The psychometric properties of the 5-item gratitude questionnaire in Chinese adolescents

Y. ZENG ¹ M. Psy, Y. LING ² Ph.D, E. S. HUEBNER ² Ph.D, Y. HE ¹ M. Psy & X. LEI ³ M. Psy

¹ College of Education, Hunan Agriculture University, Changsha, China; ² Professor, Department of Psychology, University of South Carolina, Columbia, SC, USA; ³ Medical Psychological Center, The Second Xiangya Hospital, Central South University, Changsha, China

Keywords: gender differences, GQ-5, gratitude, measurement equivalent

Accessible summary

What is known on the subject?
• The GQ-6 is one of the most widely used self-report questionnaires to evaluate the level of gratitude among adults. The GQ-5 appears suitable for adolescents.

What this paper adds to existing knowledge?
• We developed a Chinese version of the GQ-5 and examined evidence for its reliability and validity. Results demonstrated adequate reliability and validity, indicating that it is appropriate for the assessment of gratitude in Chinese adolescents.
• In addition, Chinese early adolescent females reported higher gratitude than adolescent males.

What are the implications for practice?
• Screening adolescents who have lower levels of gratitude through the GQ-5 could help identify students who may benefit from empirically validated interventions to promote higher levels of gratitude in an effort to promote positive psychosocial and academic outcomes.

Abstract

Background: This study was conducted to evaluate the psychometric properties of the Chinese version of the 5-item Gratitude Questionnaire (GQ-5). Method: The sample consisted of 2093 middle school students (46.8% males) in mainland China. Confirmatory factor analysis and multigroup confirmatory factor analysis were performed to examine the factor structure and the measurement equivalence across gender. The convergent validity, Cronbach’s α and mean interitem correlations of the GQ-5 were also evaluated. Results: The results provided evidence of internal consistency reliability through a Cronbach’s α of 0.812 and a mean interitem correlation of 0.463 for the total sample. The results also supported a one-dimensional factor structure. In addition, convergent validity was assessed by statistically significant positive correlations between the GQ-5 and the two subscales of the Children’s Hope Scale (CHS) and the Brief Multidimensional Students’ Life Satisfaction Scale (BMSLSS) total score. Finally, multigroup confirmatory factor analysis also demonstrated measurement equivalence across gender. Subsequent analyses of latent mean revealed gender differences in early adolescent male and female students. Conclusions: The Chinese version of the GQ-5 appears to be a reliable and valid measure of gratitude among Chinese early adolescents. Early adolescent female students reported higher gratitude than early adolescent male students.
Introduction

With the introduction of positive psychology, there has been a recent surge of gratitude research. McCullough et al. (2002) defined gratitude as ‘a generalized tendency to recognize and respond with grateful emotions to other people’s benevolence in the positive experiences and outcomes that one obtains’ (p. 112). According to Emmons & McCullough (2004), gratitude has been described as the greatest virtue and the mother of all other virtues, and thus closely related to the other virtues. In McCullough’s theory, gratitude is regarded as a moral barometer (indicating the value of the relationship with the benefactor), a moral motive (prompting grateful people to behave prosocially) and a moral reinforcer (influencing the benefactor to provide gifts in the future) (McCullough et al. 2001). These three functions of gratitude result in positive emotions by creating a positive interaction between the beneficiary and the benefactor (Magno & Orillosa 2012).

Recently, scholars in psychology and related disciplines have focused on the relations between gratitude and important outcome variables. Specifically, previous studies indicate that higher levels of gratitude relate to positive outcomes, such as life satisfaction (Datu & Mateo 2015), optimism, happiness, hope (McCullough et al. 2002), subjective well-being and prosocial behaviour (Tian et al. 2015a), and better mental health and school performance (Bono et al. 2014), whereas lower levels of gratitude relate to negative outcomes, such as pessimism (Cohen 2006), envy, materialism (Lambert et al. 2009) and narcissistic personality (McCullough et al. 2001).

Although gratitude is thought to emerge in children between 7 and 10 years old (Weiner & Graham 2008), its development is crucial in early adolescence (Froh et al. 2010). Adolescence is a period of both mental and physical changes, characterized for many by increased autonomy from the family, peer conflicts and decreased subjective well-being, including life satisfaction (Steinberg 1987, Silverberg & Steinberg 1990). Gratitude may help adolescents foster prosocial relationships, self-esteem and well-being, which could help protect them against adverse changes (Gilman et al. 2014, Tian et al. 2015a). Early adolescence is also regarded as the developmental period when one can first experience and benefit most from higher levels of gratitude (Froh et al. 2010) Thus, a full understanding of the development of gratitude in adolescence would be valuable in facilitating optimal psychological growth in youth. Nevertheless, compared to that of adults, the research on gratitude in children and adolescents lags far behind.

Several measures of gratitude have been developed, some of which are appropriate for adolescents. One widely used measure has been the Gratitude Questionnaire (GQ-6), which was developed by McCullough et al. (2002) based on four qualities of gratitude including intensity, frequency, density and span. The GQ-6 is a one-factor self-report questionnaire with six items, and two of them are reverse-keyed (McCullough et al. 2002). Previous studies from Taiwan (Chen et al. 2009), America (Froh et al. 2011), Chile (Langer et al. 2016) and Turkey (Yüksel & Oguz Duran 2012) have revealed that a 5-item version (omitting Item 6) is more suitable for adolescents than the 6-item version because Item 6 showed a low factor loading and was reported to be difficult to understand by adolescents. Furthermore, subsequent research using the 5-item measure has demonstrated good reliability and validity.

The GQ-5 has been studied in a number of different countries with adolescent samples (Chen et al. 2009, Froh et al. 2011, Yüksel & Oguz Duran 2012, Kobayashi 2013, Langer et al. 2016), but, to our knowledge, most samples were from Western, individualistic cultures. There has been no study which has examined its reliability and validity with adolescents from mainland China. Given that the meaning of gratitude constructs may vary because of differing cultural characteristics, gratitude measures should be validated before being used in research (or practice) with different groups (e.g. culture, age) (İrmak & Kuruiüzüm 2008). Our first major goal was thus to evaluate evidence for the reliability and validity of the GQ-5 in a large sample of Chinese early adolescents.

Our second major goal was to evaluate gender differences in gratitude levels among female and male Chinese early adolescents. Gender differences in emotional expressions have been noted for a long time, including expressions of gratitude. Numerous studies show that women are generally more likely to experience and express gratitude than men (Gordon et al. 2004, Froh et al. 2009). Although evidence suggests the existence of gender differences in gratitude, the nature and magnitude of the relation between gender and gratitude in Chinese mainland adolescents remain unknown. To study further the possible differences between gratitude levels in early adolescent females and males, it is important to evaluate the measurement invariance of the GQ-5 across the two genders so that any differences in scores observed between boys and girls can be attributed to gender differences rather than measurement error. To do so, we employed multi-group confirmatory factor analysis (MCFA), which is a frequently used method to assess measurement invariance across groups or across time (French & Finch 2008).

To recapitulate, this study was conducted to assess the factor structure, reliability and the measurement invariance of the Chinese version of the GQ-5 across gender in
early adolescents, as well as evaluate latent mean differences across gender.

Method

Participants

The study sample consisted of 2093 students recruited from four middle schools in the Hunan Province of China. A total of 979 (46.8%) of participants were males and 1114 (53.2%) were females. The participants ranged in age from 11 to 16 years (M = 12.94, SD = 1.12). Most (95.6%) students were Han, the predominant ethnic group in China, and the rest were from various ethnic minority groups. Five per cent of the students were from low-income families (<30 thousand RMB), 16.6% were from relatively low-income families (from 30 to 80 thousand RMB), 49.9% were from medium-income families (from 80 to 120 thousand RMB), 20.4% were from relatively high-income families (from 120 to 300 thousand RMB) and 8.1% were from high-income families (higher than 30 thousand RMB). All participants spoke Mandarin. There was no significant age difference across gender (P > 0.05).

Procedure

Prior to participation, participants were informed about the purpose of this research, and consent forms were sent to their parents. Adolescents could participate only if they provided written parental consent and student assent. Participants completed questionnaires administered by trained research assistants during the school day in the students’ regular classrooms without the presence of teachers. The participants took about 30 min to complete the entire set of measures and retained the right to participate or withdraw at any time during the research. A total of 2234 students returned the consent and assent forms, but 52 of them withdrew and 89 omitted answering 20% or more of the questions. Therefore, 2093 students provided usable data, yielding an effective response rate of 93.7%. Students completed a demographics form and the following measures in counterbalanced order: (1) Gratitude Questionnaire-5 (GQ-5; McCullough et al. 2002); (2) Brief Multidimensional Students’ Life Satisfaction Scale (BMSLSS; Seligson et al. 2003); and (3) Children’s Hope Scale (CHS; Snyder et al. 1997). All measures were administered in the students’ primary language of Mandarin.

Ethics committees from the four middle schools’ administrations approved the research protocol. Furthermore, to ensure the rights of participants’ confidentiality and anonymity in this study, participants were assigned and identified by a unique code known only to the investigators.

Measures

Gratitude Questionnaire-5

The GQ-5 is a 5-item self-report scale with a one-factor structure. Each item was rated on a 7-point Likert-type scale (1 = strongly disagree and 7 = strongly agree), and the third item was negatively keyed (e.g. ‘When I look at the world, I don’t see much to be grateful for’). Possible scores ranged from 5 to 35, with higher scores indicating a higher level of gratitude. This scale was adapted from the GQ-6, which was developed by McCullough et al. (2002). The translation of the GQ was originally conducted by a postgraduate student majoring in psychology who had mastered both Chinese and English. After the GQ-5 was translated into Chinese, a second translator back-translated the items into English. Both Chinese and English items were also evaluated by the authors to ensure equivalence in meaning and comparability. In previous studies, the Cronbach’s α of the GQ-5 among adolescents ranged from 0.73 to 0.80 (Chen et al. 2009, Yüksel & Oguz Duran 2012, Langer et al. 2016).

Brief multidimensional students’ life satisfaction scale

The BMSLSS is a 6-item measure (e.g. ‘I would describe my satisfaction with my overall life as . . . ’), with each item rated on a 7-point scale (1 = terrible and 7 = delighted; Seligson et al. 2003). The total scores ranged from 6 to 42, with higher scores indicating higher general life satisfaction. The BMSLSS was developed to judge the perceived quality of life or life satisfaction (Diener & Seligman 2004). It consists of five items representing five major life satisfaction domains, including family, friends, school, self, living environment and one additional item representing overall life satisfaction. Previous research showed that the Cronbach’s α of the BMSLSS among Chinese adolescents was 0.77 (Tian et al. 2015b). In our study, the Cronbach’s α for the BMSLSS was 0.841.

Children’s hope scale

The CHS was developed for children ages 8 to 16 (Snyder et al., 1997). It consists of six items (e.g. ‘I think I am doing pretty well’), which are rated using a Likert-type format from 1 (none of the time) to 6 (all of the time), with total score ranging from 6 to 36. Consistent with Snyder’s hope theory, previous factor analyses of the CHS yielded two related, but distinguishable factors representing separable agency and pathways components of hope. Agency refers to the motivational ‘will’ to pursue goals,
whereas pathways refers to the cognitive ability to generate ‘ways’ to achieve the goals. Considering that agency and pathways are distinct concepts (Snyder et al. 1991), the scores on the hope subscales were utilized instead of the total score in this study. Previous research showed that the Chinese version of the CHS has good reliability and validity (Zhao 2011). In our study, the Cronbach’s α for the Agency subscale, Pathways Thinking subscale and total scale scores were 0.653, 0.651 and 0.800, respectively.

Data analysis

In this study, confirmatory factor analysis (CFA) and multigroup confirmatory factor analysis (MCFA) were conducted with AMOS 17.0 software using maximum likelihood (ML) as the estimation method. The other statistical analyses were conducted with SPSS 17.0.

Descriptive analyses were calculated for the five items of the Chinese version of the GQ-5. CFA s were performed to examine the factor structure of the Chinese version of the GQ-5. Traditionally, values of relative χ²/df of <2 or 3 indicate a good model fit (Hu & Bentler 1999). However, given the sensitivity of the χ² statistic to sample size (Joreskog & Sorbom 1996), we employed multiple complementary fit indices used to assess model fit: standardized root mean square residual (SRMR), non-normal fit index (NNFI/TLI), comparative fit index (CFI) and root mean square error of approximation (RMSEA). NNFI/TLI and CFI values above 0.95 (0.90) and SRMR and RMSEA values less than 0.05 (0.08) were considered acceptable (Schermelleh-Engel et al. 2003).

Next, the internal consistency of the GQ-5 was evaluated by calculating Cronbach’s α coefficient and mean interitem correlations. In addition, Pearson’s correlations with other theoretically related variables were conducted to examine the convergent validity.

Multigroup confirmatory factor analysis (MCFA) procedures were used subsequently to examine whether the GQ-5 demonstrated measurement invariance across male and female students. Because the GQ-5 represents a one-factor scale, there was no covariance. Four models were estimated by default in the following order: unconstrained model, measurement weights model, measurement intercepts model and measurement residuals model. The unconstrained model was the baseline model, which relaxed all equality constraints. The measurement weights model tested the invariance of factor loadings across groups by placing equality constraints on these parameters. The third model, in addition to loadings, also tested equal intercepts across groups. In the fourth model, the factor loadings, intercepts of variables and error variances were all set to be equal across gender (Liu 2011). The level of measurement equivalence was determined by comparing indexes. Considering the large sample of our study, we used ΔCFI to evaluate equivalence among different models and a ΔCFI ≤ 0.01 was considered evidence of equivalence (Cheung & Rensvold 2002).

Finally, the latent mean differences across gender were tested to examine whether the latent means of these groups were statistically significantly different. In the structural means model, the male group was selected as the reference group and the latent mean of male group was constrained to 0, while the female group was estimated freely. The critical ratio (CR) was chosen as the index to evaluate whether the latent means were different across gender (Tsaousis & Kazi 2013). If the CR > 1.96 or < -1.96, the estimate of equality was rejected. Moreover, positive CR values indicated a higher latent mean for the comparison group relative to the reference group, while negative values indicated a lower latent mean for the comparison group.

No missing data were included in the data analysis procedure. Participants whose questionnaires were blank or had missing values were excluded from the final database.

Results

Descriptive statistics

Descriptive statistics for the Chinese version of the GQ-5 items were calculated for males, females and total sample. These statistics, which are displayed in Table 1, indicate significant differences between male and female students, especially for the third item.

Factorial validity by confirmatory factor analysis

The fit indexes of scale indicated an adequate fit for three groups (total: χ² (df) = 68.818 (5), CFI = 0.980, TLI = 0.959, RMSEA = 0.078, SRMR = 0.025; male: χ² (df) = 30.386 (5), CFI = 0.980, TLI = 0.961, RMSEA = 0.072, SRMR = 0.0167; female: χ² (df) = 49.733 (5), CFI = 0.976, TLI = 0.952, RMSEA = 0.090, SRMR = 0.0282). Additionally, the standardized factor loadings for the five items represented in Fig. 1 were all significant, although Item 3 yielded a relatively lower factor loading.

Cronbach’s α and mean interitem correlation

The Cronbach’s α for the total, male and female samples was 0.794, 0.773 and 0.812, respectively. The mean interitem correlations for the total, male and female samples were 0.436, 0.405 and 0.463 respectively.
Convergent validity

The GQ-5 score positively related to the CHS Agency subscale score \((r = 0.390, P < 0.001)\), CHS Pathways Thinking subscale score \((r = 0.395, P < 0.001)\) and BMSLSS score \((r = 0.516, P < 0.001)\). See Table 2 for all correlations.

Measurement invariance

As mentioned previously, the \(\chi^2\) statistic is sensitive to the sample size. Thus, a \(\Delta\)CFI was utilized to examine the measurement invariance of the scale. As can be seen in Table 3, all \(\Delta\)CFIs were well below the cut-off of 0.01, indicating measurement equivalence.

Latent mean differences

Given the invariance of unconstrained, measurement weights and measurement intercepts models, a comparison of latent factor mean differences across gender was conducted. Males reported significantly lower gratitude scores than females \((CR = 2.644)\).

Discussion

The overarching goal of this study was to provide evidence of the reliability and validity of the 5-item Gratitude Questionnaire in a sample of Chinese mainland adolescents. Specifically, the results of CFAs demonstrated an adequate fit of the hypothesized one-factor model among Chinese adolescents, including female and male students. The values of CFI, TLI, RMSEA and SRMR of GQ-5 all exceeded the relative standards, which is consistent with previous studies with undergraduate students from Taiwan (Chen et al. 2009) and Turkey (Yüksel & Oguz Duran 2012), and children and adolescents from America (Froh et al. 2011) and Chile (Langer et al. 2016). Additionally, although Item 3 (‘When I look at the world, I don’t see much to be grateful for’) demonstrated a lower factor loading than the other items \((0.44/0.41/0.47)\), its item-total correlation was acceptable. We thus retained Item 3 in subsequent analyses based upon Chen et al.’s (2009) argument that its inclusion in the measure is essential to the construct validity of the underlying conceptual model of gratitude.

In relation to its reliability, we calculated Cronbach’s \(\alpha\) coefficients and mean interitem correlations for the total sample as well as for the female and male samples. The \(\alpha\) coefficients all exceeded the widely used cut-off of 0.7 for research purposes (Nunnally & New 1978). The \(\alpha\) coefficient for the female sample was a little higher than for the

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The descriptive statistics of the GQ-5 items</strong></td>
</tr>
<tr>
<td>Item</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Full sample/Male/Female ((n = 2093/979/1114))</td>
</tr>
<tr>
<td>1. I have so much in life to be thankful for</td>
</tr>
<tr>
<td>2. If I had to list everything that I felt grateful for, it would be a very long list</td>
</tr>
<tr>
<td>3. When I look at the world, I don’t see much to be grateful for</td>
</tr>
<tr>
<td>4. I am grateful to a wide variety of people</td>
</tr>
<tr>
<td>5. As I get older, I find myself more able to appreciate the people, events, and situations that have been part of my life history</td>
</tr>
</tbody>
</table>

SD, standard deviation; ITC, item-total correlation.
male sample, which means that females’ responses may be more internally consistent. Also, mean interitem coefficients for all three groups were well between the lowest (0.10) and the highest cut-off values (0.50) (Bentler 1990), indicating that Chinese early adolescents’ scores on the scale demonstrate good internal consistency.

Evidence for the convergent validity of the GQ-5 was also provided. The results demonstrated the expected, positive relations between GQ-5 scores and scores on the life satisfaction and hope (agency and pathways) measures. These results are similar to those of previous studies as well (Andersson et al. 2007, Loo et al. 2014, Datu & Mateo 2015).

Furthermore, our results demonstrated measurement invariance for the GQ-5 across gender. Thus, the results provided support for meaningful comparisons of gender differences among Chinese early adolescents. This finding was similar to that of Langer’s (2016) study.

Finally, tests of latent mean differences demonstrated that gratitude levels in Chinese early adolescents females were higher than levels for males, which is in line with previous studies with students from other nations (Froh et al. 2009, Kashdan et al. 2009, Wang et al. 2015). This difference may reflect that Chinese females are more emotionally expressive and pay more attention to emotions and complex emotional experiences than males (Kring & Gordon 1998, Ciarrochi et al. 2005, Naito et al. 2005, Kashdan et al. 2009). In China, parents’ expectancies and parenting styles differ for adolescent males and adolescent females (Wang et al. 2013). Chinese parents primarily aim to raise their sons to become strong, capable and independent men, whereas Chinese parents aim to raise their daughters to be kind-hearted, honest and considerate women. As a result, it is possible that Chinese adolescent females may be more willing than males to seek help from others, which makes them more open to experiences and expressions of gratitude.

**Limitation**

Limitations of our study should be noted. First, although a large sample was utilized, our participants were all middle school students from one province in Central China. A more nationally representative sample would increase the generalizability of the findings. Furthermore, considering that age relates to gratitude differences (Bono et al. 2014), longitudinal research is needed to investigate the stability of gratitude reports among youth as well as the antecedents and consequences of gratitude differences in different age groups. Also, considering that all of the criterion measures used were drawn from the ‘positive psychology’ literature, future research may be enhanced using additional more traditional ‘negative’ psychology measures, such as symptom-focused measure of psychopathology. Last, because the understanding of the GQ-5 items may vary with other demographic characteristics, such as age, ethnicity and region, our findings should not be generalized to different populations, such as children of other ages, adults or individuals from other nations.

**Conclusion**

Our study provides preliminary evidence for the reliability and validity of the GQ-5 with Chinese early adolescents. Our results also support the applicability of the GQ-5 across female and male Chinese early adolescents. Female students reported higher levels of gratitude than males. These findings suggest comparability of the meaning of the GQ-5 across cultures and should facilitate the study of the development and consequences of gratitude differences in adolescents from different cultures. The findings also revealed gender differences in Chinese adolescents’ levels of gratitude, suggesting the need for additional research on the development, nature and consequences of such gender differences in Chinese adolescents.

**Relevance statement**

Greater insight into the development of gratitude in adolescence is needed to improve our understanding of its antecedents, correlates and consequences across the life span. As previous studies have shown, lower level of gratitude has been related to negative outcomes, such as pessimism, envy and poorer mental health and academic outcomes, while higher levels of gratitude appear to protect them from adverse life experiences. Screening adolescents who have lower levels of gratitude through the GQ-5 could help identify students who may benefit from

---

**Table 3**

Multigroup analysis by gender

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$ (df)</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>$\Delta \chi^2$ (df)</th>
<th>$P$</th>
<th>$\Delta$CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>80.118 (10)</td>
<td>0.978</td>
<td>0.955</td>
<td>0.058</td>
<td>0.0316</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Model 2</td>
<td>84.331 (14)</td>
<td>0.978</td>
<td>0.968</td>
<td>0.049</td>
<td>0.0323</td>
<td>4.213 (4)</td>
<td>0.378</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>104.399 (19)</td>
<td>0.973</td>
<td>0.971</td>
<td>0.046</td>
<td>0.0324</td>
<td>20.068 (5)</td>
<td>0.001</td>
<td>0.005</td>
</tr>
<tr>
<td>Model 4</td>
<td>139.801 (25)</td>
<td>0.964</td>
<td>0.971</td>
<td>0.047</td>
<td>0.0396</td>
<td>35.402 (6)</td>
<td>0.000</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Model 1: unconstrained model; Model 2: measurement weights model; Model 3: measurement intercepts model; Model 4: measurement residuals model.
empirically validated interventions to promote higher levels of gratitude in an effort to promote positive psychosocial and academic outcomes in adolescents.

Acknowledgments
The research reported in this paper was supported by Ministry of Education in China (MOE) Youth Project of Humanities and Social Sciences (Grant No. 15YJC190012), Scientific Research Fund of Hunan Provincial Education Department (Grant No. 15A093), Hunan Provincial Social Science Foundation of China (Grant No. 16YBA199), and General project of Hunan Provincial Social Science Achievement Review Committee (Grant No. XSPYZZO27) awarded to Dr. Yu Ling.

References

© 2017 John Wiley & Sons Ltd


