



The application of item response theory in developing and validating a shortened version of the Rotterdam Emotional Intelligence Scale

Xue Wang¹ · Yan Cai¹ · Dongbo Tu¹

Accepted: 9 June 2022 / Published online: 8 July 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The Rotterdam Emotional Intelligence Scale (REIS) has a good balance between the emotions of self and others, which solves the problem of previous scales being unable to clearly distinguish the emotional intelligence of self from that of others. In the current study, a short version of the REIS scale was developed by the item response theory (IRT) method according to item fit, item or test information function (IIF) and differential item function (DIF) indicators. The results show that the short scale with 14 items had acceptable structural validity (RMSEA = 0.05, CFI = 0.96, TLI = 0.94, SRMR = 0.04) and high reliability ($\omega = 0.90$). The high correlations (more than 0.7) between the scores of the short scale and EIS and WLEIS also verified the concurrent validity of the short scale. In conclusion, compared with the original version of REIS, the developed short version saves 50% of the items and has competitive validity and reliability. The short form will be useful for research and applied contexts where an efficient, concise version is needed.

Keywords Rotterdam Emotional Intelligence Scale · Item response theory · Bifactor model · Coefficient omega

Introduction

The development of reliable, effective and short tools or scales is an important goal of evaluation, especially in educational, psychological and clinical evaluation. In a general way, shorter scales enhance evaluation efficiency in that they can salvage response time and energy, minimize burdens, increase response rate, and reduce fatigue effects. Green and Frantom (2002) pointed out that reducing the length of the scale can reduce the variance and further affect the reliability and validity of scales, and the short and high-quality scale is the most ideal. In contrast to classical test theory (CTT), item response theory (IRT) includes but is not limited to techniques for generating sufficient information and viability at both the item and scale levels; hence, it can be used to accurately exploit and assess short versions of tests, tools, or scales.

Emotional intelligence (EI) is considered as an influential construct in personality and social psychology (Extremera et al., 2011). EI is also defined as a set of abilities, that can

be trained (Hodzic et al., 2017) by which people obtain information from their emotions and use it to guide their thinking and behavior for optimal adaptation (Salovey & Mayer, 1990). Scientific interest in the function of emotional intelligence (EI) in various fields of life is increasing (Blasco-Belled et al., 2019; Joseph & Newman, 2010; Martins et al., 2010). Specifically, EI has shown a positive correlation with a sound body and mind, task performance, and much social contact (Blasco-Belled et al., 2019; Joseph & Newman, 2010; Lopes et al., 2004; Martins et al., 2010; Tejada-Gallardo et al., 2020). As the field advances, researchers are increasingly interested in the processes that underlie the positive effects of EI (Lievens & Chan, 2017). Therefore, an important question is whether dealing with one's own emotions or those of other individuals is equally important for the prediction of criteria (Brasseur et al., 2013; Zeidner et al., 2008). Pekaar et al (2018) propose that both EI dimensions (i.e., dealing with one's own emotions and dealing with others' emotions) may have a positive effect; however, this effect may occur in different life domains. To illustrate, while effective handling of one's emotions presumably plays an important role in staying (mental and physical) health, effective handling of other's emotions may be more important in facilitating smooth social interactions. As the positive effects of EI may thus reflect different processes, it

✉ Dongbo Tu
tudongbo@aliyun.com

¹ School of Psychology, Jiangxi Normal University, 99 Ziyang Ave, Nanchang 330022, Jiangxi, China

may be relevant to differentiate self- from other-focused EI. Previous emotional intelligence scales did not clearly distinguish between self and others, making the influence of the emotional intelligence dimension on both aspects unclear. The Rotterdam Emotional Intelligence Scale (REIS; Pekaar et al., 2018), on the other hand, has a good balance between the emotions of self and others and solves the problem of the previous scales being unable to clearly distinguish between emotional intelligence of the self and the other.

There may be multiple reasons to develop a short version of REIS. First, a short version of the original scale could save time in responding and improve the response rate, which maximizes examinees' representativeness. Second, the short form of the psychological scale evaluates the overall psychological constructs as well as the original version, and they have the same receivable psychometric quality as their original form counterparts, especially validity and reliability (Thalmayer et al., 2011). Compared with the long scale, the short scale is capable of reducing the fatigue effect of the subjects. Furthermore, a short form could be accepted easily for its application in clinical research and its inclusion in studies where multiple scales need to be accomplished. In fact, many authors consider scales of perhaps fifty or more items as having no benefit, especially in studies with multiple scales, repeated measures or targeting persons who might become bored or disengaged (Austin et al., 2018). In terms of theory, a balance should be found in the time spent in responding (long scales generally have preferable psychometric properties with fair-sized territory coverage) and the requirements placed on examinees (e.g., long or repeated responses). Therefore, a shortened form of the REIS has the advantage of minimizing missing data and solves the problem of not clearly distinguishing between self- and other-focused EI.

Based on the above analysis, the main purpose of the research is to develop a short scale. However, to do this well, we first discuss some measurement characteristics of the original scale, which provides a basis for the development of the short version of the scale. On the basis of the above analysis, we then develop a short version of the test and verify its structure and measurement characteristics.

Methods

Participants

A total sample of 1,086 college students participated in this research by responding to the scales. The response data were first sifted for outliers or extremes prior to performing any statistical analysis (i.e., respondents whose response data were missing for 5 consecutive items or 10 consecutive items in the same response). Ninety-five persons were

eliminated from the response dataset on account of missing data or extreme values. The rejection rate was 8.75%, and the effective rate was 91.25%. In this final sample, there were 574 (57.92%) females and 417 (42.08%) males (Data sets can be referred to <https://osf.io/2rus5/>). The respondents' mean age was 19.25 years (SD = 1.23) with a range of 16–25 years. Table 1 reveals the distribution of the sample according to the demographic variables.

In the current study, the total examinees were optionally divided into two pieces: the development examinees (494) and the validation examinees (497), which were both approximately 500. Wolf et al. (2013) suggested that 500 respondents are sufficient for confirmatory factor analysis (CFA). Embreston and Reise (2000) proposed 500 examinees for precise parameter estimates by means of the GRM, and Tsutakawa and Johnson (1990) recommended that the sampling sizes be considered sufficient for precise IRT parameter estimates.

The development sample was used to screen items and validate the structure of REIS's original version. Descriptive statistical analysis and CFA were first conducted on the original scale, and then the original scale was simplified according to standards of item fit, differential item function and other conditions. The verification sample was applied to verify the simplified version of the REIS based on the development sample. For example, validation samples were used to verify the structure of the short scale, and the validation comparison and analysis were conducted with the criterion scale (i.e., EIS and WLEIS).

From Table 1, the development data included more females (57.49%) than males (42.51%), and the majority of participants were science majors (61.34%) from rural areas (64.17%). The average age of all respondents was approximately 19 years. In terms of the validation sample, the same consequences were shown for the demographic characteristics concerning gender, major, census register and age.

Table 1 Demographic characteristics of the participants in development sample and verification sample

Demographic Characteristics	Development sample $N_1 = 494$	Verification sample $N_2 = 497$
Gender		
Males	210(42.51%)	207(41.65%)
Females	284(57.49%)	290(58.35%)
Major		
Arts	191(38.66%)	212(42.66%)
Science	303(61.34%)	285(57.34%)
Census register		
City	177(35.83%)	201(40.44%)
Village	317(64.17%)	296(59.56%)
Age, M(SD)	19.26(1.30)	19.24(1.20)

Instrument

The Rotterdam Emotional Intelligence Scale (REIS)

The REIS is a 28-item Likert scale in which response data are evaluated on a 5-point scale from 0 (strongly disagree) to 4 (strongly agree). Higher scores on the REIS imply higher emotional intelligence levels. Factor analysis of the original scale implied that the REIS conformed to the hierarchical four-factor model, which contains self-focused emotion appraisal (Cronbach's $\alpha=0.82$), other-focused emotion appraisal (Cronbach's $\alpha=0.85$), self-focused emotion regulation (Cronbach's $\alpha=0.80$) and other-focused emotion regulation dimensions (Cronbach's $\alpha=0.82$) (Pekaar, et al., 2018). The full scale's internal consistency reliability was 0.86 (Pekaar, et al., 2018).

Under the principle of keeping the original meaning and being easy to understand, two psychology graduate students translated the scale into Chinese and formed the first draft. Then, the first draft was translated back and discussed by two graduate students majoring in English, and according to their suggestions we modified the first draft. After that, two psychological experts were consulted to make some modifications to the translation and we formed the second draft. Finally, ten graduate students were invited for cognitive interviews to test the Chinese version of REIS to ensure that there was no ambiguity, difficulty in understanding items and other problems, and according to the cognitive interviews we modified some ambiguous item expressions, then the final draft was formed. The internal consistency reliability of the whole scale (Chinese version) and its four subscales in our sample are 0.92, 0.83 (Self-focused emotion appraisal), 0.76 (Other-focused emotion appraisal), 0.77 (Self-focused emotion regulation) and 0.80 (Other-focused emotion regulation dimensions), respectively.

Emotional Intelligence Scale (EIS)

The EIS is constitutive of 33 items applied to evaluate emotional intelligence (e.g., "I have control over my emotions"). The scale is measured on a 5-point Likert scale with a range from 0 (strongly disagree) to 4 (strongly agree); moreover, higher scores imply higher levels of emotional intelligence. The scale was found to be internally consistent (Cronbach's $\alpha=0.9$; Schutte et al., 1998). The Chinese version adopts the scale translated by professor Wang (2002) of South China Normal University (Cronbach's $\alpha=0.84$). EIS's internal consistency reliability in this paper is 0.9.

The Wong-Law Emotional Intelligence Scale (WLEIS)

The WLEIS is a 16-item Likert scale in which response data are evaluated on a 7-point scale from 0 (strongly disagree)

to 6 (strongly agree) by Wong and Law (2002). The Chinese version was translated by Wang (2010) of Central South University (Cronbach's $\alpha=0.9$). WLEIS's internal consistency reliability in this paper is 0.93.

IRT in brief

Compared to the test level in the CTT framework, IRT comprises methods for generating abundant information and greater convenience for item-level applications (Embretson, 1996). A crucial drawback of CTT is that test/item statistics are dependent on a sample of candidates/participants. The other disadvantage is that it is a test-oriented theory, which provides a modicum of information at the item level for a special group of interviewees (Hambleton et al., 1991). IRT models overcome the above shortcomings by performing at the item level and providing sample-free measurements. Next, we introduce the IRT model and item selection indicators used in the study.

The REIS is a Likert multilevel scoring scale, but the multilevel scoring models include the graded response model, generalized partial credit model and nominal model. Therefore, the study compares the three scoring models and selects the most suitable scoring model for the study. The results are shown in Table 3. As shown, the GRM model has the smallest AIC, AICC, SABIC, BIC and $-2\log\text{Lik}$, so this study selected the GRM model as the scoring model.

Graded response model

The graded response model (GRM; Samejima, 1969) is a development of the two-parameter logistic model and analyses polytomous response data, for instance, letter grading and Likert-type scales. In the bifactor case, the graded response model is:

$$P^*(u_{ij} \geq k | \theta_{0i}, \theta_{si}) = \frac{1}{1 + \exp[-(d_k + a_{0j}\theta_{0j} + a_{sj}\theta_{sj})]},$$

where P^* is the probability that a participant with a potential trait of θ has a response equal to k ($k = 0, 1, 2, \dots, K - 1$) or above in item j (Gibbons et al., 2007; Mao et al., 2018). Therefore, the probability on the k is equal to the difference between the cumulative probability on the two adjacent categories:

$$P(u_{ij} = k | \theta_{0i}, \theta_{si}) = P^*(u_{ij} \geq k | \theta_{0i}, \theta_{si}) - P^*(u_{ij} \geq k + 1 | \theta_{0i}, \theta_{si}).$$

The model has a permit of items to contain a different number of categories, and each item is represented by a general slope parameter (a_{0j}), a special slope parameter (a_{sj}) and by numerous threshold parameters ($d_k, d_k = -(a_{0j}b_{jk} + a_{sj}b_{jk})$) that are one less than the number of categories within the item (Gibbons et al., 2007; Mao

et al., 2018). These threshold parameters describe the location with a probability of more than 50% on the θ scale, and the response is in the related or the higher category.

Item fit

The $S-X^2$ (Orlando & Thissen, 2000) statistic was applied here to check the item fit. In this method, the item fitting statistic is based on the observed and expected frequency of correct and incorrect scores for each summary. The prominent advantage of this method is that these expected frequencies can be directly compared with the frequencies observed in the data. The significance level of $S-X^2$ indicators can be set as 0.01 according to requirements. When the p value of $S-X^2$ of the item is less than the significance level, the item fit is poor and should be considered for deletion.

Test and item information function

In the IRT framework, reliability (the degree of error-free measurement) is evaluated by the item or test information function (IIF or TIF). The amount of information provided by each item is a function of the trait level of the subjects tested. The IIF or TIF varies with the value of the trait level. The item provided the most information when the subject's trait level matched the item's threshold value. What determines the overall level of information provided by an item at the competency trait level is also the parameter characteristics of the item itself. Among them, the most critical factor that affects the amount of information of an item is the item discrimination parameter. The larger the item discrimination parameter is, the more information it provides. The formula of the item information function is as follows:

$$\text{IIF}(\theta) = \frac{[p'_j(\theta)]^2}{p_{j(\theta)}(1 - p_{j(\theta)})}$$

where θ is the measured trait, $p_j(\theta)$ is the subject response function of item j , and $p'_j(\theta)$ is the first derivative function of $p_j(\theta)$. The IIF shows that the item is more helpful in measuring the level of respondents within the θ (Reeve & Fayers, 2005) and indicates the item's contribution to the evaluation of traits. IIF is also an index to appraise item quality (it can distinguish the abilities of respondents) and is therefore used to appraise the relative performance of each item. TIF is the aggregation of the information functions of the items contained in the test.

Differential item function

DIF is a process of evaluating whether the item is biased against a certain group and comprehensively evaluating

the fairness of the test. DIF not only affects the estimation of the ability of the subjects but also affects the validity and fairness of the test (Kim, 2001). When members of different groups have different item responses, the differences should be due to variations in the underlying trait, not to unintended, construct-irrelevant factors. In the case of gender, for example, men and women with the same trait level should have the same predicted response. When the expected response varies for groups with the same trait level, the item is said to show DIF and is, hence, biased in favor of one group (Penfield & Camilli, 2006). Criteria based fully on statistical significance can detect DIF items that are not actually relevant (Crane et al., 2007). We chose the following criteria for meaningful differential item functioning: the logistic regression square and McFadden's pseudo R^2 (Crane et al., 2006; Prati & Pietrantonio, 2016). When the change in R^2 is greater than 0.02, it indicates the existence of deviation in the item and should be deleted (Crane et al., 2006; Gu & Wen, 2017). In the current study, we conducted DIF analysis of gender to exclude the influence of factors related to personal background on the item.

Confirmatory factor analysis

The development sample and validation sample used confirmatory factor analysis to verify the structure of the original and the short version of REIS, respectively. We used the development sample to conduct CFA for the original version, while we used the validation sample to conduct CFA for the short version. Using CFA, we compared the consistency of six models: the original hierarchical four-factor structure (Pekaar, et al., 2018), the hierarchical five-factor model, the hierarchical six-factor model and their corresponding bifactor structure models. The bifactor model permits researchers to empirically determine which items accurately describe general dimensions or conceptually narrower dimensions after regulating the general factor (Gibbons et al., 2007; Li et al., 2021; Reise et al., 2010). That is, in the bifactor model, each term is loaded on both the general and the special dimensions, both of which are orthogonal. The fit of the proposed models was evaluated by four indices: the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean squared error of approximation (RMSEA), and the standardized root mean squared residual (SRMR). The fit indices were interpreted using values that should be close to 0.95 or higher, indicating an excellent or at least acceptable (> 0.90) model fit for CFI and TLI, an approach to 0.06 for RMSEA, or an approach to 0.08 for SRMR (Hu & Bentler, 1999; MacCallum et al., 1996).

In addition, based on the bifactor model, three types of reliability indicators can be calculated by omega (ω), omega subscale (ω_s) and omega hierarchical (ω_H). The coefficient omega is expressed by ω , also known as internal consistency

reliability, which reflects the correlation between all the items. When ω is above 0.70, the reliability of the synthesis score of the multidimensional test is acceptable. The omega subscale (ω_s) is the same as the coefficient omega (ω); the higher the ω_s , the higher the reliability of the subscale composite score. The coefficient omega hierarchical, expressed in ω_H , reflects the extent to which all items measure the same trait, and the larger the ω_H , the more homogeneous the test. When ω_H is greater than 0.50, the homogeneity can be considered high, in which case it makes sense to synthesize the whole test score (Gu & Wen, 2017; Rodriguez et al., 2016). Both CFA and Omega coefficient results are considered in the study.

Procedure

Before the study, the scales' purposes and regulations were explained, and we obtained students' counselors' consent to conduct the study in their classes. We assured respondents that all answers were confidential and that nobody in their classes would have access to the personalized data. In addition, respondents were notified that they could refuse to participate in the study anytime, and if they did not want to fill out the scales, they could fill in blanks.

Statistical analysis was performed using R (R Core Team, 2014) and M-plus (Wang, 2014). We conducted confirmatory factor analysis with M-plus and ran a GRM using the MIRT package (Chalmers, 2012). We analyzed item fitting, discrimination, and test and item/test information function to evaluate the psychometric characteristics of the items and tests. The sex-based DIF was also examined via the lordif package (Choi et al., 2011).

The process of simplifying the scale is as follows: first, the item that does not fit according to item fit is deleted; next, the item with DIF is deleted; then, according to item discrimination and IIF, the items that do not meet the requirements are deleted. Furthermore, the test information and correlation of REIS and two related scales (i.e., EIS and WLEIS) were analyzed through the short version scale to verify its performance.

Results

Descriptive statistics and CFA of the original REIS

Table 2 shows the descriptive statistics of each item of the original REIS with the development sample ($N_j = 494$), including the mean score, standard deviations and item response frequencies. For all items, the majority of respondents (more than 40%) agree in favor of the category (option 4), and for some items (e.g., item 5 and item 20), a relatively large proportion of those supporting the category strongly

agree, while few strongly disagree (option 1). The full-scale average was 71.51, with a standard deviation of 13.11 and a range from 11 to 110.

In order to better explore the structure of the Chinese version of the REIS, we randomly divided the development sample into sub-Sample1 ($N = 244$) and sub-Sample2 ($N = 250$). Firstly, exploratory factor analysis (EFA) is performed on sub-Sample1 to analyze the structure of Chinese REIS. The results show that the Kaiser–Meyer–Olkin (KMO) and Bartlett's spherical test is 0.91 and the hierarchical six-factor structure was extracted according to the criterion that the eigen value was greater than 1. The six-factor structure was optimized for 59.65% of the total variance. Then, we performed CFA for the hierarchical six-factor structure by sub-Sample 2, and the result was not ideal (RMSEA = 0.07, CFI = 0.82, TLI = 0.80, SRMR = 0.07). We believed that the hierarchical structure was not ideal because there were too few items on two factors, with only 3 and 2 items, respectively. Therefore, we combined the two factors into one factor because they had similar content and reperformed the CFA of the hierarchical five-factor, but the result was still unsatisfactory: RMSEA = 0.07, CFI = 0.82, TLI = 0.80, SRMR = 0.08. In addition, CFA was conducted on the original scale's hierarchical four-factor structure by sub-sample 2 (Pekaar, et al., 2018), but the fit index values did not fit well (RMSEA = 0.07, CFI = 0.78, TLI = 0.76, SRMR = 0.08).

Since bifactor models are widely used to solve dimension analysis problems in behavioral science and other related fields, we try to construct bifactor models with four special factors, five special factors and six special factor structures. Among them, the bifactor model with four special factor structures was the most ideal (RMSEA = 0.07, CFI = 0.83, TLI = 0.80, SRMR = 0.07), while the bifactor model with five special factor structures and the bifactor model with six special factor structures could not be identified. The model comparison results are shown in Table 3. Compared with other structures, the fitting index of the bifactor model with four special factor structures was the best, and the results of AIC, SABIC, BIC and $-2\log\text{Lik}$ were also the smallest, indicating that the data fit the bifactor model with four special factor structures. Therefore, the following studies are based on a bifactor model with four special factors.

Table 4 presents the coefficient omega, coefficient omega hierarchical and omega subscale of the REIS based on the bifactor model. The coefficient omega (ω) of the whole scale is 0.92, indicating that the reliability of the total scale is above 0.70, which meets the psychometric standard. The omega subscales (ω_s) of the four subscales are 0.81, 0.75, 0.79 and 0.76, respectively, which are all greater than 0.7, indicating a highly reliable multidimensional composite scale. The coefficient omega hierarchical (ω_H) is 0.83 (above 0.50), which indicates that the total score is meaningful.

Table 2 Abbreviated item content and response percentages of the 28 Items of the REIS ($N_f = 494$)

No	Item content	Scale value (%)				
		0	1	2	3	4
Self-focused emotion appraisal						
1	I always know how I feel	1.20	4.10	27.50	51.80	15.40
2	I know which emotions I experience	1.00	2.60	20.50	50.00	25.90
3	Mostly, I am able to explain exactly how I feel	1.20	7.30	29.40	45.30	16.80
4	I am aware of my own emotions	1.20	3.20	18.80	54.30	22.50
5	I understand why I feel the way I feel	0.80	3.20	22.10	54.90	19.00
6	I can judge well if events touch me emotionally	2.00	9.30	38.50	41.10	9.10
7	I can distinguish my own emotions well	1.20	5.30	27.90	52.80	12.80
Other-focused emotion appraisal						
8	I am aware of the emotions of the people around me	1.20	3.60	22.10	52.50	20.60
9	I know which feelings others experience	2.40	5.30	43.10	40.50	8.70
10	I understand why other people feel the way they feel	1.60	6.30	35.00	46.60	10.50
11	I can judge well if events touch others emotionally	1.20	7.90	34.20	43.70	13.00
12	I can empathize with the people around me	1.00	6.30	33.00	47.10	12.60
13	When I look at other people, I can see how they feel	2.60	8.30	50.40	34.00	4.70
14	I can distinguish well between other people's emotions	2.40	15.40	41.70	32.00	8.50
Self-focused emotion regulation						
15	I do not let my emotions take over	2.40	14.80	39.10	34.40	9.30
16	I only show my emotions when it is appropriate	1.20	6.50	25.10	49.00	18.20
17	If I want to, I put on my poker face	4.30	12.90	22.90	38.00	21.90
18	Even when I am angry, I can stay calm	4.50	13.70	32.60	34.60	14.60
19	I am in control of my own emotions	1.20	7.30	35.40	41.50	14.60
20	I adjust my emotions when necessary	0.00	2.60	16.20	55.30	25.90
21	I can suppress my emotions easily	3.20	14.80	41.30	32.40	8.30
Other-focused emotion regulation						
22	I can boost or temper the emotions of others	1.60	5.50	35.40	45.40	12.10
23	I have great influence on how others feel	4.50	13.50	51.40	24.90	5.70
24	I know how to influence people	3.60	10.90	47.00	30.20	8.30
25	I can make someone else feel differently	2.00	7.50	44.50	36.70	9.30
26	I am able to calm others down	1.80	11.10	51.80	30.40	4.90
27	I can alter another person's emotional state	1.60	13.80	55.70	24.60	4.30
28	I know what to do to improve people's mood	2.20	8.70	40.90	41.10	7.10

Development and verification of the short version of REIS

Development of the REIS short form

The GRM with a bifactor model was used for calibrating the 28 items of the REIS utilizing the data from the development sample. The factor load path diagram of the original scale can be seen in the [Appendix](#). The item parameters of GRM, item information, IIFC, model fit and DIF are displayed in [Table 5](#) and [Fig. 1](#), respectively. Five slope parameters (a, a_1, a_2, a_3 and a_4) and four threshold parameters (b_1, b_2, b_3 and b_4) were estimated for each five-category item. The discrimination parameter measures the strength

of the relationship between the item and the construct being measured—the stronger the relationship, the better the item. In addition, it is used as an indicator of item quality—the higher the slope, the more discriminating the item. The value range of general factor discrimination (a) of the original scale is from 0.97 to 2.33. The slope parameter a_1 ranges from -0.09 to 1.68, a_2 ranges from -0.20 to 1.37, a_3 ranges from 0.24 to 1.37, and a_4 ranges from 0.64 to 1.33, which shows deviation in the discrimination values of the items.

The threshold parameters (b_1, b_2, b_3 and b_4) represent four locations on the θ scale (REIS). The parameter b_1 represents the location where the probability exceeds 50% that the response is in the category strongly disagree or in the higher category disagree. Similarly, b_2 represents the location

Table 3 Comparison of multidimensional item response theory models using confirmatory factor analyses (sub-Sample 2)

Model		AIC	SABIC	BIC	-2LogLik
GRM	bifactor model with four special factors	15464.89	15523.26	15523.26	15128.89
	hierarchical four-factor	15907.17	15955.80	16399.61	15627.17
	hierarchical five-factor	15925.30	15973.93	16417.74	15645.30
	hierarchical six-factor	15964.48	16013.11	16456.92	15684.48
GPCM	bifactor model with four special factors	15584.68	15643.04	16175.61	15248.68
	hierarchical four-factor	16014.57	16063.21	16507.02	15734.57
	hierarchical five-factor	16013.78	16062.42	16506.22	15733.78
	hierarchical six-factor	16053.23	16101.87	16545.67	15773.23
Nominal	bifactor model with four special factors	15607.02	15694.56	16493.42	15103.02
	hierarchical four-factor	16018.61	16096.43	16806.52	15570.62
	hierarchical five-factor	16020.90	16098.72	16808.81	15572.90
	hierarchical six-factor	16064.43	16142.25	16852.34	15616.43

GRM Graded response model; GPCM Generalized partial credit model; Nominal Nominal model; AIC Akaike’s information criterion; BIC Bayesian information criterion; SABIC Sample-Size adjusted BIC; -2LogLik -2*Log-Likelihood

Table 4 Omega (ω), Omega subscale (ω_s) and Omega hierarchical (ω_H) for the full Scale (sub-Sample 2)

Dimension	ω_s	ω	ω_H
1-Self-focused emotion appraisal	0.81	0.92	0.83
2-Other-focused emotion appraisal	0.75		
3-Self-focused emotion regulation	0.79		
4-Other-focused emotion regulation	0.76		

where the probability exceeds 50% that the response is in the category disagree or in the higher category agree; and b_3 represents the location where the probability exceeds 50% that the response is in the category not sure. Finally, b_4 represents the location where the probability exceeds 50% that the response is in the category agree or in the higher category strongly agree. The values of the four threshold parameters (b_1, b_2, b_3 and b_4) were very similar to all items. According to Dodeen and Darmaki (2016), if the intercept parameter ('difficulty' parameter) of all items shows a consistent trend, this parameter is not helpful in distinguishing between items. Parameter b_1 ranges from -4.58 to -2.01 for all items, parameter b_2 is greater than b_1 (ranging from -2.84 to -1.06) for all items, parameter b_3 ranges from -0.82 to 1.04, and b_4 ranges from 0 to 2.57. These results show how the items in the scale work. Over the scale items, respondents with high EI tend to select strongly agree, while respondents who have lower EI tend to select strongly disagree. Because the results of these items were similar, the intercept parameter ('difficulty' parameter) had no effect on filtering items. The difficulty of the 28 items is shown in Fig. 2, which ranges from 0.42 to 0.71, meaning that the test can maximize the difference between individuals.

The mean and standard deviation of IIF of the items were 0.16 to 1.23 and 0.02 to 0.70, respectively. The value of $S-X^2$ ranges from 54.06 to 113.59, with a p value between 0.04 and 0.83, and the R^2 change degree of DIF is between 0.00 and 0.02. In the self-focused emotion appraisal dimension, items 2, 4, 5 and 7 provide significantly more information than items 1, 3 and 6. In the other-focused emotion appraisal dimension, items 8 and 9 provide more information than other items. In the self-focused emotion regulation dimension, items 18, 19 and 21 obviously provide more information than the other items. In the other-focused emotion regulation dimension, items 24, 25, 26 and 27 provide more information than the remaining items (see Table 5).

Based on the above information, four indicators were used for selecting the suitable items for the short form of the scale: item fit (p value of $S-X^2$), DIF, item discrimination (a) and IIF. First, all items fit the GRM well, with p values of $S-X^2$ less than 0.01. Therefore, the item-fit index does not delete items. According to the second judgment criteria, only item 15 has DIF, whose R^2 change is greater than 0.02, and then we delete item 15.

Item discrimination is a very important index for item quality under the framework of IRT. According to the item slope (a) parameter, the item with the highest slope value was identified. Considering the slope parameters of the general and special factors together, the self-focused emotion appraisal dimension was reserved for items 2, 4, 5 and 7; the other-focused emotion appraisal dimension was reserved for items 8 and 9; the self-focused emotion regulation dimension was reserved for items 18, 19 and 21; and the other-focused emotion regulation dimension was reserved for items 23–28. Existing studies indicate that at least three items should be reserved for each dimension to

Table 5 IIF, item parameters, model-fit($S-X^2$, df , p) and R^2 for the 28 Items of the REIS ($N_I = 494$)

Item	Mean(IIF)	SD	a	a1	a2	a3	a4	b1	b2	b3	b4	$S-X^2$	df	p	R^2
1	0.28	0.05	1.09	0.48				-4.21	-2.84	-0.72	1.73	58.41	63.00	0.64	0.00
2	0.78	0.37	1.84	1.05				-2.99	-2.20	-0.82	0.78	78.34	59.00	0.05	0.00
3	0.59	0.20	1.57	0.74				-3.20	-1.81	-0.40	1.28	82.78	71.00	0.16	0.00
4	1.23	0.70	2.33	1.60				-2.64	-1.87	-0.76	0.83	70.34	58.00	0.13	0.00
5	1.22	0.64	1.98	1.68				-2.91	-1.99	-0.69	0.99	66.68	60.00	0.26	0.00
6	0.45	0.15	2.26	-0.09				-2.46	-1.35	0.00	1.61	72.64	61.00	0.15	0.01
7	0.92	0.39	2.25	0.89				-2.71	-1.72	-0.44	1.32	67.44	53.00	0.09	0.00
8	0.92	0.44	1.38		1.31			-2.97	-2.15	-0.78	1.07	80.83	65.00	0.09	0.01
9	0.92	0.35	1.63		1.37			-2.49	-1.74	0.03	1.74	54.81	59.00	0.63	0.00
10	0.55	0.14	1.63		0.72			-2.99	-1.84	-0.22	1.70	76.97	65.00	0.15	0.00
11	0.25	0.07	1.81		-0.20			-3.06	-1.72	-0.21	1.50	55.21	65.00	0.80	0.00
12	0.20	0.02	1.07		0.28			-4.58	-2.65	-0.42	2.06	86.78	72.00	0.11	0.01
13	0.45	0.06	1.45		0.51			-2.88	-1.75	0.39	2.57	60.57	57.00	0.35	0.00
14	0.30	0.07	1.79		-0.06			-2.67	-1.17	0.30	1.85	82.26	69.00	0.13	0.00
15	0.57	0.13	1.25			1.09		-2.80	-1.28	0.20	1.85	83.20	78.00	0.32	0.02
16	0.34	0.08	1.12			0.66		-3.84	-2.24	-0.66	1.47	84.63	75.00	0.21	0.00
17	0.16	0.03	0.97			0.16		-3.47	-1.79	-0.47	1.50	113.59	95.00	0.09	0.00
18	0.99	0.49	1.76			1.35		-2.01	-1.06	0.01	1.26	113.23	85.00	0.02	0.00
19	1.12	0.50	1.99			1.37		-2.63	-1.53	-0.16	1.24	54.06	65.00	0.83	0.01
20	0.24	0.07	1.27			0.24		-3.18	-1.34	1.04	0.00	78.25	61.00	0.07	0.01
21	0.80	0.26	1.64			0.98		-2.35	-1.11	0.31	1.84	71.59	71.00	0.46	0.00
22	0.39	0.08	1.29				0.64	-3.39	-2.15	-0.25	1.80	80.77	67.00	0.12	0.00
23	0.51	0.10	1.08				0.96	-2.62	-1.31	0.75	2.46	73.24	69.00	0.34	0.00
24	0.89	0.33	1.71				1.22	-2.19	-1.21	0.35	1.74	66.14	70.00	0.61	0.00
25	0.73	0.22	1.47				1.12	-2.74	-1.63	0.13	1.76	90.67	69.00	0.04	0.00
26	0.73	0.17	1.37				1.13	-2.93	-1.46	0.50	2.32	81.25	64.00	0.07	0.00
27	0.79	0.18	1.34				1.33	-2.90	-1.25	0.71	2.35	74.23	65.00	0.20	0.00
28	0.65	0.16	1.37				1.05	-2.79	-1.57	0.08	2.05	79.29	63.00	0.08	0.00

SD Standard deviation; *IIF* Item information function; R^2 R-squared change for pseudo R-squared criterion

ensure the reliability and validity of the subscale (Prati & Pietrantonio, 2016). Hou et al. (2004) also showed that each dimension requires at least three items to ensure model identification. Therefore, we set the slope parameter of the other-focused emotion appraisal dimension to 0.72 (below 0.80) but retained item 10, which contains more information. That is, the other-focused emotion appraisal dimension and the self-focused emotion regulation dimension each contain 3 items (items 8–10, item 18, item 19 and item 21), and items are no longer deleted according to the amount of information.

Next, we remove the items with content redundancy by considering the IIF and the item information function curves. The rule of IIF is that the more information the item gives, the better; then we selected the items to maximize the total amount of information across the entire continuum with minimal content overlap. Specifically,

in the case of content redundancy, we selected items that reached high levels of information or that were the highest at the higher or lower end of the continuum and, therefore, were informative for differentiating among participants with high or low scores on a scale. As seen from Fig. 1 and Table 5, items 2, 4, 5 and 7 can provide the maximum amount of information for the subjects in the inability range. Therefore, these four items are retained within the self-focused emotion appraisal dimension. In the other-focused emotion regulation dimension, it can be found from Fig. 1 that the item information provided by items 24–27 overlaps with that provided by item 23 and 28. Combined with the item information provided by Table 5, we choose item 24–27 that can provide higher item information. Based on the above analysis, we retained 14 items (2, 4, 5, 7, 8, 9, 10, 18, 19, 21, 24, 25, 26, and 27) as short versions of REIS.

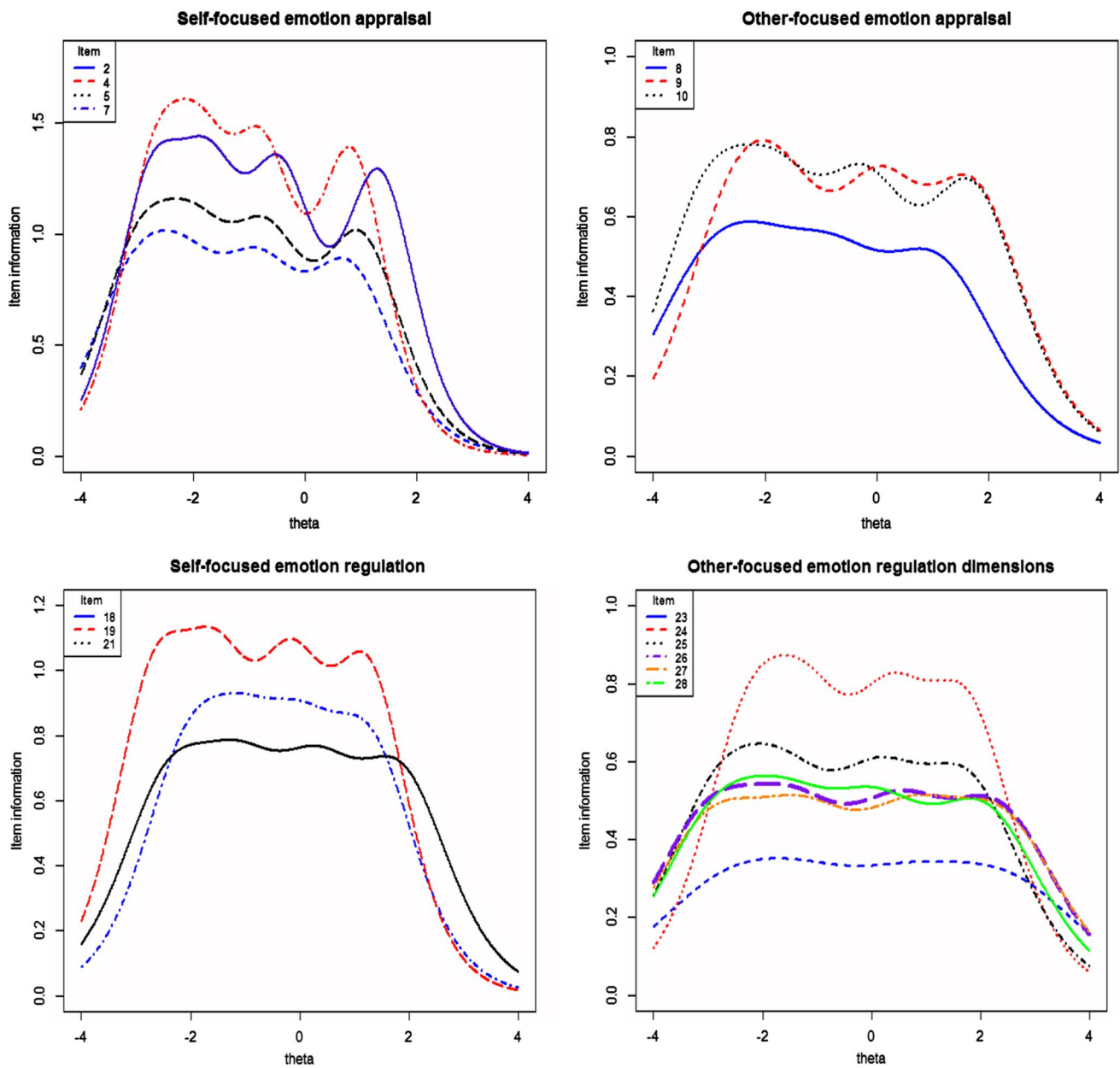


Fig. 1 Item information function curves of the items ($N_I=494$)

Validation of the REIS short form

According to the above analysis, the final short scale of the REIS contains fourteen items. The short scale fits the bifactor model well (CFI = 0.96, TLI = 0.94, RMSEA = 0.05, SRMR = 0.04), which was higher than the full scale (CFI = 0.88, TLI = 0.86, RMSEA = 0.06, SRMR = 0.05). Table 6 reports the short-scale reliability coefficients. The coefficient omega (ω) of the whole scale is 0.90 (above 0.70), which meets the standard of

psychometrics. The omega subscales (ω_s) of the four subscales are 0.80, 0.71, 0.77 and 0.73, respectively, which are all greater than 0.7, indicating a highly reliable multidimensional composite scale. The coefficient omega hierarchical (ω_H) is 0.77 (above 0.50), which indicates that the total score is meaningful.

The standardized factor load coefficient of the shortened scale is shown in Fig. 3 with the validation sample. In the general factor, the load value was greater than 0.4. The value range of general factor discrimination (a) of

Fig. 2 Item difficulty distribution curve of 28 items ($N_j=494$)

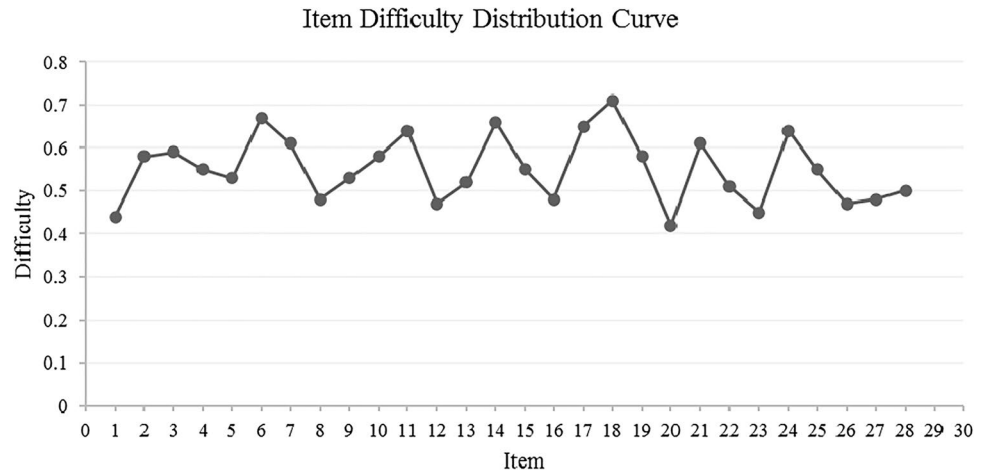
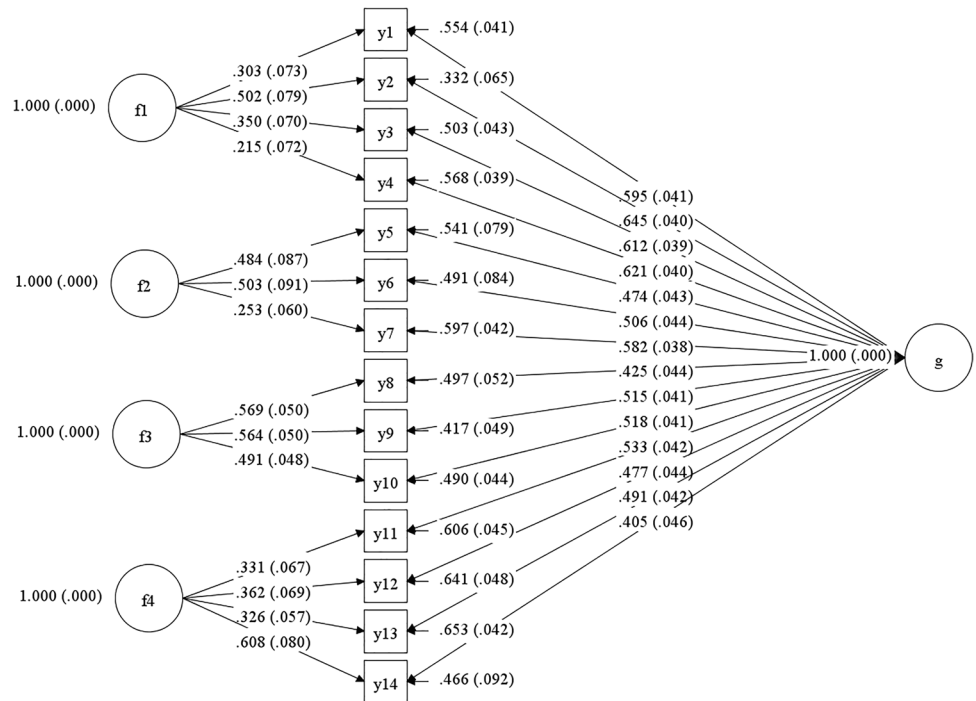


Table 6 Comparison of Omega (ω), Omega subscale (ω_s), Omega hierarchical (ω_H) and CFA of short version ($N_2=497$)

Version	ω_H	ω_s				CFA				
		ω	one	two	three	four	RMSEA	CFI	TLI	SRMR
REIS-short-14	0.77	0.90	0.80	0.71	0.77	0.73	0.05	0.96	0.94	0.04

one = Self-focused emotion appraisal; two = Other-focused emotion appraisal; three = Self-focused emotion regulation; four = Other-focused emotion regulation dimensions

Fig. 3 Path coefficient of the 14 Items of the REIS ($N_2=497$). Note: All the path coefficients in the figure are statistically significant at the level of 0.01. f1 = Self-focused emotion appraisal; f2 = Other-focused emotion appraisal; f3 = Self-focused emotion regulation; f4 = Other-focused emotion regulation



the short scale is from 1.23 to 2.67. The slope parameter a_1 ranged from 0.69 to 1.89, a_2 ranged from 0.65 to 1.65, a_3 ranged from 1.46 to 1.81, and a_4 ranged from 0.78 to 1.54, which shows that the items of the short scale were highly differentiated and could well distinguish subjects

with different abilities. The values of the four threshold parameters (b_1, b_2, b_3 and b_4) were very similar across all items (see Table 7). The difficulty of the short scale is shown in Fig. 4, which ranges from 0.46 to 0.69, meaning that the test can maximize the difference between

Table 7 Means and standard deviations of IIF, Slope (*a*) Parameter and Location (*b*) Parameter Estimates for the 14 Items of the REIS ($N_2=497$)

Item	Mean(IIF)	SD	a	a1	a2	a3	a4	b1	b2	b3	b4
1	0.77	0.33	1.80	0.99				-3.86	-2.21	-0.80	1.02
2	1.57	1.00	2.67	1.89				-2.76	-1.93	-0.81	1.01
3	1.00	0.46	1.99	1.27				-3.32	-1.95	-0.72	1.07
4	0.66	0.21	1.79	0.69				-3.17	-2.11	-0.48	1.61
5	0.77	0.29	1.40		1.36			-3.81	-2.29	-0.89	1.29
6	1.05	0.38	1.58		1.65			-2.70	-1.64	0.17	1.84
7	0.55	0.13	1.53		0.65			-3.90	-1.99	-0.14	1.87
8	0.96	0.42	1.31			1.73		-2.47	-1.31	-0.08	1.39
9	1.21	0.52	1.76			1.81		-2.90	-1.56	-0.21	1.47
10	0.99	0.33	1.57			1.46		-2.56	-1.01	0.45	1.97
11	0.71	0.16	1.41				1.08	-3.18	-1.47	0.36	2.22
12	0.70	0.18	1.29				1.19	-3.18	-1.80	0.30	1.96
13	0.50	0.08	1.23				0.78	-3.30	-1.73	0.69	2.47
14	0.85	0.24	1.26				1.54	-2.84	-1.40	0.74	2.26

IIF Item information function

individuals. Over the scale items, respondents with high EI tend to select strongly agree, while respondents with lower levels of EI tend to select strongly disagree, which is the same as the full scale.

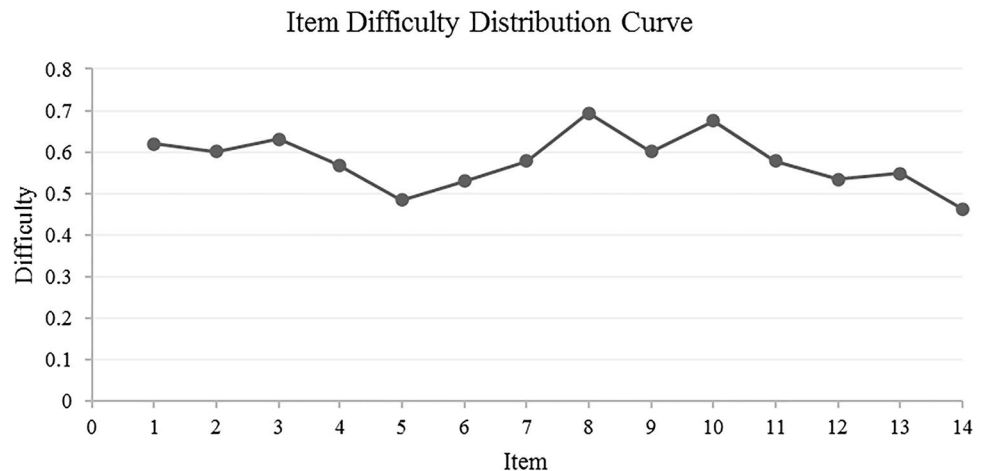
The IIF mean and standard deviation ranged from 0.50 to 1.57 and 0.08 to 1, respectively. In addition, we also calculated $S-X^2$ and DIF for the short scale, and the results showed that the items of the short scale all met these two standards.

To compare the long and shorter versions of REIS, we calculated and compared two forms of TIF. The data are graphically represented by two curves in Figs. 5 and 6. Each curve represents the information about the emotional intelligence levels (θ) offered by each version. The figure clearly shows that, of all the theta values,

the long version of REIS supplies further information than the short form. Additionally, the mean of the TIF of the long form was 18.13, while the mean of the TIF of the short form was 12.29. This is an expected result, since TIF is the collection of all IIFs, and there are more items in the long version; thus, we expect to obtain more information.

Finally, two criterion scales were used to verify the short version of the REIS: the EIS and the WLEIS. Based on the results of coefficient omega analysis, we only conducted correlation analysis on the scale total scores of REIS, EIS and WLEIS and did not conduct comparative analysis among subscales. The correlations between the EIS, WLEIS and REIS using the long REIS version were compared with those using the short

Fig. 4 Item Difficulty Distribution Curve of 14 items ($N_2=497$)



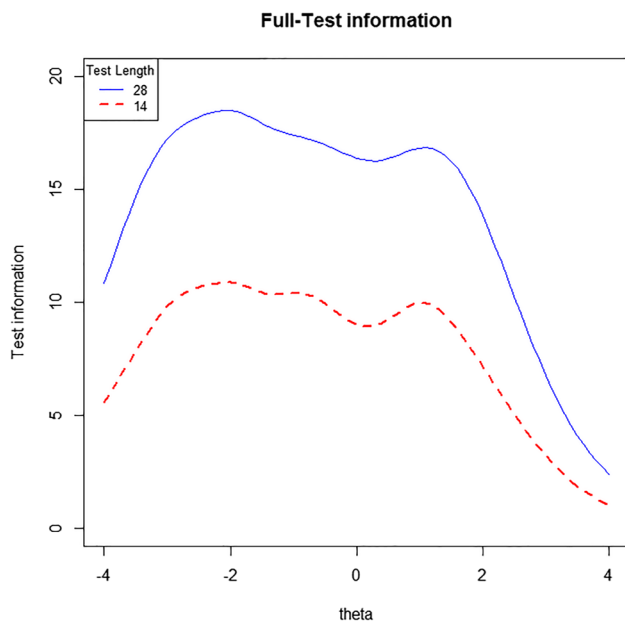


Fig. 5 Test information function curves of the original version and the short version of the REIS for the general factor ($N_2 = 497$)

form, respectively. The correlation values (see Table 8) were also very close, which verifies the performance in the short version.

Discussion

The objective of this research was to develop a short version of REIS and to evaluate its performance. First, we explore a new EI structure, the bifactor model, which has the better fitting indices with CFA. The bifactor model consists of a general factor and four conceptually different EI dimensions (such as the original scale), which include self-focused emotion appraisal, other-focused emotion appraisal, self-focused emotion regulation, and other-focused emotion regulation, which indicates that the short version covers all four dimensions of the original version. Next, the IRT method was used to shorten the REIS based on two factors: one is that the IRT method plays an important role in the development of modern psychometric tests and scales (Mungas & Reed, 2000), and the other is that it is especially suited for scale construction and refinement, containing the reduction of scales (Petersen, et al., 2006). By comprehensively considering $S-X^2$, DIF, item discrimination and IIF, this study obtained a short scale with 14 items. Compared with the original scale, although there are fewer items, the short scale has a perfect fitting index,

which is important because reducing the scale should not be achieved at the cost of losing important content. In addition, the $S-X^2$ of the short scale was acceptable, and there was no item with DIF. TIF shows that the short scale is good at distinguishing individuals within the range of potential traits. More importantly, the convergence validity of the short REIS was established by its high correlation with other EI scales. The total score of the REIS scale was significantly moderately or highly positively correlated with the total score of the EI and WLEIS scales, respectively, indicating that the REIS had sufficient convergence validity. For the usefulness of the scale, considering the small number of items in each subdimension, we believe that, although the reliability of the subscale reaches 0.7, which meets the standard of psychometry (Gu & Wen, 2017), caution should be exercised when using the current subscale as a measurement tool by itself. In terms of coverage of test information, the short scale retains 66.96% of the original scale's information, which indicates that the short REIS could systematically capture self- and other-focused EI similar to the original scale and provide a wider scope of EI dimensions. In conclusion, compared with the original version of REIS, the developed short version saves 50% of the items and has competitive validity and reliability.

The current study is not without limitations. First, the REIS is a self-reported instrument that has the potential influence of a social desirability bias. That is, participants can easily disguise or deliberately distort their responses to make themselves present a favorable self-representation. The second limitation is that this study focuses on college students, but the scale has been widely applied to all kinds of people. Therefore, it is necessary to be cautious in interpreting the results of the study because the respondents were not recruited through probability sampling and were concentrated among college students, which also leads to the limited generalizability of our study results. Sampling bias may have reduced the attitudes of different groups toward REIS because of differences in emotional intelligence between groups. The third limitation was that participants were biased against female participants. We note that the gender distribution (58% women) is skewed; although most items do not have DIF on gender, the gender ratio is not balanced. The last limitation is the use of subscales. Due to the small number of topics in each subscale, caution should be taken when using the current subscale as a measurement tool by itself. We suggest that the short scale should be used as a whole.

Despite the limitations of this research, we believe that the development of the scale has a great number of potential meanings for future research. Although previous studies have proven that REIS are

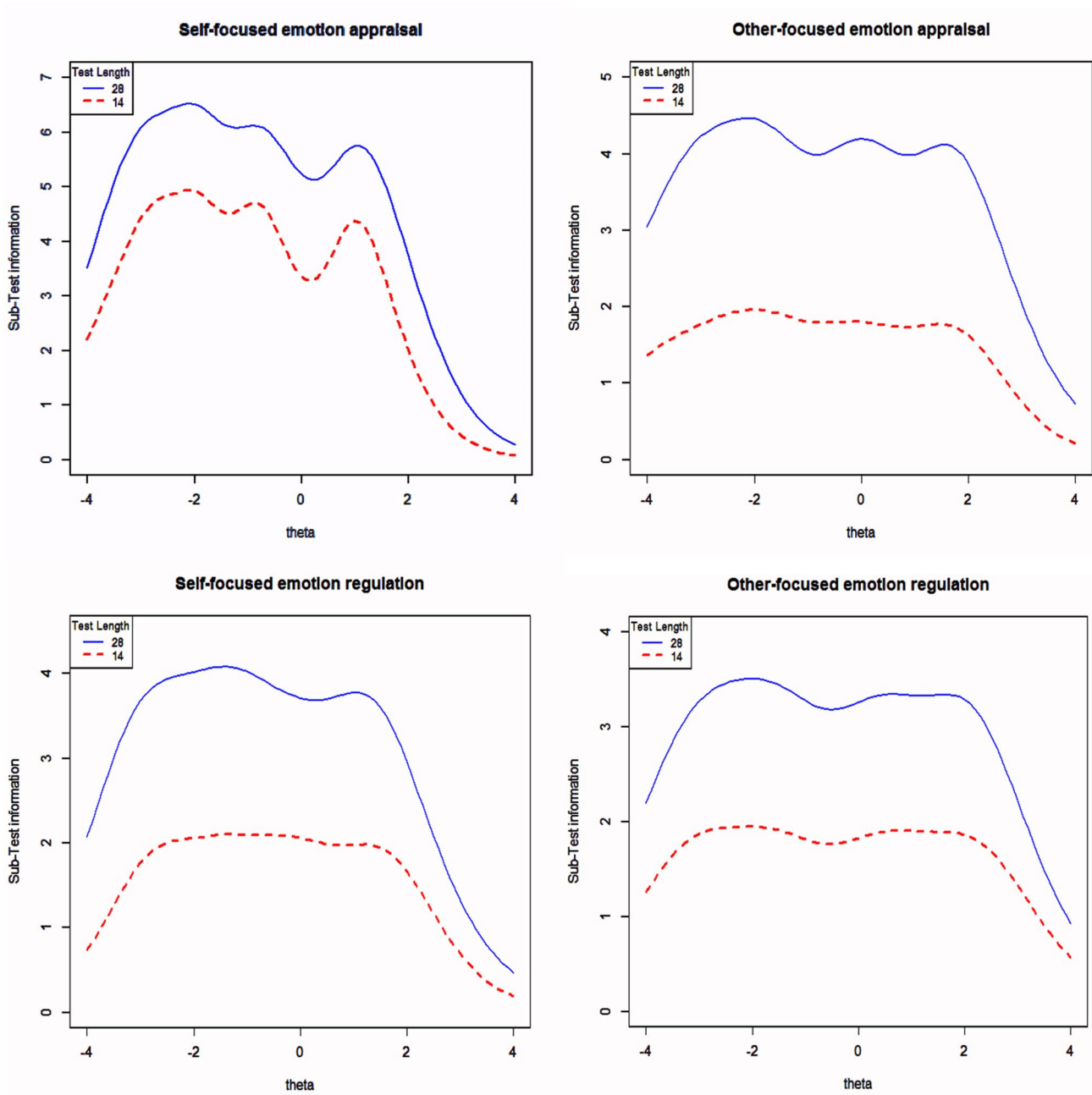


Fig. 6 Test information function curves of the original version and the short version of the REIS for the four special factors ($N_2 = 497$)

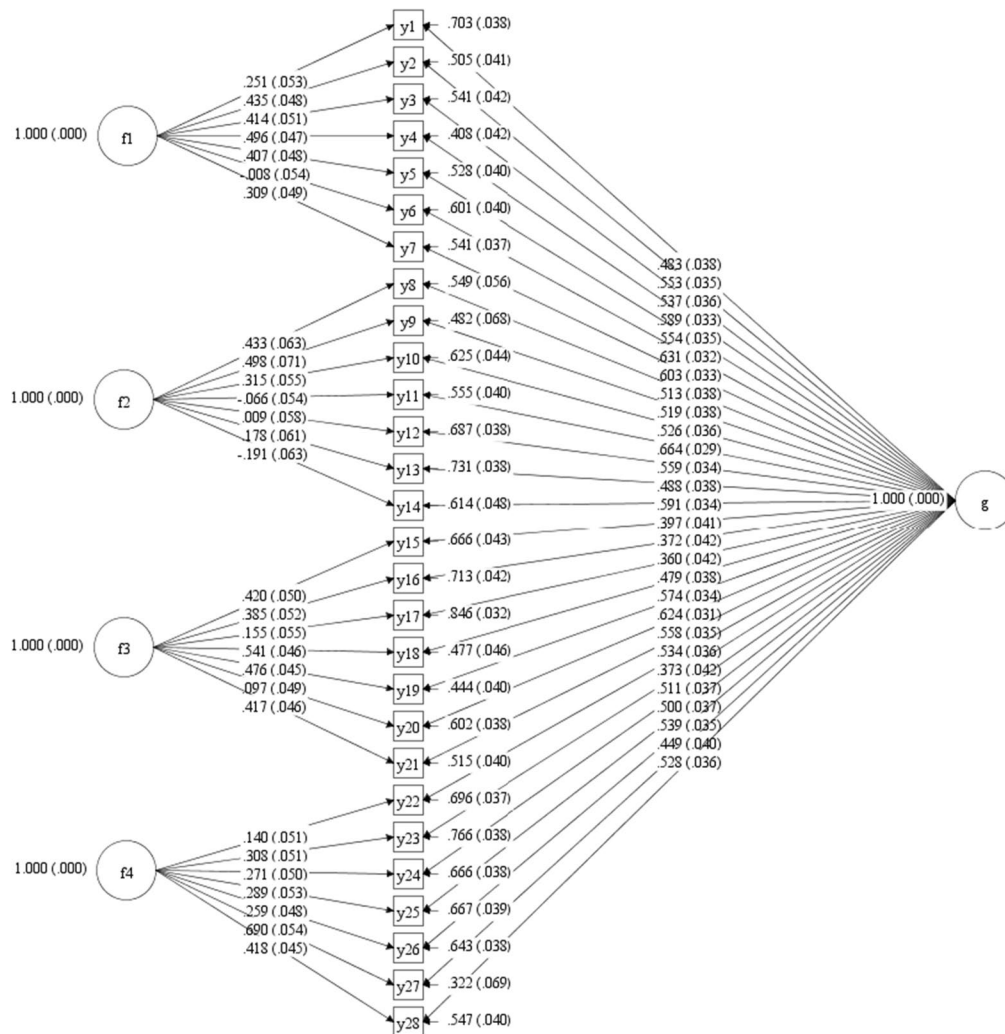
Table 8 Correlations between the REIS (Original and Short Form), EIS, and WLEIS

Version	EIS	WLEIS
REIS-long	0.75**	0.73**
REIS-short	0.70**	0.72**

* $p < 0.05$; ** $p < 0.01$

multidimensional, the current data distinctly indicate evidence in favor of a bifactor model. The research result of a bifactorial model supplies the basis for current practice.

Appendix



Path Coefficient of the 28 Items of the REIS ($N_1 = 494$)

Note: All the path coefficients in the figure are statistically significant at the level of 0.01. f1 = Self-focused emotion appraisal; f2 = Other-focused emotion appraisal; f3 = Self-focused emotion regulation; f4 = Other-focused emotion regulation.

Funding This work was supported by the National Natural Science Foundation of China (62167004, 31960186, 32160203).

Data availability The datasets generated and analyzed during the current study are available at <https://osf.io/2rus5/>.

Declarations

Ethical approval All procedures carried out in the studies involving human participants were in accordance with the ethical standards of the

Research Center of Psychological Health Education, Jiangxi Normal University and the 1964 Helsinki Declaration and its subsequent revisions or similar ethical standards.

Informed consent Informed consent was given to all individual respondents included in the research.

Conflict of interest The authors states that there is no conflict of interest.

References

- Austin, E. J., Saklofske, D. H., & Smith, M. M. (2018). Development and Validation of Two Short Forms of the Managing the Emotions of Others (MEOS) Scale. *Frontiers in Psychology*, 9, 974. <https://doi.org/10.3389/fpsyg.2018.00974>
- Blasco-Belled, A., Rogoza, R., Torrelles-Nadal, C., & Alsinet, C. (2019). Emotional intelligence structure and its relationship with

- life satisfaction and happiness: new findings from the bifactor model. *Journal of Happiness Studies*, 21(4), 2031–2049. <https://doi.org/10.1007/s10902-019-00167-x>
- Brasseur, S., Goire, J. G., Bourdu, R., & Mikolajczak, M. (2013). The profile of emotional competence (pec): Development and validation of a self-reported measure that fits dimensions of emotional competence theory. *PLoS ONE*, 8(5), e62635. <https://doi.org/10.1371/journal.pone.0062635>
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48, 1–29. <https://doi.org/10.18637/jss.v048.i06>
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39, 1–30. <https://doi.org/10.18637/jss.v039.i08>
- Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIF detect and difwithpar. *Medical Care*, 44(11 Suppl 3), S115–S123. <https://doi.org/10.1097/01.mlr.0000245183.28384.ed>
- Crane, P. K., Gibbons, L. E., Ocepek-Welikson, K., Cook, K., Cella, D., Narasimhalu, K., Hays, R. D., & Teresi, J. A. (2007). A Comparison of Three sets of Criteria for determining the presence of differential item functioning using ordinal logistic regression. *Quality of Life Research*, 16, 69–84. <https://doi.org/10.1007/s11136-007-9185-5>
- Dodeen, H., & Al-Darmaki, F. (2016). The application of item response theory in developing and validating a shortened version of the Emirate Marital Satisfaction Scale. *Psychological Assessment*, 28(12), 1625–1633. <https://doi.org/10.1037/pas0000296>
- Embreston, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Erlbaum Associate. <https://doi.org/10.4324/9781410605269>
- Embretson, S. E. (1996). The New rules of Measurement. *Psychological Assessment*, 8, 341–349. <https://doi.org/10.1037/1040-3590.8.4.341>
- Extremera, N., Ruiz-Aranda, D., Pineda-Galán, C., & Salguero, J. M. (2011). Emotional intelligence and its relation with hedonic and eudaimonic well-being: A prospective study. *Personality & Individual Differences*, 51(1), 11–16. <https://doi.org/10.1016/j.paid.2011.02.029>
- Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E., Grochocinski, V. J., & Stover, A. (2007). Full-information Item Bifactor Analysis of graded response data. *Applied Psychological Measurement*, 31, 4–19. <https://doi.org/10.1177/0146621606289485>
- Green, K. E., & Frantom, C. G. (2002). Survey Development and Validation With the Rasch Model. *International Conference on Questionnaire Development, Evaluation, and Testing*, 42.
- Gu, H. C., & Wen, Z. L. (2017). Reporting and Interpreting Multidimensional Test Scores: A Bi-factor Perspective. *Psychological Development and Education*, 33(4), 504–512. <https://doi.org/10.16187/j.cnki.issn1001-4918.2017.04.15>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. *Contemporary Sociology*, 21(2), 289–290. <https://doi.org/10.2307/2075521>
- Hodzic, S., Scharfen, J., Ripoll, P., et al. (2017). How efficient are emotional intelligence trainings: a meta-analysis. *Emotion review: journal of the International Society for Research on Emotions*, 10(2), 138–148. <https://doi.org/10.1177/1754073917708613>
- Hou, J. T., Wen, Z. L., & Cheng, Z. J. (2004). *Structural equation model and its applications*. Education Science Press.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>
- Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, 95, 54–78. <https://doi.org/10.1037/a0017286>
- Kim, M. (2001). Detecting DIF across the different language groups in a speaking test. *Language Testing*, 18(1), 89–114. <https://doi.org/10.1191/026553201675366418>
- Li, Y., Li, Y., Duan, W., Guan, Q., & Tao, Y. (2021). Testing the contribution of general factor of character strengths to well-being: an exploratory bifactor approach. *Current Psychology* (2). <https://doi.org/10.1007/s12144-020-01223-z>
- Lievens, F., & Chan, D. (2017). Practical intelligence, emotional intelligence, and social intelligence. In *Handbook of Employee Selection* (2nd ed., pp. 342–364). Taylor and Francis.
- Lopes, P. N., Brackett, M. A., Nezlek, J. B., Schutz, A., Sellin, I., & Salovey, P. (2004). Emotional Intelligence and Social Interaction. *Personality and Social Psychology Bulletin*, 30, 1018–1034. <https://doi.org/10.1177/0146167204264762>
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- Mao, X. Z., Xia, M. L., & Xin, T. (2018). Full-information item bifactor analysis: Model, parameter estimation and application. *Advances in Psychological Science*, 26(2), 10. <https://doi.org/10.3724/SPJ.1042.2018.00358>
- Martins, A., Ramalho, N., & Morin, E. (2010). A Comprehensive Meta-Analysis of the Relationship between emotional intelligence and health. *Personality and Individual Differences*, 49, 554–564. <https://doi.org/10.1016/j.paid.2010.05.029>
- Mungas, D., & Reed, B. R. (2000). Application of item response theory for development of a global functioning measure of dementia with linear measurement properties. *Statistics in Medicine*, 19(11–12), 1631–1644. [https://doi.org/10.1002/\(SICI\)1097-0258\(20000615/30\)19:11/12%3c1631::AID-SIM451%3e3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0258(20000615/30)19:11/12%3c1631::AID-SIM451%3e3.0.CO;2-P)
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50–64. <https://doi.org/10.1177/01466216000241003>
- Pekaar, K. A., Bakker, A. B., Dimitri, V. D. L., & Born, M. P. (2018). Self- and other-focused emotional intelligence: Development and validation of the Rotterdam Emotional Intelligence Scale (REIS). *Personality and Individual Differences*, 120, 222–233. <https://doi.org/10.1016/j.paid.2017.08.045>
- Penfield, R. D., & Camilli, G. (2006). Differential Item Functioning and Item Bias. *Handbook of Statistics*, 26, 125–167. [https://doi.org/10.1016/S0169-7161\(06\)26005-X](https://doi.org/10.1016/S0169-7161(06)26005-X)
- Petersen, M. A., Groenvold, M., Aaronson, N., Blazebly, J., Brandberg, Y., Graeff, A. D., Fayers, P., Hammerlid, E., Sprangers, M., & Velikova, G. (2006). Item response theory was used to shorten EORTC QLQ-C30 scales for use in palliative care. *Journal of Clinical Epidemiology*, 59(1), 36–44. <https://doi.org/10.1016/j.jclinepi.2005.04.010>
- Prati, G., & Pietrantonio, L. (2016). Bifactor and Item Response Theory Analyses of Health care workers' attitudes toward HIV-infected people. *European Journal of Psychological Assessment*, 1–8. <https://doi.org/10.1027/1015-5759/a000379>
- R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Reeve, B. B., & Fayers, P. M. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. M. Fayers & R. Hays (Eds.), *Assessing Quality of Life in Clinical Trials: Methods of Practice*, (2nd ed., pp. 55–73). Oxford University Press.

- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92, 544–559. <https://doi.org/10.1080/00223891.2010.496477>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment*, 98, 223–237. <https://doi.org/10.1080/00223891.2015.1089249>
- Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination Cognition and Personality*, 9(6), 217–236. <https://doi.org/10.2190/DUGG-P24E-52WK-6CDG>
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 34(1), 1–97. <https://doi.org/10.1007/BF03372160>
- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality & Individual Differences*, 25(2), 167–177. [https://doi.org/10.1016/S0191-8869\(98\)00001-4](https://doi.org/10.1016/S0191-8869(98)00001-4)
- Tejada-Gallardo, C., Blasco-Belled, A., Torrelles-Nadal, C., & Alsinet, C. (2020). How does emotional intelligence predict happiness, optimism, and pessimism in adolescence? investigating the relationship from the bifactor model. *Current Psychology* (4). <https://doi.org/10.1007/s12144-020-01061-z>
- Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief to medium-length Big Five and Big Six Personality Questionnaires. *Psychological Assessment*, 23, 995–1009. <https://doi.org/10.1037/a0024165>
- Tsutakawa, R. K., & Johnson, J. C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*, 55, 371–390. <https://doi.org/10.1007/BF02295293>
- Wang, C. K. (2002). Emotional intelligence, general self-efficacy and coping style of delinquent teenagers. *Chinese Mental Health Journal*, 16(08), 566–567. <https://doi.org/10.3969/j.issn.1005-3611.2002.04.021>
- Wang, M. C. (2014). *Latent variable modeling using Mplus*. Chongqing University Press.
- Wang, Y. F. (2010). *Reliability and Validity of Chinese Version of Emotional Intelligence Scale*. (Doctoral dissertation, Central South University).
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational & Psychological Measurement*, 76(6), 913–934. <https://doi.org/10.1177/0013164413495237>
- Wong, C. S., & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *The Leadership Quarterly*, 13, 243–274. [https://doi.org/10.1016/S1048-9843\(02\)00099-1](https://doi.org/10.1016/S1048-9843(02)00099-1)
- Zeidner, M., Roberts, R. D., Matthews, G. (2008). The science of emotional intelligence. *European Psychologist*, 13, 64–78. <https://doi.org/10.1027/1016-9040.13.1.64>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.