

Validation of the Empathy Quotient in Mainland China

Qing Zhao, David L. Neumann , Xiaoyan Cao, Simon Baron-Cohen, Xiang Sun, Yuan Cao, Chao Yan, Yuna Wang, Lin Shao & David H. K. Shum

To cite this article: Qing Zhao, David L. Neumann , Xiaoyan Cao, Simon Baron-Cohen, Xiang Sun, Yuan Cao, Chao Yan, Yuna Wang, Lin Shao & David H. K. Shum (2017): Validation of the Empathy Quotient in Mainland China, Journal of Personality Assessment, DOI: [10.1080/00223891.2017.1324458](https://doi.org/10.1080/00223891.2017.1324458)

To link to this article: <http://dx.doi.org/10.1080/00223891.2017.1324458>



Published online: 24 May 2017.



Submit your article to this journal [↗](#)



Article views: 31



View related articles [↗](#)



View Crossmark data [↗](#)



Validation of the Empathy Quotient in Mainland China

Qing Zhao,¹ David L. Neumann,² Xiaoyan Cao,³ Simon Baron-Cohen,⁴ Xiang Sun,^{4,5,6} Yuan Cao,⁷ Chao Yan,⁸ Yuna Wang,⁹ Lin Shao,¹⁰ and David H. K. Shum²

¹School of Applied Psychology and Menzies Health Institute Queensland, Griffith University, Brisbane, Australia; ²School of Applied Psychology and Menzies Health Institute Queensland, Griffith University, Gold Coast Campus, Southport, Australia; ³Department of Psychology, Southwest Medical University, Luzhou, Sichuan, China; ⁴Department of Psychiatry, University of Cambridge, United Kingdom; ⁵MIND Institute, University of California, Davis; ⁶Psychology Department, The Chinese University of Hong Kong, Hong Kong, China; ⁷School of Psychology, University of Queensland, Brisbane, Australia; ⁸Key Laboratory of Brain Functional Genomics, Ministry of Education, Shanghai Key Laboratory of Brain Functional Genomics (MOE & STCSM), East China Normal University, Shanghai, China; ⁹Flight Technology College, Civil Aviation University of China, Tianjin, China; ¹⁰CapitalBio Corporation, Beijing, China

ABSTRACT

This research aimed to validate a simplified Chinese version of the Empathy Quotient (EQ; 60 items) for use with Mainland Chinese people. The original English version of the EQ was translated into simplified Chinese. Through an online survey, 588 Mainland Chinese participants completed the EQ and 3 other questionnaires: the Interpersonal Reactivity Index (IRI), the Autism-Spectrum Quotient (AQ), and the 20-item Toronto Alexithymia Scale (TAS-20). Thirty-five participants completed retesting of the EQ 3 to 4 weeks later. Sex differences on the EQ scores and psychometric properties of the EQ items were examined. Confirmatory factor analysis suggested that an EQ 15-item structural model fitted the data quite well. Self-report empathy, as assessed by the current simplified Chinese version of the EQ, appeared to relate to participants' autistic and alexithymic traits but not sex.

ARTICLE HISTORY

Received 4 February 2016
Revised 6 March 2017

The Empathy Quotient (EQ) is a self-report instrument developed by Baron-Cohen and Wheelwright (2004) to measure empathy in both healthy individuals and those with autism spectrum conditions (ASCs). To date, the original English version of the EQ has been translated into a number of languages and validated in different cultures and populations (Dimitrijević, Hanak, Vukosavljević-Gvozden, & Opačić, 2012; Groen, Fuermaier, Den Heijer, Tucha, & Althaus, 2015; Kim & Lee, 2010; Lepage, Lortie, Taschereau-Dumouchel, & Théoret, 2009). Scores on the EQ have been found to have good reliability and validity across these linguistic and cultural differences in measuring self-report empathy (Groen et al., 2015). However, a full-length (60-item) version of the EQ has not been adapted or fully validated for use in Mainland China.

Empathy is “understanding and sharing in another’s emotional state or context” (Cohen & Strayer, 1996, p. 988). It is an essential social communication skill (Baron-Cohen & Wheelwright, 2004) and consists of two main components (Cohen & Strayer, 1996; Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004; Shamay-Tsoory, 2011). One is emotional or affective empathy (Decety & Moriguchi, 2007), which is sharing and responding to another person’s emotional state with an appropriate emotion (Baron-Cohen & Wheelwright, 2004; Smith, 2006). The other is cognitive empathy, or perspective taking, which is the ability to recognize another person’s mental state, including his or her emotions, thoughts, intentions, and perceptions (Baron-Cohen & Wheelwright, 2004; Shamay-

Tsoory, 2011; Smith, 2006). There is some evidence that these two components have dissociated brain networks (Cox et al., 2012; Fan, Duncan, de Greck, & Northoff, 2011; Shamay-Tsoory, 2011; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009).

The original EQ (60 items) has 40 items that measure empathy as a single construct and another 20 filler items (Baron-Cohen & Wheelwright, 2004). Each item is rated on a 4-point Likert scale. The 20 filler items, which are not scored, were designed by the authors of the EQ to prevent participants from constantly answering empathy questions (Baron-Cohen & Wheelwright, 2004). Examples of the 40 empathy items include: “I can tell if someone is masking their true emotion” (EQ 55), “Seeing people cry doesn’t really upset me” (EQ 32, reverse item), and “I find it hard to know what to do in a social situation” (EQ 8, reverse item; Baron-Cohen & Wheelwright, 2004). According to the instructions given by Baron-Cohen and Wheelwright (2004), each empathy item is scored on a reduced basis (viz., from 0, 1, to 2 points). For a forward item (e.g., EQ 55), 2 points are provided for strong agreement, 1 point for mild agreement, and 0 points for both mild and strong disagreement; for a reverse item (e.g., EQ 32), the preceding four conditions are scored with 0, 0, 1, and 2 points, respectively. The total score of the EQ ranges from 0 to 80, with higher scores reflecting greater empathy. Subsequent research has suggested different structural models for the EQ. Some have identified emotional and cognitive empathy as two factors (Andrew, Cooke, & Muncer, 2008; Lawrence et al.,

2004; Muncer & Ling, 2006), whereas others have found empathy as measured by the EQ items to be unidimensional (Allison, Baron-Cohen, Wheelwright, Stone, & Muncer, 2011; Guan, Jin, & Qian, 2012; Wakabayashi et al., 2006).

Lawrence et al. (2004) validated the EQ in a group of British participants ($N = 172$, M age = 34.1 years, $SD = 10.4$). Factor analyses indicated that EQ items loaded on three factors, namely, cognitive empathy (11 items, e.g., EQ 55), emotional reactivity (11 items, e.g., EQ 32), and social skills (6 items, e.g., EQ 8). The first two factors were used to measure cognitive and emotional empathy separately (Lawrence et al., 2004). The three-factor model has been found to have a good fit to the observed data in several EQ validation studies based on different populations and translated versions, such as Dutch (Groen et al., 2015) and French (Berthoz, Wessa, Kedia, Wicker, & Grèzes, 2008).

Allison et al. (2011) tested the EQ in a group of ethnically diverse participants ($N = 5,377$; age range = 16–78 years). Results of factor analyses indicated that EQ items loaded on two factors. The two factors were grouped according to the items' response directions, namely, items requiring agreement (13 items; e.g., EQ 55) and items requiring disagreement (13 items; e.g., EQ 32) to indicate empathy (Allison et al., 2011). Wakabayashi et al. (2006) found with a group of British university students ($N = 1,761$; age range = 18–26 years) that the EQ items did not separate in terms of the theoretical components of empathy nor in terms of the items' response directions, and thus recommended a one-factor model.

To date, the best fit structural model of the EQ in the Mainland Chinese context is unclear. There have been two attempts to validate the EQ in Mainland China, with inconsistent results: Yang, Xiao, Qian, Mo, and Zhuo (2013) aimed to validate a full-length EQ (60 items) with a group of Mainland Chinese participants ($N = 426$; M age = 35.1 years, $SD = 4.6$). However, Yang et al. did not follow a standard translation and validation process in validating the EQ for use in Mainland China. First, they did not provide any information about their translation process, which must be reported according to a standard cross-cultural validation process (Beaton, Bombardier, Guillemin, & Ferraz, 2000). Second, they did not conduct a confirmatory factor analysis (CFA), which is a procedure required to check the fit of previously established structural models to a new data set (Burnett & Dart, 1997; Levine, 2005). In all, Yang et al. considered their study an early exploratory investigation and suggested that the formal validation of the EQ in a Chinese context requires further study.

Guan et al. (2012) validated a short version of the EQ (22 items) in Mainland China using a group of health care trainees or professionals ($N = 840$; age range = 17–52 years). They confirmed that the best fit model for their translated version was a one-factor model with 15 EQ items (Guan et al., 2012). However, it should be noted that health-care training and work environments can change individuals' self-reported empathy levels (Dehning et al., 2013; Nunes, Williams, Sa, & Stevenson, 2011; Penprase, Oakley, Ternes, & Driscoll, 2013). Furthermore, the one-factor model provided by Guan et al. only summarized the psychometric properties of the items of the short version of the EQ. Therefore, further investigation is needed to test the psychometric properties of the full-length version of the EQ items in a Mainland Chinese sample.

In addition to the factor structure of the EQ, sex differences in empathy are relevant to validation of the EQ (Baron-Cohen & Wheelwright, 2004). Females on average score higher than males on self-reported empathy (Baron-Cohen & Wheelwright, 2004; Groen et al., 2015). This might reflect prenatal biological influences on the sexes, including genetics (Wu, Li, & Su, 2012), hormone modulation (Hurlemann et al., 2010), and neural differences (Derntl et al., 2010; Rueckert & Naybar, 2008; Schulte-Rüther, Markowitsch, Shah, Fink, & Piefke, 2008). It might also be related to postnatal experiences, such as cultural influences (Dehning et al., 2013) and social expectations (Ickes, Gesn, & Graham, 2000). However, previous self-report studies of empathy based on Chinese populations have not found the expected sex difference on the EQ scores (Guan et al., 2012; Yang et al., 2013). These studies can be questioned in that Guan et al. (2012) used a short version of the EQ with health care professionals, and Yang et al. (2013) did not validate the EQ using a standard translation and validation process. These limitations might conceal a genuine sex difference in empathy for Mainland Chinese participants, and point to the need to validate the full-length EQ (60 items) in Mainland China following a recommended cross-cultural validation procedure (Beaton et al., 2000).

This study aimed to validate a full-length (60 items) simplified Chinese version of the EQ in Mainland China. There are two forms of Chinese written text, traditional and simplified. The former has a longer history and is used in Hong Kong, Macau, and Taiwan; the latter is used in Mainland China and Singapore. The factor structure, internal consistency, and test-retest coefficients of the EQ scores were assessed, and sex differences were examined. Finally, three other self-report scales were selected with reference to previous EQ validation studies to provide evidence to support the construct validity of the simplified Chinese version of the EQ scores in measuring self-report empathy of the Mainland Chinese participants (Baron-Cohen & Wheelwright, 2004; Groen et al., 2015; Lawrence et al., 2004; Preti et al., 2011). The three scales included an empathy scale, namely, the Interpersonal Reactivity Index (IRI; Chan, 1986; Davis, 1980; Melchers, Montag, Markett, & Reuter, 2015; Siu & Shek, 2005); an autism scale, namely, the Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001); and an alexithymia scale, namely, the 20-item Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994; Zhu et al., 2007). We predicted significant positive correlations between scores on the EQ and the IRI perspective taking (IRI-PT) and empathic concern (IRI-EC) subscales (Dimitrijević et al., 2012; Kim & Lee, 2010; Lawrence et al., 2004; Preti et al., 2011), and significant negative correlations between scores on the EQ and the AQ and the TAS-20 subscales (Baron-Cohen & Wheelwright, 2004; Bird et al., 2010; Jonason & Krause, 2013; Lombardo, Barnes, Wheelwright, & Baron-Cohen, 2007; Melchers et al., 2015; Preti et al., 2011; Wheelwright et al., 2006; Williams & Wood, 2010).

Method

Participants

Participants voluntarily completed an anonymous online survey on the Sojump platform (see <http://www.sojump.com>) in

Mainland China. Advertisements were broadcast through two Mainland Chinese popular online chat tools (Wechat and QQ Software), three Mainland Chinese popular public information Web sites (taobao.com, weibo.com, and qq.com), and several online forums. The first author's institution granted ethical approval. All participants provided their informed consent online prior to taking part in the study. A 25 RMB cash (about US\$4) or equivalent gift was provided to each participant to compensate for his or her time.

This study included a test and a retest phase. During the first administration, participants were informed that the study was restricted to individuals who satisfied the following inclusion criteria: (a) nationality is Chinese; (b) ethnicity is Han Chinese or minority Chinese; (c) place of birth was Mainland China; (d) primary residence while growing up was Mainland China; (e) current primary residence is Mainland China; and (f) was 18 years or older. At the end of the first administration, each participant was asked to provide a six-digit password. Retest participants were randomly selected from the individuals who finished the first administration. During the retest administration, participants were asked to meet two additional criteria: a valid password consistent with that created at their first administration, and a time interval between the test and the retest phase of 3 to 4 weeks.

In all, 634 participants ($N = 588$ final participants; see later) completed the first administration. Of the original sample, 40 participants were randomly selected and invited to take part in the retest administration, and 38 of them ($N = 35$ final participants; see later) completed the retest administration. Demographic information for the final participants of the test and retest phases is provided in Table 1. The primary residences reported by the 588 final participants were 30 out of the 31 provinces of Mainland China. The only province not represented was the Tibet Autonomous Region. Among the 588 final participants, 560 (95.2%) were Han Chinese and 28 (4.8%) were minority Chinese. The 28 minority Chinese participants came from 12 different ethnic minorities in Mainland China: 6 Zhuang Chinese, 5 Manchu Chinese, 4 Uygur Chinese, 3 Hui Chinese, 2 Bai Chinese, 2 Gelo Chinese, 1 Bouyei Chinese, 1 Hani Chinese, 1 Li Chinese, 1 Mongolian Chinese, 1 Tibetan Chinese, and 1 Yi Chinese.

Excluded and missing data

For the first administration, 46 participants were excluded for the following five reasons: cultural background was not Mainland Chinese (2 participants); age was younger than 18 years (8 participants); if multiple questionnaires were submitted by any

one participant using the same Internet Protocol address and with the same demographic information, only one of the questionnaires (the one with the longest finishing time) was included and the others were excluded (5 participants); if individuals selected the same answer for all items on a questionnaire, they were excluded (6 participants); and individuals who finished their first administration in less than 15 min were excluded (25 participants). For the retest administration, 3 participants were excluded because they did not provide valid passwords (their passwords and demographic information in the retest administration did not match those submitted in the first administration).

The online survey was designed in such a way that participants could not submit their results if any of the questions had not been answered. Therefore, there were no missing data. However, one participant answered "dropout" for the education level question and did not give any further explanation. Therefore, the education level of this participant was coded as missing. All analyses involving education level were based on the other 587 final participants of the first administration or the 35 final retest responses. All of the other analyses were based on the 588 final participants of the first administration or the 35 final retest responses.

Measures

Empathy Quotient

The EQ consists of 60 items, including 40 that measure empathy and 20 filler items (Baron-Cohen & Wheelwright, 2004). Each item is rated on a 4-point Likert scale ranging from 1 (*strongly agree*) to 4 (*strongly disagree*). The 40 empathy items were scored according to the original instructions, namely, for a forward item (e.g., EQ 55), 2 points are provided for a response of *strongly agree*, 1 point for *slightly agree*, and 0 points for both *slightly disagree* and *strongly disagree*; for a reverse item (e.g., EQ 32), these four records are scored with 0, 0, 1, and 2 points, respectively (Baron-Cohen & Wheelwright, 2004). The 20 filler items were designed by the authors of the EQ to prevent participants from constantly answering empathy questions and these were not scored (Baron-Cohen & Wheelwright, 2004). The total EQ score ranges from 0 to 80, with higher scores reflecting greater empathy. The Cronbach's α for scores on the 40 empathy items of the original version of the EQ was .92 (Baron-Cohen & Wheelwright, 2004). According to Groen et al. (2015), the Cronbach's α reported for other validation studies using other populations ranged from .78 to .89. The Cronbach's α for the scores of the 40 empathy items of the simplified Chinese version of the EQ based on the final sample of this study was .86.

Permission for translation for this cross-cultural adaptation of the EQ into simplified Chinese is based on the Autism Research Centre terms and conditions (www.autismresearchcentre.com). The overall validation processes followed the guidelines for cross-cultural adaptation of self-report measures (Beaton et al., 2000). The original English version of the EQ was translated into simplified Chinese by an English-Chinese bilingual researcher. The translated version was proofread by another two English-Chinese bilingual researchers. A third independent English-Chinese bilingual researcher back-

Table 1. Demographic information for test and retest participants.

Characteristic	First administration ($N = 588$)	Retest ($n = 35$)
Sex (n for males, male %)	213 (36.2%)	12 (34.3%)
Student or employee (n for student, student %)	420 (71.4%)	15 (42.9%)
Mean age (years, SD)	24.12 (6.20)	30.14 (8.70)
Mean education (years, SD)	15.43 (2.22) ^a	16.91 (1.88)

^a $n = 587$. One participant's education level was indicated as "dropout" without a further explanation. This participant's education level was treated as missing.

translated the simplified Chinese statements into English. The simplified Chinese version of the EQ and its English back-translation were provided to the EQ's original author for checking. The final translation was approved by all members of the translation panel and was tested in a pilot study with 10 Mainland Chinese participants. All reported that the simplified Chinese version was clear and readable.

Interpersonal Reactivity Index

The IRI includes 28 items and measures an individual's empathy based on four subscales (7 items each), namely, perspective taking (IRI-PT), empathic concern (IRI-EC), personal distress (IRI-PD), and fantasy (IRI-FS; Davis, 1980). Each item is rated on a 5-point Likert scale ranging from 0 (*does not describe me well*) to 4 (*describes me very well*). The IRI items were scored and the values for the four subscales were computed according to the original instruction (Davis, 1980). The total score on each subscale ranges from 0 to 28 with higher scores reflecting greater empathy. The IRI has been translated into traditional (Chan, 1986; Siu & Shek, 2005) and simplified Chinese (Huang, Li, Sun, Chen, & Davis, 2012). In Mainland China, researchers have frequently adapted the Chan (1986) traditional Chinese version into simplified Chinese to measure empathy, as it was the first available translation of the IRI (Neumann, Chan, Wang, & Boyle, 2016; Y. Wang et al., 2013; Yang et al., 2013; F. Zhang, Dong, Wang, Zhan, & Xie, 2010; Q. Zhang et al., 2014). Permission was obtained from the authors of the Y. Wang et al. (2013) study to use their adapted simplified Chinese version of the IRI in this study. The Cronbach's α values for the scores on the four IRI subscales (*viz.* IRI-PT, IRI-EC, IRI-PD, and IRI-FS) for our final participants ($N = 588$) were .66, .69, .79, and .72, respectively. These values are consistent with the original English and previous Chinese versions of the IRI (Cronbach's α ranged from .68–.79; Davis, 1980; Y. Wang et al., 2013).

Autism-Spectrum Quotient

The AQ consists of 50 items and assesses an individual's autistic traits (Baron-Cohen et al., 2001). It includes five subscales (10 items each), namely, poor social skills (AQ-SS), poor communication skills (AQ-CM), exceptional attention to detail (AQ-AD), poor imagination (AQ-IM), and poor attention switching (AQ-AS; Baron-Cohen et al., 2001). Each item is rated on a 4-point Likert scale ranging from 1 (*definitely agree*) to 4 (*definitely disagree*). According to the original instructions, each item "scores 1 point if the respondent records the abnormal or autistic-like behavior either mildly or strongly" (Baron-Cohen et al., 2001, p. 6). The total score on each subscale of AQ ranges from 0 to 10, with higher scores suggesting stronger autistic traits.

The cross-cultural translation of the AQ into simplified Chinese was based on the Autism Research Centre terms and conditions. The overall translation process of the AQ was identical to that of the EQ. The final translation was approved by all members of the translation panel. The final version of the AQ was tested with seven Mainland Chinese participants, and all of them reported that the translation was clear and readable. The Cronbach's α values for the five subscales as reported by Baron-Cohen et al. (2001) were .77, .65, .63, .65, and .67, respectively. The Cronbach's α values calculated based on our

final participants ($N = 588$) were .69, .57, .63, .33, and .32, respectively. Given the low values of alpha for AQ-IM and AQ-AS in this study and given that these two scales do not relate theoretically to empathy, they were not used in subsequent analyses.

The 20-Item Toronto Alexithymia Scale

The TAS-20 has 20 items and assesses an individual's alexithymic traits. The TAS-20 has three subscales, namely, difficulties in identifying one's own feelings (TAS-IF, 7 items), difficulties in describing one's feelings to other people (TAS-DF, 5 items), and externally oriented thinking (TAS-EOT, 8 items; Bagby et al., 1994). Each item is rated on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The TAS-20 was scored according to the original instructions (Zhu et al., 2007). The total scores for the three subscales of the TAS-20 range from 7 to 35, 5 to 25, and 8 to 40, respectively, with higher scores indicating stronger alexithymic traits (Zhu et al., 2007).

The TAS-20 has been translated into simplified Chinese for university students to self-report their alexithymic traits (Zhu et al., 2007). Cronbach's α values for scores on the three TAS-20 subscales reported by Zhu et al. (2007) were .77, .65, and .52, respectively. These values were lower than in the original English version (Cronbach's $\alpha = .80, .76, \text{ and } .71$, respectively; Parker, Taylor, & Bagby, 2003). Two English-Chinese bilingual researchers in this study proofread the Zhu et al. (2007) translation and agreed that seven translated items (1, 3, 8, 10, 12, 16, and 20) needed to be modified. The seven items were retranslated by the panel for this study using the same process as described for the EQ.

The modified translation of the TAS-20 and its back-translation were provided to both the original English author (Graeme J. Taylor) and the corresponding author of Zhu et al. (2007). Both provided their permission to use the modified version of the TAS-20 in this study. The final translation was approved by all members of the translation panel. The final translation of the TAS-20 was tested with seven Mainland Chinese participants, and all of them reported that the translation was clear and readable. The Cronbach's α values for the scores on the three subscales of the TAS-20 based on the final participants ($N = 588$) were .83, .70, and .55, respectively. Considering the TAS-EOT is not theoretically correlated with the EQ and has a very low Cronbach's α , it was not used in subsequent analyses.

Demographic information questionnaire

A demographic questionnaire was designed to collect the following information: personal demographic characteristics (date of birth, sex, education level), cultural background (nationality, place of birth and childhood, and primary residence), and occupation (employee or student). Type of work and academic major were also asked of the employees and students, respectively.

Procedure

All participants were instructed to read the introduction to the study and the inclusion criteria at the beginning of testing. It was explained that this study expected them to satisfy all the inclusion criteria, provide their demographic information

honestly, and complete the whole task carefully. Next, we asked participants to provide their informed consent. Following that, participants began to answer the questionnaires.

During the first administration, data for nine questionnaires were collected. The first five questionnaires were the demographic questionnaire, the EQ, the IRI, the AQ, and the TAS-20. The other four questionnaires were included for use as part of another study. At the end of the first administration, each participant was required to leave a six-digit password. During the retest administration, participants were only asked to enter their date of birth, sex, and their six-digit password on a short demographic questionnaire. With the exception of the demographic questionnaire, the other eight questionnaires included in the retest administration were identical to those in the first administration.

Data analysis

The CFA was conducted using *Mplus* 7.4 (Muthén & Muthén, 1998–2012). The weighted least squares with mean and variance adjustment (WLSMV) estimation method was used. The WLSMV is a robust estimator appropriate for ordered categorical data (Sass, 2011). The best fit parameters of CFA were set as the comparative fit index (CFI) ≥ .95, Tucker–Lewis Index (TLI) ≥ .90, root mean square error of approximation (RMSEA) ≤ .08, and weighted root mean square residual (WRMR) ≤ 1.00 (J. Wang & Wang, 2012).

Internal consistency of the EQ scores was calculated using Cronbach’s α. The stability of the EQ scores was examined using the two-way, random-effects, single-measure intraclass correlation coefficient (ICC type 2, 1; Shrout & Fleiss, 1979) between participants’ test and retest responses. The ICC is a ratio reflecting the proportion of total variance that is due to the variance between participants, and it is more sensitive to systematic error than Pearson’s correlation coefficient (Bédard, Martin, Krueger, & Brazil, 2000; Weir, 2005). The ICC (type 2, 1) with 95% confidence interval (CI) is frequently used for reporting the test–retest correlations of scores on self-report questionnaires (Fritz & Irrgang, 2001; Gremigni, Damásio, & Borsa, 2013; Hart, 2003). A single-measure ICC of equal to or greater than .75 is considered excellent reliability (Fleiss, 1999).

Bivariate linear correlation coefficients were calculated between scores on the EQ and the other three scales (IRI, AQ,

and TAS-20) to provide evidence of construct validity. The Pearson correlation coefficient (*r*) was chosen after linear relationships were confirmed by scatter plot inspection. Independent-sample *t* tests were used to test for sex differences in empathy (EQ and IRI subscales scores) and other scales. Cohen’s *d* was calculated as an effect size. To control for the possible impact of age and education level on empathy, sex differences were also checked using these two variables as covariates in the univariate analyses. Apart from the CFAs, all statistical analyses were conducted using SPSS.

Results

Demographic information

The demographic information for the final participants of the test and the retest phases are shown in Table 1. For the first administration, 28.6% of participants were employed (*n* = 168) in 12 different types of work (e.g., office workers, public servants). The other 71.4% of participants were students (*n* = 420) studying 23 different academic majors (e.g., aerospace, medicine).

Confirmatory factor analysis

Six structural models have been reported for the EQ. The model description and CFA results for each model are provided in Table 2. Results showed that Guan et al.’s (2012) 15-item one-factor model is a reasonable fit to these data. The model modification indexes suggested that the Guan et al. model might provide a better fit to the data if a covariance were added between EQ 43 (i.e., “Friends usually talk to me about their problems as they say that I am very understanding”) and 36 (i.e., “Other people tell me I am good at understanding how they are feeling and what they are thinking”). With the pair of residuals correlated, the final modified model showed a good fit to the data (see Table 2). The CFA standardized estimates of the 15-item model are illustrated in Figure 1. The standardized factor loading regression weights (*b*) of the 15 items ranged from .33 to .82. The squared multiple correlations (*R*²) of these items ranged from .11 to .68. For the subsequent analyses, the total scores on the 40 items of the EQ (EQ-40) and on the 15 items of the Guan et al. scale (EQ-15) were calculated. The

Table 2. Descriptions of Empathy Quotient (EQ) structural models and confirmatory factor analysis (CFA) results based on a Mainland Chinese sample.

Reference	Model	Factors (item <i>n</i> for each factor)			CFA results						Cronbach’s α for each factor		
		1	2	3	χ ²	<i>df</i>	CFI	TLI	RMSEA [90% CI]	WRMR	1	2	3
Baron-Cohen and Wheelwright (2004)	EQ-40-item one-factor	EM (40)	/	/	3243.12	740	.73	.71	.076 [.073, .079]	2.18	.86	/	/
Lawrence et al. (2004)	EQ-28-item three-factor	CE (11)	ER (11)	SS (6)	1588.28	347	.84	.83	.078 [.074, .082]	1.93	.87	.69	.57
Wakabayashi et al. (2006)	EQ-22-item one-factor	EM (22)	/	/	1200.55	209	.86	.84	.090 [.085, .095]	1.93	.86	/	/
Muncer and Ling (2006)	EQ-15-item three-factor	CE (5)	ER (5)	SS (5)	360.04	87	.90	.88	.073 [.065, .081]	1.54	.78	.55	.56
Allison et al. (2011)	EQ-26-item two-factor	AG (13)	DI (13)	/	732.93	298	.91	.90	.050 [.045, .054]	1.40	.80	.74	/
Guan, Jin, and Qian (2012)	EQ-15-item one-factor	EM (15)	/	/	422.11	90	.94	.93	.079 [.072, .087]	1.44	.86	/	/
This study	Modified Guan et al. model ^a	EM (15)	/	/	359.16	89	.95	.95	.072 [.064, .080]	1.31	.86	/	/

Note. *N* = 588. CFI = comparative fit index; TLI = Tucker–Lewis Index; RMSEA = root mean square error of approximation; WRMR = weighted root mean square residual; EM = empathy; CE = cognitive empathy; ER = emotional reactivity; SS = social skills; AG = agreement; DI = disagreement.

^aIn the modified Guan et al. model, a covariance was added between the errors of EQ 43 and 36.

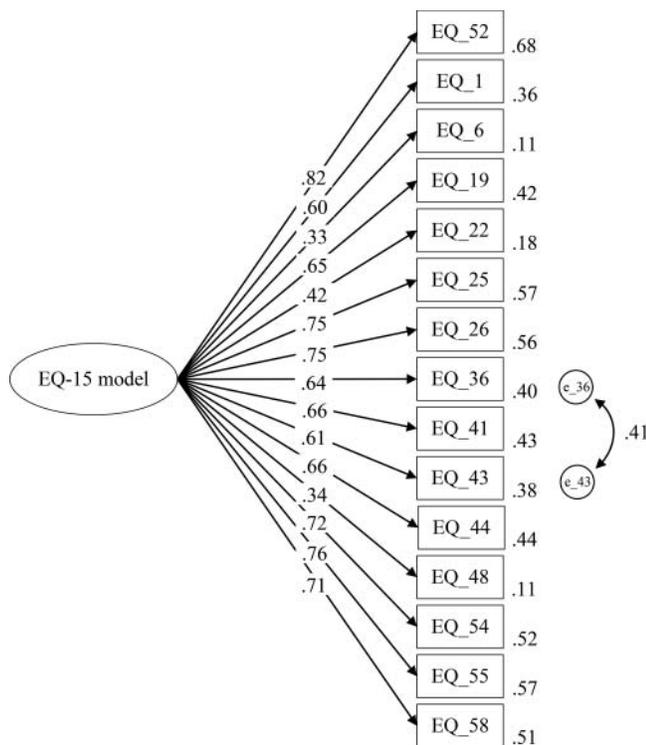


Figure 1. Confirmatory factor analysis standardized estimates of the EQ 15-item one-factor model. The EQ-15 model is based on Guan et al. (2012). Values to the left of the observed item variables represent standardized factor loading regression weights (b). Values to the right of the observed item variables represent the squared multiple correlations (R^2). The value to the far right on the error covariance pathway represents the correlation coefficient (r).

EQ-40 was calculated to provide comparability with previous studies (e.g., Groen et al., 2015).

Internal consistency and test-retest coefficients

Cronbach's α values for the scores on the EQ-40 and EQ-15 were both .86. Cronbach's α values for the scores on the other EQ models are provided in Table 2. The ICC (type 2, 1) between participants' test and retest total scores on the EQ-40 was .82 ($p < .001$), 95% CI [.67, .90] and on the EQ-15 was .68 ($p < .001$), 95% CI [.45, .82].

Evidence of construct validity

The intercorrelations between scores on the scales are shown in Table 3. Both EQ-40 and EQ-15 were positively correlated with IRI-PT, IRI-EC, and AQ-AD. Both of the EQ scores were negatively correlated with AQ-SS, AQ-CM, TAS-IF, and TAS-DF.

Sex differences

Mean scores on all scales for the final participants ($N = 588$) and the mean scores by sex are provided in Table 4. The female group was found to have a significantly higher EQ-40 score, but a similar EQ-15 score compared with the male group. We found that participants' age and education level correlated significantly with the scales used in the study. Therefore, the comparisons between males and females on these scales were

Table 3. Pearson correlation coefficients between the EQ, IRI, AQ, and TAS-20 scores based on a Mainland Chinese sample.

Scale	EQ-40	EQ-15
IRI-PT	.48	.38
IRI-EC	.34	.17
IRI-FS	.29	.20
IRI-PD	-.26	-.24
AQ-SS	-.40	-.42
AQ-CM	-.45	-.34
AQ-AD	.24	.32
TAS-IF	-.29	-.16
TAS-DF	-.36	-.26

Note. $N = 588$. EQ = Empathy Quotient; IRI = Interpersonal Reactivity Index; AQ = Autism-Spectrum Quotient; TAS-20 = 20-item Toronto Alexithymia Scale; EQ-40 = total score for the 40-item EQ; EQ-15 = total score for the 15-item EQ based on the Guan et al. (2012) model; IRI-PT = total score for the IRI's perspective-taking items; IRI-EC = total score for the IRI's empathic concern items; IRI-FS = total score for the IRI's fantasy items; IRI-PD = total score for the IRI's personal distress items; AQ-SS = total score for the AQ's social skill items; AQ-CM = total score for the AQ's communication skill items; AQ-AD = total score for the AQ's exceptional attention to detail items; TAS-IF = total score for the TAS's identifying feeling items; TAS-DF = total score for the TAS's describing feeling items. All correlations are significant at $p < .001$.

conducted again by controlling these variables as covariates. Results of the analyses with and without the covariates were similar and led to the same conclusions for the sex comparisons.

Discussion

A simplified Chinese version of the EQ (60 items) was validated in this study with a Mainland Chinese sample. The one-factor model with 15 EQ items (Guan et al., 2012) described the psychometric properties of the EQ based on the current Mainland Chinese participants quite well. Significant sex differences on EQ scores were found. The construct validity found in this study supports that the underlying concept measured by the EQ scores is empathy.

This study suggests using the EQ-15 model provided by Guan et al. (2012) as the structural model for the current simplified Chinese version of the EQ. This one-factor model supports the original proposal that Baron-Cohen and Wheelwright (2004) made for the EQ, namely, "an initial attempt to separate items into purely affective and cognitive categories was abandoned because in most instances of empathy, the affective and cognitive components co-occur and cannot be easily disentangled" (p. 166). This study, along with previous studies based on Chinese populations (Guan et al., 2012; Siu & Shek, 2005), provides evidence to support the notion of cooccurring emotional and cognitive empathy. However, this finding is different from several previous EQ validation studies based on other populations, such as British (Lawrence et al., 2004) and Italian (Preti et al., 2011) samples. These previous validation studies reported that EQ items could be psychometrically divided according to emotional and cognitive empathic components (Lawrence et al., 2004).

Researchers have considered that the blurring of the line between emotional and cognitive empathy found in the Chinese samples might be an adaptation to their cultural requirements for emotional communication (Neumann et al., 2016; Siu & Shek, 2005). On the one hand, Chinese people might be

Table 4. Descriptive statistics and *t*-test results for males and females on the EQ, IRI, AQ, and TAS-20 scores.

Scale	Overall (<i>N</i> = 588)		Females (<i>n</i> = 375)		Males (<i>n</i> = 213)		<i>t</i> ^a	<i>p</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
EQ-40	38.67	10.42	39.57	10.31	37.08	10.43	2.80	.005	0.24
EQ-15	14.70	5.54	14.66	5.45	14.76	5.72	-0.20	.839	-0.02
IRI-PT	17.27	3.49	17.19	3.56	17.42	3.37	-0.77	.440	-0.07
IRI-EC	18.63	3.68	18.93	3.67	18.10	3.65	2.65	.008	0.23
IRI-FS	19.06	4.10	19.66	4.00	18.00	4.05	4.82	< .001	0.41
IRI-PD	13.59	4.38	14.36	4.34	12.24	4.12	5.79	< .001	0.50
AQ-SS	4.34	2.48	4.43	2.44	4.19	2.54	1.11	.266	0.10
AQ-CM	2.74	1.93	2.61	1.76	2.97	2.19	-2.02	.044	-0.18
AQ-AD	5.07	2.27	4.99	2.32	5.20	2.19	-1.04	.300	-0.09
TAS-IF	18.14	5.09	18.29	5.05	17.87	5.15	0.94	.346	0.08
TAS-DF	13.62	3.55	13.63	3.53	13.59	3.58	0.16	.875	0.01

Note. EQ = Empathy Quotient; IRI = Interpersonal Reactivity Index; AQ = Autism-Spectrum Quotient; TAS-20 = 20-item Toronto Alexithymia Scale; EQ-40 = total score for the 40-item EQ; EQ-15 = total score for the 15-item EQ based on the Guan et al. (2012) model; IRI-PT = total score for the IRI's perspective-taking items; IRI-EC = total score for the IRI's empathic concern items; IRI-FS = total score for the IRI's fantasy items; IRI-PD = total score for the IRI's personal distress items; AQ-SS = total score for the AQ's social skill items; AQ-CM = total score for the AQ's communication skill items; AQ-AD = total score for the AQ's exceptional attention to detail items; TAS-IF = total score for the TAS's identifying feeling items; TAS-DF = total score for the TAS's describing feeling items.

^aEqual variances between sex groups could be assumed for most *t* tests (*df* = 586), except the one of the AQ-CM (*df* = 368.34).

influenced by Confucius's "Golden Mean" philosophy and this could lead to the control of emotional expression (Bond, 1993; Frijda & Sundararajan, 2007; Ho, Fu, & Ng, 2004). Furthermore, Chinese people are not encouraged to express their feelings verbally (Zhu et al., 2007). On the other hand, Chinese are expected to empathize with and respond to other persons' inner emotions and needs, using perspective taking, and being aware of other peoples' subtle emotional changes (Ho et al., 2004; Siu & Shek, 2005; Q. Wang, 2001). This study found there is a positive correlation between participants' EQ scores and their traits concerning exceptional attention to detail (i.e., AQ-AD). As a consequence, the dissociation between emotional and cognitive empathy could be attenuated in Chinese populations (Siu & Shek, 2005). Nevertheless, the direct correlation between empathy and emotional suppression needs more investigation.

In this study, female Mainland Chinese were found to have significantly higher scores on the EQ-40 than male Mainland Chinese. This is consistent with the commonly reported sex difference for the EQ-40 in Western cultures (Baron-Cohen & Wheelwright, 2004; Groen et al., 2015). However, the current effect size for the sex difference was small (Cohen's *d* = .24), which was only larger than that for the Korean version (Cohen's *d* = .11; Kim & Lee, 2010), and was lower than those for the original and for the other translated versions of the EQ (Cohen's *d* range = .39-.88; Groen et al., 2015). In contrast, this study did not find a significant sex difference based on the best fit EQ-15 model (Cohen's *d* = -.02). The finding of no significant sex difference on self-report empathy is consistent with previous findings in Chinese populations using both EQ and IRI (Guan et al., 2012; Siu & Shek, 2005; Yang et al., 2013). Results of this study seem to indicate a culture and sex interaction on self-report empathy. A cross-cultural comparison study confirmed the interaction in German and Mainland Chinese participant groups (i.e., the sex difference was larger in the former than the latter group; Melchers et al., 2015). Further study is required to explore the possible explanations for the culture and sex interaction on self-report empathy.

The mean value of the EQ-40 items for the current Mainland Chinese participants was 38.67 ± 10.42 . This value was

lower than the one reported in the original study conducted with British participants (Baron-Cohen & Wheelwright, 2004), and lower than most of the values reported for participants from other Western countries and cultures (Groen et al., 2015), but similar to those of East Asians, including Koreans (Kim & Lee, 2010) and Japanese (Wakabayashi et al., 2007). The Western-Asian differences on the EQ scores can be further investigated using a meta-analysis. These cultural differences on scores of self-report empathy might reflect that these cultures have different social expectations or requirements for empathy (Dehning et al., 2013).

Researchers have provided cutoffs of the total EQ score to provide the best discrimination point to separate clinical individuals (e.g., ASC) from general populations (Baron-Cohen & Wheelwright, 2004). The values of cutoffs were different between populations; less than 30 for British (Baron-Cohen & Wheelwright, 2004) and less than 33 for French Canadian populations (Lepage et al., 2009). Without including a group of clinical participants, this study cannot provide a value of the best cutoff for Mainland Chinese people. However, it should be noted that empathy is a trait that is continuously distributed in the general population (Chakrabarti, Bullmore, & Baron-Cohen, 2006). When researchers have tried to apply any specific cutoffs, they found that 12% to 54% of typical controls were defined as low EQ, whereas 19% to 59% of individuals with ASC were categorized as high EQ (Baron-Cohen & Wheelwright, 2004; Groen et al., 2015; Lepage et al., 2009). An increasing number of researchers have recommended not using cutoffs to divide behaviors into psychologically healthy and unhealthy (Keyes, 2002; Lichtenberg, Cassetta, & Scanlon, 1960; Melchers et al., 2015). Rather, they recommend using a concept of the mental health continuum. Therefore, it might be more meaningful to treat the EQ score as a continuous variable rather than as a taxonomic index.

This validation study has several limitations. The study was based on a convenience sample recruited online. It was not recruited based on census data in terms of proportional representation on ethnicity, age, sex, province, and so forth. As such, it is not a representative sample of Mainland Chinese.

Nevertheless, our sample included participants from all but one of the 31 provinces in Mainland China and included 28 minority Chinese from 12 ethnic groups. Further study is recommended to investigate empathy using a representative sample of Mainland Chinese. This sample also did not include a group of clinical participants. Further research is needed to investigate the utility and validity of scores on this translated version of the EQ in measuring self-report empathy in Chinese clinical populations (e.g., ASC). However, this study adopted the AQ and the TAS-20 to measure participants' autistic and alexithymic traits and found scores on these two scales were both negatively correlated with the EQ. The sample size of the retest administration was small and the test-retest duration was short. Therefore, more evidence on the stability of EQ scores in Mainland Chinese samples is required.

Due to the unavailability of Chinese indigenous supporting scales to validate the EQ, this study used a set of questionnaires originally developed in Western cultures. Some subscales were found to have low reliability based on this sample. This is a limitation of this study because the concept of empathy might not be equivalent between Mainland Chinese and Western cultures. Therefore, a concern might be raised that this study imposed on Mainland Chinese participants self-evaluation on a concept of empathy that is more suitable for Western culture than their own. Empirical cross-cultural comparison studies are needed to further answer this important question, namely, whether empathy measured by the EQ scores has the same meaning across cultures.

A further limitation is that this study did not test the best fitting models for each of the supporting scales (IRI, AQ, and TAS-20). The official scoring recommendations of these supporting scales were used instead. The correlations between scores on these supporting scales and the EQ were consistent with our hypothesis, research theories, and previous findings. However, the structural models of these three supporting scales need further investigation in a Mainland Chinese context.

Acknowledgments

We are grateful to all the participants who took part in this study, and thank Professor Raymond C. K. Chan for granting permission to use the Simplified Chinese version of the IRI and his feedback on this article. We thank Professor John O'Gorman for his feedback on this article, and Professors Graeme J. Taylor and Shuqiao Yao for their help in adapting the TAS-20 and providing permission to use the updated version of the TAS-20. We thank Dr. Kate Li and Ms. Yunhua Xiao for their help with the back-translation of the EQ and the TAS-20. Finally, we thank Dr. Amanda Biggs for providing us with technical support in conducting the confirmatory factor analyses.

Funding

This research was supported by a Griffith University Postgraduate Research Scholarship and a Griffith University International Postgraduate Research Scholarship to Qing Zhao.

ORCID

David L. Neumann  <http://orcid.org/0000-0001-5400-462X>
David H. K. Shum  <http://orcid.org/0000-0002-4810-9262>

References

- Allison, C., Baron-Cohen, S., Wheelwright, S., Stone, M. H., & Muncer, S. J. (2011). Psychometric analysis of the Empathy Quotient (EQ). *Personality and Individual Differences, 51*, 829–835. doi:10.1016/j.paid.2011.07.005
- Andrew, J., Cooke, M., & Muncer, S. (2008). The relationship between empathy and Machiavellianism: An alternative to empathizing–systemizing theory. *Personality and Individual Differences, 44*, 1203–1211. doi:10.1016/j.paid.2007.11.014
- Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I: Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*, 23–32. doi:10.1016/0022-3999(94)90005-1
- Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*, 163–175. doi:10.1023/B:JADD.0000022607.19833.00
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*, 5–17. doi:10.1023/A:1005653411471
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*, 3186–3191.
- Bédard, M., Martin, N. J., Krueger, P., & Brazil, K. (2000). Assessing reproducibility of data obtained with instruments based on continuous measurements. *Experimental Aging Research, 26*, 353–365. doi:10.1080/036107300750015741
- Berthoz, S., Wessa, M., Kedia, G., Wicker, B., & Grèzes, J. (2008). Cross-cultural validation of the Empathy Quotient in a French-speaking sample. *Canadian Journal of Psychiatry, 53*, 469–477. doi:10.1177/070674370805300712
- Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain, 133*, 1515–1525. doi:10.1093/brain/awq060
- Bond, M. H. (1993). Emotions and their expression in Chinese culture. *Journal of Nonverbal Behavior, 17*, 245–262. doi:10.1007/BF00987240
- Burnett, P. C., & Dart, B. C. (1997). Conventional versus confirmatory factor analysis: Methods for validating the structure of existing scales. *Journal of Research and Development in Education, 30*, 126–131.
- Chakrabarti, B., Bullmore, E., & Baron-Cohen, S. (2006). Empathizing with basic emotions: Common and discrete neural substrates. *Social Neuroscience, 1*, 364–384. doi:10.1080/17470910601041317
- Chan, C. (1986). *The relations between age, sex-role, orientation of human and empathy* (Unpublished master's thesis). Department of Education, National Chengchi University, Taipei.
- Cohen, D., & Strayer, J. (1996). Empathy in conduct-disordered and comparison youth. *Developmental Psychology, 32*, 988–998. doi:10.1037/0012-1649.32.6.988
- Cox, C. L., Uddin, L. Q., Di Martino, A., Castellanos, F. X., Milham, M. P., & Kelly, C. (2012). The balance between feeling and knowing: Affective and cognitive empathy are reflected in the brain's intrinsic functional dynamics. *Social Cognitive and Affective Neuroscience, 7*, 727–737. doi:10.1093/scan/nsr051
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Journal Supplement Abstract Service Catalog of Selected Documents in Psychology, 10*, 85.
- Decety, J., & Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: Implications for intervention across different clinical conditions. *BioPsychoSocial Medicine, 1*, 22–65. doi:10.1186/1751-0759-1-22
- Dehning, S., Gasperi, S., Tesfaye, M., Girma, E., Meyer, S., Krahl, W., ... Siebeck, M. (2013). Empathy without borders? Cross-cultural heart and mind-reading in first-year medical students. *Ethiopian Journal of Health Sciences, 23*, 113–122.
- Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D. I., Schneider, F., & Habel, U. (2010). Multidimensional assessment of empathic abilities: Neural correlates and gender differences. *Psychoneuroendocrinology, 35*, 67–82. doi:10.1016/j.psyneuen.2009.10.006

- Dimitrijević, A., Hanak, N., Vukosavljević-Gvozden, T., & Opačić, G. (2012). Psychometric properties of the Serbian version of the Empathy Quotient (S-EQ). *Psihologija*, *45*, 257–276. doi:10.2298/PSI1203257D
- Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews*, *35*, 903–911. doi:10.1016/j.neubiorev.2010.10.009
- Fleiss, J. L. (1999). *Design and analysis of clinical experiments*. New York, NY: Wiley.
- Frijda, N. H., & Sundararajan, L. (2007). Emotion refinement: A theory inspired by Chinese poetics. *Perspectives on Psychological Science*, *2*, 227–241. doi:10.1111/j.1745-6916.2007.00042.x
- Fritz, J. M., & Irrgang, J. J. (2001). A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Physical Therapy*, *81*, 776–788.
- Gremigni, P., Damásio, B. F., & Borsa, J. C. (2013). Development and validation of a questionnaire to evaluate overt aggression and reactions to peer aggression. *Psicologia: Reflexão e Crítica*, *26*, 311–318. doi:10.1590/S0102-79722013000200011
- Groen, Y., Fuermaier, A., Den Heijer, A., Tucha, O., & Althaus, M. (2015). The Empathy and Systemizing Quotient: The psychometric properties of the Dutch version and a review of the cross-cultural stability. *Journal of Autism and Developmental Disorders*, *45*, 2848–2864. doi:10.1007/s10803-015-2448-z
- Guan, R., Jin, L., & Qian, M. (2012). Validation of the Empathy Quotient–Short Form among Chinese healthcare professionals. *Social Behavior and Personality*, *40*, 75–84. doi:10.2224/sbp.2012.40.1.75
- Hart, D. L. (2003). Test–retest reliability of an abbreviated self-report overall health status measure. *Journal of Orthopaedic and Sports Physical Therapy*, *33*, 734–744.
- Ho, D. Y., Fu, W., & Ng, S. (2004). Guilt, shame and embarrassment: Revelations of face and self. *Culture & Psychology*, *10*, 64–84. doi:10.1177/1354067X04044166
- Huang, X., Li, W., Sun, B., Chen, H., & Davis, M. H. (2012). The validation of the Interpersonal Reactivity Index for Chinese teachers from primary and middle schools. *Journal of Psychoeducational Assessment*, *30*, 194–204. doi:10.1177/0734282911410588
- Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., ... Maier, W. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *The Journal of Neuroscience*, *30*, 4999–5007. doi:10.1523/JNEUROSCI.5538-09.2010
- Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, *7*, 95–109. doi:10.1111/j.1475-6811.2000.tb00006.x
- Jonason, P. K., & Krause, L. (2013). The emotional deficits associated with the Dark Triad traits: Cognitive empathy, affective empathy, and alexithymia. *Personality and Individual Differences*, *55*, 532–537. doi:10.1016/j.paid.2013.04.027
- Keyes, C. L. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior*, *43*, 207–222.
- Kim, J., & Lee, S. J. (2010). Reliability and validity of the Korean version of the Empathy Quotient scale. *Psychiatry Investigation*, *7*, 24–30. doi:10.4306/pi.2010.7.1.24
- Lawrence, E., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine*, *34*, 911–924. doi:10.1017/S0033291703001624
- Lepage, J.-F., Lortie, M., Taschereau-Dumouchel, V., & Théoret, H. (2009). Validation of French-Canadian versions of the Empathy Quotient and Autism Spectrum Quotient. *Canadian Journal of Behavioural Science*, *41*, 272–276. doi:10.1037/a0016248
- Levine, T. R. (2005). Confirmatory factor analysis and scale validation in communication research. *Communication Research Reports*, *22*, 335–338. doi:10.1080/00036810500317730
- Lichtenberg, P., Cassetta, R. K., & Scanlon, J. C. (1960). One description of mental health and disorder. *Archives of General Psychiatry*, *3*, 575–582. doi:10.1001/archpsyc.1960.01710060007002
- Lombardo, M. V., Barnes, J. L., Wheelwright, S. J., & Baron-Cohen, S. (2007). Self-referential cognition and empathy in autism. *PLoS ONE*, *2*, e883–e883. doi:10.1371/journal.pone.0000883
- Melchers, M., Montag, C., Markett, S., & Reuter, M. (2015). Assessment of empathy via self-report and behavioural paradigms: Data on convergent and discriminant validity. *Cognitive Neuropsychiatry*, *20*, 157–171. doi:10.1080/13546805.2014.991781
- Muncer, S. J., & Ling, J. (2006). Psychometric analysis of the Empathy Quotient (EQ) scale. *Personality and Individual Differences*, *40*, 1111–1119. doi:10.1016/j.paid.2005.09.020
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Neumann, D. L., Chan, R. C., Wang, Y., & Boyle, G. J. (2016). Cognitive and affective components of empathy and their relationship with personality dimensions in a Chinese sample. *Asian Journal of Social Psychology*, *19*, 244–253. doi:10.1111/ajsp.12138
- Nunes, P., Williams, S., Sa, B., & Stevenson, K. (2011). A study of empathy decline in students from five health disciplines during their first year of training. *International Journal of Medical Education*, *2*, 12–17. doi:10.5116/ijme.4d47.ddb0
- Parker, J. D., Taylor, G. J., & Bagby, R. M. (2003). The 20-Item Toronto Alexithymia Scale: III. Reliability and factorial validity in a community population. *Journal of Psychosomatic Research*, *55*, 269–275. doi:10.1016/S0022-3999(02)00578-0
- Penprase, B., Oakley, B., Ternes, R., & Driscoll, D. (2013). Empathy as a determining factor for nursing career selection. *Journal of Nursing Education*, *52*, 192–197. doi:10.3928/01484834-20130314-02
- Preti, A., Vellante, M., Baron-Cohen, S., Zucca, G., Petretto, D. R., & Masala, C. (2011). The Empathy Quotient: A cross-cultural comparison of the Italian version. *Cognitive Neuropsychiatry*, *16*, 50–70. doi:10.1080/13546801003790982
- Rueckert, L., & Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain and Cognition*, *67*, 162–167. doi:10.1016/j.bandc.2008.01.002
- Sass, D. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, *29*, 347–363. doi:10.1177/0734282911406661
- Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *Neuroimage*, *42*, 393–403. doi:10.1016/j.neuroimage.2008.04.180
- Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist*, *17*, 18–24. doi:10.1177/1073858410379268
- Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*, 617–627. doi:10.1093/brain/awn279
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. doi:10.1037/0033-2909.86.2.420
- Siu, A. M., & Shek, D. T. (2005). Validation of the Interpersonal Reactivity Index in a Chinese context. *Research on Social Work Practice*, *15*, 118–126. doi:10.1177/1049731504270384
- Smith, A. (2006). Cognitive empathy and emotional empathy in human behavior and evolution. *Psychological Record*, *56*, 3–21.
- Wakabayashi, A., Baron-Cohen, S., Uchiyama, T., Yoshida, Y., Kuroda, M., & Wheelwright, S. (2007). Empathizing and systemizing in adults with and without autism spectrum conditions: Cross-cultural stability. *Journal of Autism and Developmental Disorders*, *37*, 1823–1832. doi:10.1007/s10803-006-0316-6
- Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., ... Weil, L. (2006). Development of short forms of the Empathy Quotient (EQ–Short) and the Systemizing Quotient (SQ–Short). *Personality and Individual Differences*, *41*, 929–940. doi:10.1016/j.paid.2006.03.017
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. West Sussex, UK: Wiley.
- Wang, Q. (2001). “Did you have fun?": American and Chinese mother-child conversations about shared emotional experiences. *Cognitive Development*, *16*, 693–715. doi:10.1016/S0885-2014(01)00055-7
- Wang, Y., Neumann, D. L., Shum, D. H., Liu, W., Shi, H., Yan, C., ... Cheung, E. F. (2013). Cognitive empathy partially mediates the association between negative schizotypy traits and social functioning. *Psychiatry Research*, *210*, 62–68. doi:10.1016/j.psychres.2013.03.015

- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19*, 231–240.
- Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., ... Wakabayashi, A. (2006). Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Research, 1079*, 47–56. doi:10.1016/j.brainres.2006.01.012
- Williams, C., & Wood, R. L. (2010). Alexithymia and emotional empathy following traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 32*, 259–267. doi:10.1080/13803390902976940
- Wu, N., Li, Z., & Su, Y. (2012). The association between oxytocin receptor gene polymorphism (OXTR) and trait empathy. *Journal of Affective Disorders, 138*, 468–472. doi:10.1016/j.jad.2012.01.009
- Yang, N., Xiao, X., Qian, L., Mo, X., & Zhuo, S. (2013). Reliability and validity of the Chinese version of the Empathy Quotient. *Chinese Journal of Clinical Psychology, 20*, 760.
- Zhang, F., Dong, Y., Wang, K., Zhan, Z., & Xie, L. (2010). Reliability and validity of the Chinese version of the Interpersonal Reactivity Index-C. *Chinese Journal of Clinical Psychology, 18*, 155–157. doi: 10.16128/j.cnki.1005-3611.2010.02.019
- Zhang, Q., Wang, Y., Lui, S. S., Cheung, E. F., Neumann, D. L., Shum, D. H., & Chan, R. C. (2014). Validation of the Griffith Empathy Measure in the Chinese context. *Brain Impairment, 15*, 10–17. doi:10.1017/BrImp.2014.1
- Zhu, X., Yi, J., Yao, S., Ryder, A. G., Taylor, G. J., & Bagby, R. M. (2007). Cross-cultural validation of a Chinese translation of the 20-item Toronto Alexithymia Scale. *Comprehensive Psychiatry, 48*, 489–496. doi:10.1016/j.comppsy.2007.04.007