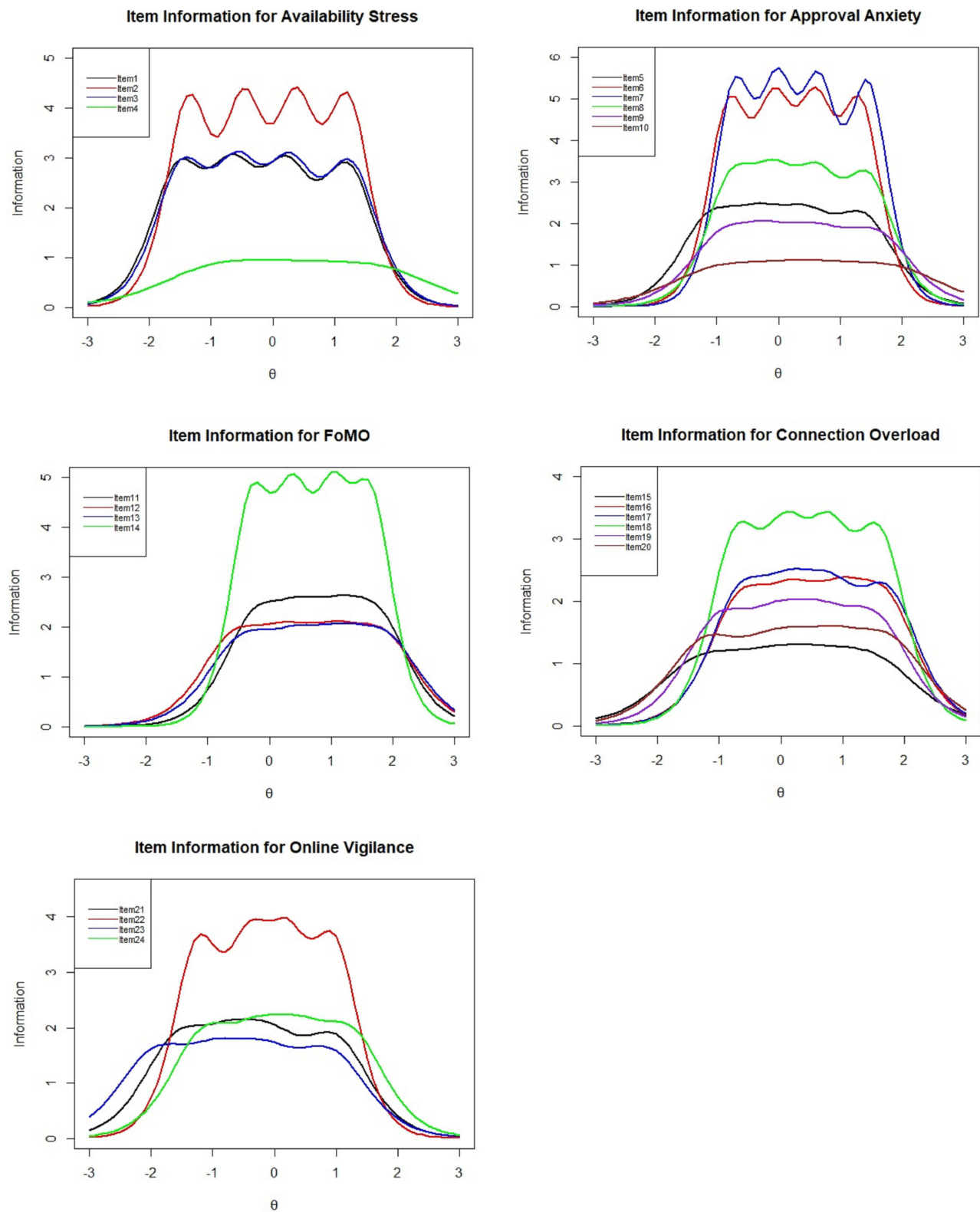
Figure 1 IIFs for each dimension.

Test Measurement Characteristics Analysis

TIF is shown in Figure 2. For the participants with $-1.9<\theta<2.3$, the overall information of the test is above 15; that is, for those with $-1.9<\theta<2.3$, the accuracy of DSS-C is high, while for those with $\theta<-1.9$ and $\theta>2.3$, the accuracy of DSS-C is poor. The TRC is shown in Figure 3. According to TRC, for the participants with $-1.9<\theta<2.3$, the DSS-C reliability is greater than 0.9.
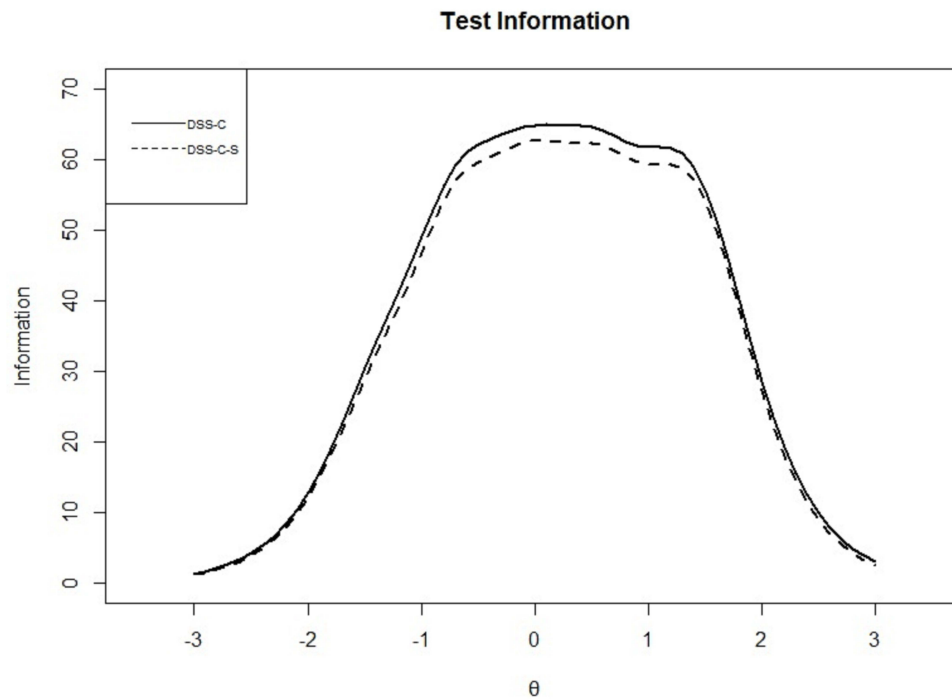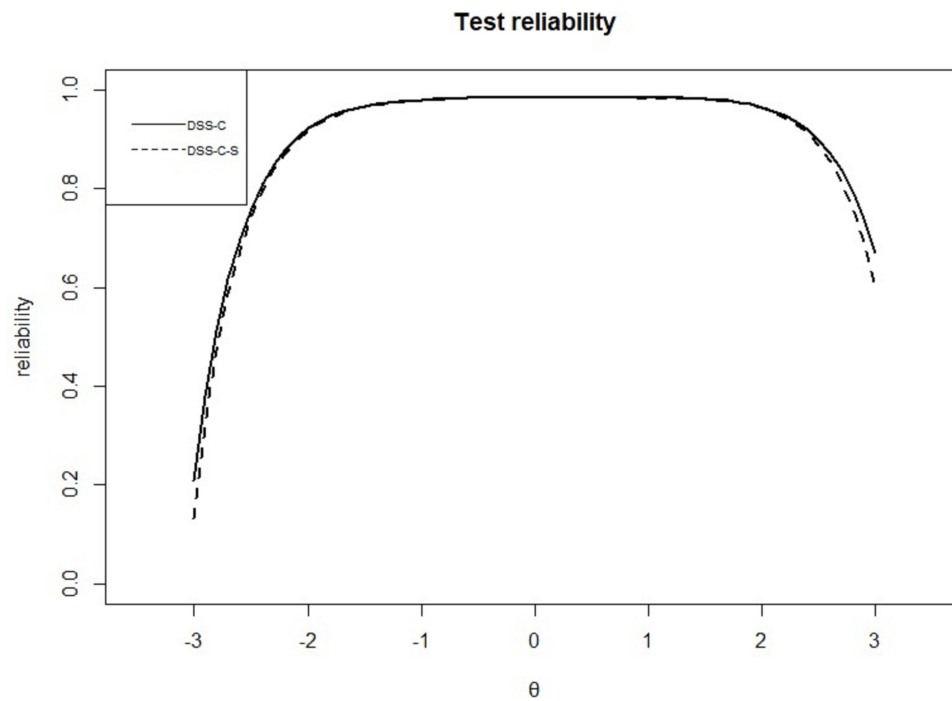
Test Information

Figure 2 TIF of DSS-C and DSS-C-S.

Test reliability

Figure 3 TRC of DSS-C and DSS-C-S.

Construction of Short Form Scale

At least 3 items should be reserved for each dimension when constructing the short form scale, and the reserved items should meet (1) high factor loading; (2) high discrimination parameter; (3) reasonable location parameters; (4) reasonable ICC curve; and (5) no DIF. Since the short-form scale needs to retain sufficient measurement accuracy, items with high information value are especially useful in its construction.[45]

According to the results of the previous analysis, the information values provided by Items 4 and 10 in the dimensions of availability stress and approval anxiety are less than 1 at all θ levels, so Items 4 and 10 are deleted and other items are reserved. The items of other dimensions meet the above requirements, so they are all reserved. In the end, the DSS-C retained 22 items to form the DSS-C-S.

The psychometric characteristics of the DSS-C-S were analyzed, and the short form was compared with the full form. The CFA results showed that the DSS-C-S was well fitted to all indexes of the five-factor first-order model (CFI = 0.956, TLI = 0.949, RMSEA = 0.056, SRMR = 0.036) and had better structural validity than the DSS-C. The Cronbach coefficient (α) of the DSS-C-S was 0.934, and the Cronbach coefficients (α) of availability stress and approval anxiety were 0.886 and 0.911, respectively, which were very close to those of the full form of the DSS-C, and both were higher than 0.80. This indicates that the DSS-C-S also has good internal consistency reliability. The factor loading of DSS-C-S items was higher than 0.65, indicating that each item had a high correlation with its corresponding factor. The DSS-C-S five-factor structure has strict measurement invariance between urban and rural college students in China. The results of the IRT analysis showed that the DSS-C-S was also more suitable for the GRM model, and no gender DIF was found in any items. The DSS-C-S was the same as the DSS-C, and all items had high discrimination parameters ranging from 2.09 to 4.80. The location parameters and ICCs were reasonably distributed, and no items with information less than 1 for all θ levels existed. The comparison of the TIF and TRC of the short form and of the full form is shown in Figures 2 and 3. The test accuracy and reliability of the short form are slightly lower than those of the full form, but there is no significant difference between them. Therefore, the DSS-C-S can be considered a concise and reliable alternative to the DSS-C with good test accuracy and reliability.

Discussion

The Evaluation of the DSS-C

In this study, a Chinese version of the DSS was developed, and CTT and IRT analysis methods were used to analyze the psychometric characteristics of the DSS-C among Chinese college students. In the CTT, the results of the DSS-C in this study are basically consistent with those of previous studies.[16] The five-factor first-order model and five-factor high-order model of the DSS-C are well fitted. The total scores of the DSS-C and four specific dimensions (ie, approval anxiety, FoMO, connection overload and online vigilance) have moderate positive correlations with anxiety score. However, the dimension of availability stress is weakly related to anxiety. This is consistent with the results of Hall et al[1] in a sample of young people. This may be because most young people have established good social relations with their friends, so the experience of availability stress may not be so serious, and the relationship with their anxiety may not be so close.[16] Hall et al[16] found that the experience of availability stress is more obvious among adolescents, because they are in a stage of eagerness to develop social relations with their peers. The Cronbach coefficients (α) of each dimension and the total scale are greater than 0.8. The DSS-C has good reliability and validity. In addition, the five-factor structure of the DSS-C has strict measurement invariance between urban and rural college students in China, which means that although there is a certain degree of development difference between urban and rural digital media in China today, the digital stress experienced by college students still has the same internal structure. This may be because most Chinese universities are located in urban areas. The experience of digital stress is highly dependent on situational contexts.[19] After college students enter universities, the similar digital media contexts in urban areas gradually weaken the possible differences in digital stress. Future research could further explore the measured invariance of the digital stress structure among adolescents who have been attending school in rural or urban areas.

In the IRT, we analyzed the items and test quality of the DSS-C. All items had good discrimination parameters and reasonable location parameters, and there was no gender DIF. It can be seen from the item information curve that Items 4

and 10 provide less information; that is, the parameters of these two items are of poor quality under the IRT framework. Individuals in countries with collectivistic cultures (such as China) pay more attention to maintaining and cultivating relationships with members of other groups,[46] so staying connected with people is advocated as a social obligation in China. The phrase "social obligation" in Item 4 might have led the participants to give the answer of social desirability rather than their actual situation; thus, Item 4 provided less than 1 of information to all θ levels participants. TIF and TRC showed that the information provided by low and high θ levels participants was less than that provided by those with medium θ levels (−1.9<θ<2.3), and the reliability was lower than that provided by those with medium θ levels. This means that the test was slightly less accurate for those with other θ levels than for those with medium θ levels.

In addition, after the completion of this study, we found that Xie et al[47] also revised the Chinese version of the DSS. However, their study still has several limitations which have been well addressed in this study. First, the Chinese version of DSS revised by Xie et al[47] only tested its psychometric characteristics under the CTT framework. CTT is easy to evaluate the overall testing information of individuals and scales, and the parameters and estimation methods are easy to understand and master, but the analysis results are easily influenced by the samples.[48] In IRT, the estimation of ability parameters and item parameters is independent of the sample, and it can accurately estimate the measurement error of each item and test for different levels of individuals.[48] Therefore, more and more scholars suggest using both CTT and IRT to analyze the psychometric characteristics of measurement tools to comprehensively evaluate the applicability of measurement tools.[26,49] Second, the participants of Xie et al[47] were recruited from only two universities by a convenient sampling method, which may reduce representativeness. The applicability of DSS among other Chinese university students needs further testing. Third, the test-retest reliability is important for longitudinal and intervention study designs.[50] But their study did not examine the test-retest reliability of the DSS in measuring digital stress. Finally, studies that rely only on self-reported measures are likely to be affected by CMB,[51] and produce spurious correlations between variables.[52] But this was not tested in their study. In contrast, this study recruited college students from 31 provincial administrative regions in China as participants to comprehensively examine the psychometric characteristics of the DSS-C under the framework of the CTT and IRT, and further validate its applicability.

The Evaluation of the DSS-C-S

Considering the results of CTT and IRT analysis, Items 4 and 10 showed poor item quality. If the scale is as concise as possible, it will help not only to reduce the psychological burden of participants but also to improve the rationality of the study design and the validity of participants' answers and reduce the cost of data collection.[53] Therefore, we deleted the above two items and constructed the DSS-C-S containing 22 items.

The results of CTT showed that the DSS-C-S, similar to the DSS-C, had good reliability and validity and strict measurement invariance between urban and rural college students in China. The results of IRT showed that all items of the DSS-C-S had good discrimination parameters, a reasonable distribution of location parameters, no gender DIF, and a reasonable distribution of the ICC curve. Therefore, the DSS-C-S generally performed well, and the quality of each item was high. Comparing the TIF and TRC of DSS-C and DSS-C-S, it can be found that the trends of the two curves are consistent, and the TIF curve of DSS-C-S is slightly lower than that of DSS-C. As the number of items decreased, so did the information provided by the tests, and the information decreased more at extreme θ levels than at medium θ levels; that is, the extreme θ levels of DSS-C-S were affected more than the moderate θ levels, and the DSS-C-S was almost as reliable as the DSS-C at medium θ levels. Overall, the DSS-C-S is a short and reliable scale that can effectively replace the DSS-C.

Limitations and Outlook

Hall et al[16] tested the reliability and validity of the DSS in adolescents and college students and found that adolescents and college students presented inconsistent results in the criterion association validity. The present study only tested the reliability and validity of the DSS-C in Chinese college students, so it cannot prove whether the DSS-C also has good reliability and validity in Chinese adolescents. With the increasing use of digital media among adolescents, future studies need to further test the reliability, validity and item quality of the DSS-C in Chinese adolescents and develop effective tools to effectively measure digital stress among them.

In addition, a previous study has shown that digital stress can affect not only negative results such as anxiety but also positive results such as happiness and life satisfaction.[4] The present study only verifies the criterion-related validity of the DSS-C from negative perspectives such as anxiety. In future studies, happiness and life satisfaction can be used as criteria to more comprehensively test the criterion-related validity of the DSS-C in the context of Chinese culture.

Conclusions

This study showed that the DSS-C and DSS-C-S have good reliability and validity under the framework of CTT and IRT, and that they can be used for group comparisons across gender and rural/urban areas. The DSS-C and DSS-C-S can be used as the tools for digital stress assessment among Chinese college students. Due to the smaller number of items, the DSS-C-S can be used as a rapid assessment tool for digital stress in a busy large-scale screening.

Statement

The scales used in this study have obtained copyright permission from APA and PROMIS Health Organization. ©2008-2022 PROMIS Health Organization (PHO). This material can be reproduced without permission by clinicians for use with their own patients. Any other use, including electronic use, requires written permission of the PHO.

Ethics Approval and Informed Consent

This study was conducted in accordance with the Declaration of Helsinki, and the questionnaire and methodology for this study were approved by the Research Center of Mental Health of Jiangxi Normal University (Ethics approval number: HM20220180043). Participants provided written informed consent.

Consent for Publication

All authors have approved this manuscript for publication.

Acknowledgments

We acknowledge the individuals who participated in this study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

The authors report no conflicts of interest in this work.

References

1. Reinecke L, Aufenanger S, Beutel ME, et al. Digital stress over the life span: the effects of communication load and internet multitasking on perceived stress and psychological health impairments in a German probability sample. *Media Psychol*. 2017;20(1):90–115. doi:10.1080/15213269.2015.1121832
2. Zhu CL, Ding XH, Lu HZ. Relationship between excessive cellphone use and mental health among teenagers. *Chin J Sch Health*. 2015;31 (11):1618–1620.
3. Shensa A, Escobar-Viera CG, Sidani JE, Bowman ND, Marshal MP, Primack BA. Problematic social media use and depressive symptoms among U.S. young adults: a nationally-representative study. *Soc Sci Med*. 2017;182:150–157.
4. Liu D, Baumeister RF, Yang CC, Hu B. Digital communication media use and psychological well-being: a meta-analysis. *J Comput Mediat Commun*. 2019;24(5):259–273.

5. Hefner D, Vorderer P. Digital stress: permanent connectedness and multitasking. In: Reinecke L, Oliver MB, editors. *Handbook of Media Use and Well-Being*. London: Routledge; 2016:237–249.
6. Thomée S, Dellve L, Härenstam A, Hagberg M. Perceived connections between information and communication technology use and mental symptoms among young adults - A qualitative study. *BMC Public Health*. 2010;10:66.
7. Steele RG, Hall JA, Christofferson JL. Conceptualizing digital stress in adolescents and young adults: toward the development of an empirically based model. *Clin Child Fam Psychol Rev*. 2020;23(1):15–26.
8. Hall J. The experience of mobile entrapment in daily life. *J Media Psychol Theor Methods Appl*. 2017;29:148–158.
9. LaRose R, Connolly R, Lee H, Li K, Hales KD. Connection overload? A cross cultural study of the consequences of social media connection. *Inf Syst Manag*. 2014;31(1):59–73.
10. Halfmann A, Rieger D. Permanently on call: the effects of social pressure on smartphone users' self-control, need satisfaction, and well-being. *J Comput Mediat Commun*. 2019;24(4):165–181. doi:10.1093/jcmc/zmz008
11. Hall JA, Baym NK. Calling and texting (too much): mobile maintenance expectations, (over)dependence, entrapment, and friendship satisfaction. *New Media & Society*. 2012;14(2):316–331. doi:10.1177/1461444811415047
12. Valkenburg PM, Peter J. The differential susceptibility to media effects model. *J Commun*. 2013;63(2):221–243.
13. Wu WL, Wang P, Zhu LJ, Zhang YL, Luo ZR. Reliability and validity test of the multidimensional fear of missing out scale in Chinese college students. *Chin J Clin Psychol*. 2022;30(4):932–935.
14. Perrone MA. *FoMO: Establishing Validity of the Fear of Missing Out Scale with an Adolescent Population [Doctoral Dissertation]*. New York: Alfred University; 2016.
15. Misra S, Stokols D. Psychological and health outcomes of perceived information overload. *Environ Behav*. 2012;44(6):737–759.
16. Hall JA, Steele RG, Christofferson JL, Mihailova T. Development and initial evaluation of a multidimensional digital stress scale. *Psychol Assess*. 2021;33(3):230–242.
17. Cella D, Riley W, Stone A, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63(11):1179–1194.
18. Jin CZ, Li L. Research on spatial heterogeneity patterns of digital divide of Chinese internet. *Econ Geogr*. 2016;36(8):106–112.
19. Nitsch C, Kinnebrock S. Well-known phenomenon, new setting: digital stress in times of the COVID-19 pandemic. *SCM Stud Commun Media*. 2022;10(4):533–556.
20. Ji XQ. An investigation and analysis on the use values of mobile media in contemporary college students. *Educ Mod*. 2020;7(45):185–188.
21. Heng S, Zhao H, Zhou Z. Nomophobia: why can't we be separated from mobile phones? *Psychol Dev Educ*. 2023;39(1):140–152.
22. Reinecke L, Klimmt C, Meier A, et al. Permanently online and permanently connected: development and validation of the online vigilance scale. *PLoS One*. 2018;13(10):e0205384.
23. De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. *Med Educ*. 2010;44(1):109–117.
24. Becker SP, Schindler DN, Luebbe AM, Tamm L, Epstein JN. Psychometric validation of the revised child anxiety and depression scales-parent version (RCADS-P) in children evaluated for ADHD. *Assessment*. 2019;26(5):811–824.
25. Ma J, Batterham PJ, Calear AL, Sunderland M. The development and validation of the thwarted belongingness scale (TBS) for interpersonal suicide risk. *J Psychopathol Behav Assess*. 2019;41(3):456–469.
26. Xie M, Dai B. Evaluation of the psychometric properties of the violent ideations scale and construction of a short form among Chinese university students. *Curr Psychol*. 2022. doi:10.1007/s12144-022-03156-1
27. Reise SP, Henson JM. A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *J Pers Assess*. 2003;81(2):93–103.
28. Barry CT, Sidoti CL, Briggs SM, Reiter SR, Lindsey RA. Adolescent social media use and mental health from adolescent and parent perspectives. *J Adolesc*. 2017;61:1–11.
29. Przybylski AK, Murayama K, DeHaan CR, Gladwell V. Motivational, emotional, and behavioral correlates of fear of missing out. *Comput Hum Behav*. 2013;29(4):1841–1848.
30. Wu FL. *The Development of a Phase-Specific Patient-Reported Outcomes Measurement System-Breast Cancer [Doctoral Dissertation]*. Shanghai: Naval Medical University; 2019.
31. Tang DD, Wen ZL. Statistical approaches for testing common method bias: problems and suggestions. *J Psychol Sci*. 2020;43(1):215–223.
32. Wen ZL, Huang BB, Tang DD. Preliminary work for modeling questionnaire data. *J Psychol Sci*. 2018;41(1):204–210.
33. Gomez R, Watson SD. Confirmatory factor analysis of the combined social phobia scale and social interaction anxiety scale: support for a bifactor model. *Front Psychol*. 2017;8:70.
34. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model Multidiscip J*. 1999;6(1):1–55.
35. Wen ZL, Hau KT, Marsh HW. Structural equation model test: cutoff criteria for goodness of fit index and chi-square test. *Acta Psychol Sin*. 2004;36(2):186–194.
36. Arias VB, Ponce FP, Núñez DE. Bifactor models of attention-deficit/hyperactivity disorder (ADHD): an evaluation of three necessary but underused psychometric indexes. *Assessment*. 2018;25(7):885–897.
37. Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychol Bull*. 1993;114(3):552–566.
38. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Model Multidiscip J*. 2002;9(2):233–255.
39. Muraki E. A generalized partial credit model: application of an EM algorithm. *Appl Psychol Meas*. 1992;16(2):159–176.
40. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika*. 1969;34(1):1–97.
41. McFadden D. The measurement of urban travel demand. *J Public Econ*. 1974;3(4):303–328.
42. Dai B, Zhang W, Wang Y, Jian X. Comparison of trust assessment scales based on item response theory. *Front Psychol*. 2020;11:10.
43. Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res*. 2005;14(10):2277–2291.

44. Reise SP, Ainsworth AT, Haviland MG. Item response theory: fundamentals, applications, and promise in psychological research. *Curr Dir Psychol Sci*. 2005;14(2):95–101.

45. Zhao Y, Kwok AWY, Chan W, Chan CKY. Evaluation of the psychometric properties of the Asian adolescent depression scale and construction of a short form: an item response theory analysis. *Assessment*. 2017;24(5):660–676.

46. Kim KH, Yun H. Cying for me, cying for us: relational dialectics in a Korean social network site. *J Comput Mediat Commun*. 2007;13(1):298–318.

47. Xie P, Mu W, Li Y, Li X, Wang Y. The Chinese version of the digital stress scale: evaluation of psychometric properties. *Curr Psychol*. 2022. doi:10.1007/s12144-022-03533-w

48. Reise SP, Waller NG. Item response theory and clinical measurement. *Annu Rev Clin Psychol*. 2009;5(1):27–48. doi:10.1146/annurev.clinpsy.032408.153553

49. Huang Q, Luo L, Xia BQ, et al. Refinement and Evaluation of a Chinese and Western Medication Adherence Scale for Patients with Chronic Kidney Disease: item Response Theory Analyses. *Patient Prefer Adherence*. 2020;14:2243–2252. doi:10.2147/PPA.S269255

50. Duan W, Mu W, Xiong H. Cross-cultural adaptation and validation of the physical disability resiliency scale in a sample of Chinese with physical disability. *Front Psychol*. 2020;11:602736. doi:10.3389/fpsyg.2020.602736

51. Lindell MK, Whitney DJ. Accounting for common method variance in cross-sectional research designs. *J Appl Psychol*. 2001;86(1):114–121. doi:10.1037/0021-9010.86.1.114

52. Spector PE. Method variance as an artifact in self-reported affect and perceptions at work: myth or significant problem? *J Appl Psychol*. 1987;72:438–443.

53. Chai XY, Li XY, Cao J, Lin DH. A short form of the Chinese positive youth development scale: development and validation in a large sample. *Stud Psychol Behav*. 2020;18(5):631–637.