ORIGINAL ARTICLE

Measurement invariances of the PHQ-9 across gender and age groups in Chinese adolescents

Doris Y.P. Leung PhD (ID) | **Yim Wah Mak PhD** (ID) | **Sau Fong Leung PhD** |
Vico C.L. Chiang PhD (ID) | **Alice Yuen Loke PhD**

School of Nursing, The Hong Kong Polytechnic University, Hunghom, Hong Kong

Correspondence
Dr Yim Wah Mak, School of Nursing, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong.
Email: yw.mak@polyu.edu.hk

Funding information
Central Research Grant, The Hong Kong Polytechnic University, Grant/Award Number: G-YL13

Abstract

Introduction: The nine-item Patient Health Questionnaire (PHQ-9) is widely used to determine the severity of depression in adult populations, but its psychometric properties with regard to adolescents has been poorly explored. The present study aims to identify the factor structure and examine the measurement invariance of this instrument across genders and age groups in a Chinese adolescent sample.

Methods: A large sample of Chinese schoolchildren completed the PHQ-9 in a cross-sectional survey in Hong Kong (N = 10 933). A confirmatory factor analysis (CFA) to test the factor structure and a multiple group CFA to test the gender and age invariances of the PHQ-9 in adolescents were conducted. Cronbach alpha was used to assess the reliability of the questionnaire, and Pearson correlations with anxiety, self-esteem, and perceived control were used to assess its construct validity.

Results: The CFA results indicate that a one-factor model with three pairs of item correlations fitted the PHQ-9 data well, and measurement invariances by age and gender were supported. The PHQ-9 also possesses adequate internal consistency (>.84) and is strongly correlated with anxiety (>.77), self-esteem (<−.57), and perceived control (<−.56) in the expected directions in the overall sample and in the gender and age subsamples.

Discussion: The results support the claim that the PHQ-9 is a reliable and valid scale and can be used to assess and compare depressive severity across ages and genders during the period of adolescence.

KEYWORDS

sage, Chinese adolescents, gender, measurement invariances, PHQ-9

1 | INTRODUCTION

The onset of depression often occurs in adolescence (Avenevoli, Swendsen, He, Burstein, & Merikangas, 2015; Rohde, Lewinsohn, Klein, Seely, & Gau, 2013), and those who experience the early onset of depression are at an elevated risk of subsequently manifesting psychosocial and health adversities (Jonsson et al., 2011; Naicker, Galambos, Zeng, Senthilselvan, & Colman, 2013; Quiroga, Janosz, Bisset, & Morin, 2013). Unfortunately, depression in youths is often undetected, and the undiagnosed are often misunderstood and

regarded as troublemakers or as unwilling to try hard enough to perform academically (Belfer, 2008). Mental disorders account for a large proportion of the disease burden in young people of all societies (Gore et al., 2011). Therefore, early recognition is crucial in order to initiate timely treatment and implement appropriate follow-up measures.

The Patient Health Questionnaire-9 (PHQ-9) is a self-administered instrument that was developed by Kroenke, Spitzer, and Williams (2001) to measure the severity of an individual's depression by evaluating the depressive episodes experienced by that individual in the past 2 weeks, using nine criteria based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). Over the past decades, the PHQ-9 is increasingly being used to determine the mental health of the general population and patient groups. The scale has been extensively validated in various adult populations, including clinical and community samples, and is well documented as having desirable psychometric properties. A recent systematic review concluded that the PHQ-9 is a valid instrument for use in identifying, differentiating, and monitoring depressive disorders (Kroenke, Spitzer, Williams, & Löwe, 2010).

The favorable psychometric properties of the PHQ-9 have also been demonstrated across culturally diverse groups through various translated versions, including Chinese, Arabic, German, Portuguese, Spanish, Tanzanian, Thai, and Korean. However, such psychometric properties in adolescent samples remain underresearched. To date, there have been few studies on the diagnostic accuracy of the PHQ-9 in identifying depression in adolescents (Allgaier, Pietsch, Frühe, Sigl-Glöckner, & Schutle-Körne, 2012; Ganguly et al., 2013; Richardson et al., 2010; Tsai et al., 2014). In addition, the consistency of its factor structure remains unclear. Three previous studies using adolescent samples reported that the PHQ-9 has a single factor (Burdzovic Andreas & Brunborg, 2017; Tsai et al., 2014; Yu, Tam, Wong, Lam, & Stewart, 2012), but mixed results have been reported on the equivalence of the meaning of items across genders. One study showed gender invariants in the PHQ-9 in a Chinese sample from the general population, which included adolescents (Yu et al., 2012), whereas another study reported that the PHQ-9 was measurement and scalar noninvariant across genders in a Norwegian adolescent sample (Burdzovic Andreas & Brunborg, 2017). To date, none of these studies have examined measurement invariants across different age subgroups within the period of adolescence, given that younger adolescents may not interpret the meaning of an item in the same way as their older counterparts.

Although the PHQ-9 can potentially facilitate comparisons across subpopulations, such as by gender, country, and age subgroup, it is a prerequisite for the items to have an equivalent meaning across the subpopulations under consideration. Thus, the current study sought to assess the usefulness of the PHQ-9 by examining its underlying factor structure using a large community sample of adolescents in Hong Kong. The aim was to (a) confirm the single-factor structure of the PHQ-9 and (b) examine the factorial invariance of the one-factor model across males and females and different age subgroups.

2 | METHODS

2.1 | Participants

This study is an analysis of secondary data derived from a cross-sectional survey aimed at exploring adolescent mental health risks and protective factors (Mak & et al, 2014). Students from 16 secondary schools from two districts in Hong Kong were randomly selected and were recruited to take part in the study between 2011 and 2014. After written informed consent was obtained from both the students and their parents, the students were instructed to anonymously complete the self-reported questionnaire. The survey was conducted by trained research assistants in a classroom setting with no teachers present. The students were assured that they could withdraw from the study at any time without penalty, that teachers and school principals would have no access to their responses, and that their data would be kept strictly confidential and used only for research purposes. Students who could not speak and read Chinese were excluded. The study was approved by the Human Ethics Committee of the participating university (the ethical approval letter was obtained on 11 January 2011) and was conducted according to the Declaration of Helsinki.

Of the 12 518 questionnaires that were collected, 522 did not contain all of the requested information on the PHQ-9, and 1063 did not contain information about the respondent's age or contained information indicating that the respondent was aged ≥18. Consequently, the sample for the current study comprised 10 933 adolescents. The average age of the respondents was 14.8 years (standard deviation [SD] = 1.6), with the following age distribution: 11 (0.4%), 12 (10.7%), 13 (14.8%), 14 (16.4%), 15 (18.6%), 16 (21.4%), and 17 (17.8%). The sample consisted of 4847 males, 6038 females, and 48 respondents who did not provide information about their gender.

2.2 | Measures

The PHQ-9 was designed to measure the presence and severity of states of depression (Kroenke et al., 2001). The scale consists of nine items that assess the frequency with which nine depressive symptoms had occurred in the past 2 weeks, with items rated on a 4-point Likert scale from "not at all" (0) to "nearly every day" (3). The responses to each item were summed to provide a single score ranging from 0 to 27, with higher scores indicating more severe depressive symptoms. According to the recommendation of Kroenke et al. (2001), depressive symptoms were classified by severity into five groups, namely, minimal (scores of 0-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20-27). The Chinese version of the PHQ-9 has demonstrated good psychometric properties in both adult and adolescent samples (Tsai et al., 2014; Yu et al., 2012; Zhang et al., 2013).

Anxiety was measured using the seven-item Generalized Anxiety Disorder Scale (GAD-7) developed by Spitzer, Kroenke, Williams, and Lowe (2006), with items rated on a 4-point Likert scale ranging from "not at all" (0) to "nearly every day" (3). An overall score of GAD-7

was calculated by summing the responses from the seven items, giving a possible range of 0 to 21, with higher scores indicating a higher level of anxiety. The scale is reliable ($\alpha = .89$) according to a previous study (Lowe et al., 2008), and the Cronbach alpha value of GAD-7 in the present study was .917.

Self-esteem was measured using the Rosenberg Self-esteem Scale (RSES; Rosenberg, 1986), which consists of 10 items rated on a 4-point Likert scale ranging from "completely disagree" (1) to "completely agree" (4). RSES scores can range from 10 to 40, with higher scores indicating a higher level of self-esteem. The Chinese version of the RSES is a reliable and valid measure in Chinese adolescent samples (Song, Thompson, & Ferrer, 2009; Chen et al., 2017). The RSES has acceptable internal consistency in the present study ($\alpha = .868$).

Perceived control was assessed using the seven-item Mastery Scale (Pearlin, Menaghan, Lieberma, & Mullan, 1981), which measures a person's general feeling of personal control over life events. Responses to all of the items were collected using a 5-point Likert scale, ranging from "strongly disagree" (0) to "strongly agree" (4). Responses from all of the items were summed to produce the overall score, which had a possible range of 0 to 24, with higher scores indicating a higher level of perceived control. The psychometric properties of the scale were satisfactory in Chinese adolescent samples (Shek, 2006). The Cronbach alpha value of the scale was .749 in the current study.

2.3 | Data analysis

The PHQ-9 was evaluated in terms of its factor-analytic structure, reliability, and construct validity. Confirmatory factor analyses (CFA) using Satorra-Bentler robust maximum likelihood estimation procedure in the EQS package (Bentler, 2006) were conducted to fit a single-factor model to the PHQ-9 items. The model was first fitted without allowing for any residual correlations among items. Then, Lagrange multiplier (LM) tests were used to identify item pairs that would improve the fit of the model if they were allowed to correlate. The modification was theoretically plausible. Then, multiple-group CFAs were performed to examine configural (an equal number of factors), metric (equal factor loadings), scalar (equal intercepts), and strict (equal error variances) invariances separately by gender and age. It was determined that there was a good fit between the model and the data, through the use of several fit indices, including (a) standardized root mean square residual ($SRMR$) ≤ 0.08, (b) robust comparative fit index (CFI) (R-CFI) ≥ 0.95, and (c) robust root mean square approximation (RMSEA) (R-RMSEA) ≤ 0.08 (Hu & Bentler, 1999). Measurement invariance was supported if reductions in the R-CFI were: ΔR-CFI ≤ 0.002 (Meade, Johnson, & Braddy, 2008). We combined the two age groups of 11 and 12 years old in the multiple group analysis, because only a few adolescents in the current sample were 11 years old.

Cronbach alpha was used to assess the internal consistency of the PHQ-9, and an alpha value of >.7 was considered a criterion of good reliability. The construct validity of the PHQ-9, with the total sample, gender subsamples, and age subsamples, was examined using Pearson correlation coefficients between the PHQ-9 score and the relevant measures, including GAD-7, RSES, and Mastery. We hypothesized that PHQ-9 scores would positively correlate with GAD-7 scores and negatively correlate with RSES and Mastery scores.

3 | RESULTS

Table 1 shows that the mean PHQ-9 score was 6.65 ($SD = 4.95$) for the whole sample. Males obtained slightly lower mean scores than females in the PHQ-9 (Cohen $d = -.15$; $P < .001$). Furthermore, the mean scores increased with age. The rates of the different severities of depressive symptoms, which were classified based on

TABLE 1 Descriptive statistics of PHQ-9 scores

| | N | Mean | SD | Level of Severity | | | | |
				Minimal (0-4)	Mild (5-9)	Moderate (10-14)	Moderately severe (15-19)	Severe (≥20)
Whole sample	10 933	6.65	4.95	38.9%	37.3%	16.1%	5.5%	2.1%
Gender								
Male	4847	6.26	5.09	43.4%	34.6%	14.5%	5.2%	2.4%
Female	6038	7.00	4.80	35.3%	39.4%	17.5%	5.7%	2.0%
Age group, year								
11-12	1207	5.20	4.82	54.9%	29.5%	9.7%	4.1%	1.8%
13	1615	5.81	4.93	48.9%	32.5%	11.7%	5.0%	2.0%
14	1789	6.37	4.79	41.7%	35.7%	16.2%	4.8%	1.7%
15	2036	6.87	4.93	36.9%	38.8%	16.2%	5.8%	2.3%
16	2338	7.22	4.91	31.4%	41.0%	19.0%	6.3%	2.3%
17	1948	7.57	4.95	29.4%	41.4%	20.2%	6.3%	2.6%

Abbreviation: PHQ-9, nine-item Patient Health Questionnaire.

TABLE 2 Fit indices for the PHQ-9 from confirmatory factor analyses by gender

	N	R-χ^2	SRMR	R-RMSEA	R-CFI	ΔR-CFI
Male	4847	313.494	0.030	0.050	0.966	-
Female	6086	405.736	0.029	0.051	0.963	-
Invariance test						
Configural	48	714.705	0.029	0.050	0.964	-
Metric	56	746.699	0.034	0.048	0.963	0.000
Scalar	65	1009.266	0.034	0.049	0.964	−0.001
Strict	75	1057.459	0.047	0.049	0.961	0.003
Strict + Error covariances	78	1054.767	0.048	0.048	0.961	0.000

Abbreviations: CFI, comparative fit index; PHQ-9, nine-item Patient Health Questionnaire; R-CFI, robust CFI; R-RMSEA, robust root mean square error approximation; SRMR, standardized root mean square residual.

TABLE 3 Fit indices for the PHQ-9 from confirmatory factor analyses by age

	N	R-χ^2	SRMR	R-RMSEA	R-CFI	ΔR-CFI
Age group						
11-12	1207	82.49	0.030	0.045	0.944	-
13	1615	122.30	0.032	0.050	0.962	-
14	1789	123.49	0.060	0.048	0.964	-
15	2036	108.48	0.026	0.042	0.975	-
16	2338	145.20	0.028	0.046	0.972	-
17	1948	219.32	0.036	0.065	0.948	-
Invariance test						
Configural	144	789.94	0.031	0.050	0.963	-
Metric	184	858.66	0.039	0.045	0.962	−0.000
Scalar	229	1403.31	0.045	0.047	0.963	−0.001
Strict	279	1498.53	0.054	0.049	0.961	0.002
Strict + Error covariances	294	1493.40	0.053	0.042	0.962	−0.001

Abbreviations: CFI, comparative fit index; PHQ-9, nine-item Patient Health Questionnaire; R-CFI, robust CFI; R-RMSEA, robust root mean square error approximation; SRMR, standardized root mean square residual.

TABLE 4 Correlations of PHQ-9 with criteria measures by gender and age group subsamples

	Whole sample	Male	Female	Aged 11–12	Aged 13	Aged 14	Aged 15	Aged 16	Aged 17
Self-esteem	−.593	−.581	−.601	−.601	−.602	−.580	−.602	−.590	−.578
Mastery	−.600	−.591	−.605	−.593	−.617	−.569	−.601	−.608	−.588
Anxiety	.791	.795	.784	.786	.795	.793	.783	.778	.792

Note: All P values are <.001.
Abbreviation: PHQ-9, nine-item Patient Health Questionnaire.

the recommendation by Kroenke et al. (2001) as minimal (scores of 0-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20-27), were 38.9%, 37.3%, 16.1%, 5.5%, and 2.1%, respectively, for the whole sample. Males and females were similar in terms of rates of moderately severe/severe depressive symptoms (7.6% vs 5.7%, P = .81), but the rates increased with age, from 5.9% in adolescents aged 11 to 12 to 8.9% in adolescents aged 17.

A one-factor model was fitted for the PHQ-9 items with no specification of residual correlations (robust χ^2 = 1624.43, df = 27, SRMR = 0.043, R-CFI = 0.916, and R-RMSEA = 0.074). LM tests indicated an improvement in model fit if the residuals of three pairs of item errors were allowed to correlate: items 1 and 2, items 3 and 4, and items 7 and 8. The model was refitted with specifications of the three correlated item errors, thereby improving the fit of the model (robust χ^2 = 700.69, df = 24, SRMR = 0.029, R-CFI = 0.957, and R-RMSEA = 0.051). Based on

the whole sample, standardized factor loadings for the items were moderately high, ranging from 0.531 to 0.742, and the three error correlations were minimal, ranging from 0.195 to 0.258.

Table 2 shows that the one-factor model with correlated item residuals for the PHQ-9 items fit the data well in both gender samples and that the factor structure (configural), factor loadings (metric), item intercept (scalar), and error variances (strict) were invariant across genders. Another model with equal error covariances for males and females was tested, and the noninvariance of the three error covariances between males and females was also supported. Similarly, Table 3 shows that the one-factor model with three error covariances for the PHQ-9 items also fits the data well in all six age groups. In addition, configural, metric, scalar, and strict invariances with error covariances invariant across age groups were observed.

The coefficient values of alpha for the whole sample was .859; those for the subsample of males and females were .866 and .852, respectively; and those for the age subsamples were .860, .860, .846, .857, .856, and .859 for the age groups 11, 12, 13, 14, 15, 16, and 17, respectively. As hypothesized, the PHQ-9 scores correlated positively and strongly with anxiety ($rs > .77$, $Ps < .001$) and negatively and moderately with self-esteem ($rs < -.57$, $P < .001$) and perceived control ($rs < -.56$, $Ps < .001$) in the whole sample, in the two gender subsamples, and in the six age subgroups (Table 4). The PHQ-9 demonstrated satisfactory construct validity in this study sample.

4 | DISCUSSION

We examined the factor structure of the PHQ-9 and investigated whether adolescents interpret and correspond to the items of the scale differently, depending on their gender and age. With a large community sample of adolescents, the CFA results supported a single-factor structure for the PHQ-9. We successfully replicated the result of the factor structure of the PHQ-9 in female and male subsamples and in the six age-group subsamples. Thus, the adolescents in this study showed that they interpreted the items on the scale in a similar manner, regardless of their gender and age. The internal consistency estimate and construct validity results for the PHQ-9 were also satisfactory for the overall, gender, and age-group subsamples.

Most previous studies showed measurement invariances of the PHQ-9 in adult samples (Galekamp, Stronks, Snijder, & Derks, 2017; Peterson et al., 2015; Schuler et al., 2018) and in the general population (Yu et al., 2012). Only one study examined the consistency of the factor structure of the PHQ-9 using adolescent samples (Burdzovic Andreas & Brunborg, 2017). Hence, there have been few investigations on the measurement invariances of the scale according to both gender and age. Our results provide important information on the psychometric properties of the PHQ-9. By separately demonstrating configural, measurement, scalar, and strict invariances across both genders and age subgroups, our study shows that the scale has a similar one-factor structure for both males and females and for all six age groups of Chinese adolescents. Thus, the observed difference in gender or age in the PHQ-9 scores among Chinese adolescents reflects a real difference, rather than factors related to the measurement of these symptoms.

Similar to previous population-based studies in Europe and Hong Kong (Galenkamp, Stronks, Snijder, & Derks 2017; Yu et al., 2012), the current study also found that the model fit improved substantially when the three error covariances (items 1 and 2, items 3 and 4, and items 7 and 8) were added to the model. Yu et al. (2012) had suggested the two possibilities—the wording of the items and the coexistence of symptoms—might have led to the observation of error covariances between the corresponding items. Another study of patients with chronic obstructive pulmonary disease (COPD) reported that the addition of the error covariance between items 3 and 4 to the one-factor model yielded a good fit to the data on the PHQ-9 items (Schuler et al., 2018). Meanwhile, Ishihara et al. (2019), using data from 7850 English-speaking participants from 20 studies, also reported that the fit of the model improved when error covariances among items 3, 4, and 5 were added to the one-factor model. A similar finding from these studies suggested that the coexistence of depressive symptoms may occur and that these symptoms may vary according to culture. Nevertheless, a further analysis of the coexistence of depressive symptoms is needed.

With regard to the strength of this study, the representative sample of the target population of adolescents in Hong Kong, which contains large and diverse age groups, could ensure the generalizability of the findings of this study. However, the study has several limitations. First, it is a secondary analysis of a study with other primary research questions, and no golden standard, such as DSM-IV, was used to assess depression. Therefore, criterion validity for the PHQ-9 was not examined in this study. Second, considering the nature of the cross-sectional design, we did not investigate the predictive power of the PHQ-9 in the case of health-related outcomes. Predictive power is one of the important psychometric properties of an instrument. Further studies to validate the scale using a more advanced study design, such as a longitudinal approach, are needed.

ORCID

Doris Y.P. Leung https://orcid.org/0000-0003-0138-8839
Yim Wah Mak https://orcid.org/0000-0001-7330-7001
Vico C.L. Chiang https://orcid.org/0000-0002-6973-7241

REFERENCES

Allgaier, A. K., Pietsch, K., Frühe, B., Sigl-Glöckner, J., & Schutle-Körne, G. (2012). Screening for depression in adolescents: Validity of the Patient Health Questionnaire in pediatric care. *Depression and Anxiety*, *29*, 906–913. http://doi.org/10.1002/da/1971

Avenevoli, S., Swendsen, J., He, J. P., Burstein, M., & Merikangas, K. R. (2015). Major depression in the National Comorbidity Survey–Adolescent Supplement: Prevalence, correlates, and treatment. *Journal*

of the American Academy of Child & Adolescent Psychiatry, *54*, 37–44. https://doi.org/10.1016/j.jaac.2014.10.010

Belfer, M. L. (2008). Child and adolescent mental disorders: The magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry*, *49*, 226–236. https://doi.org/10.1111/j.1469-7610.2007.01855.x

Bentler, P. M. (2006). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Burdzovic Andreas, J., & Brunborg, G. S. (2017). Depressive symptomatology among Norwegian adolescent boys and girls: The Patient Health Questionnaire-9 (PHQ-9) psychometric properties and correlates. *Frontiers in Psychology*, *8*, Article 887. http://doi.org/10.3389/fpsyg.2017.00887.

Chen, W., Zhang, D., Pan, Y., Hu, T., Jiu, G., & Luo, S. (2017). Perceived social support and self-esteem as mediators of the relationship between parental attachment and life satisfaction among Chinese adolescent. *Personality & Individual Differences*, *108*, 98–102. https://doi.org/10.1016/j.paid.2016.12.009

Galenkamp, H., Stronks, K., Snijder, M. B., & Derks, E. M. (2017). Measurement invariance testing of the PHQ-9 in a multi-ethnic population in Europe: The HELIUS study. *BMC Psychiatry*, *17*, 349. https://dx.doi.org/10.1186/s12888-017-1506-9

Ganguly, S., Samanta, M., Roy, P., Chatterjee, S., Kaplan, D. W., & Basu, B. (2013). Patient Health Questionnaire-9 as an effective tool for screening of depression among Indian adolescents. *Journal of Adolescent Health*, *52*, 546–551. https://doi.org/10.1016/j.jadohealth.2012.09.012

Gore, F. M., Bloem, P. J., Patton, G. C., Ferguson, J., Joseph, V., Coffey, C., ... Mathers, C. D. (2011). Global burden of disease in young people aged 10–24 years: A systematic analysis. *Lancet*, *377*(9783), 2093–2102. https://doi.org/10.1016/S0140-6736(11)60512-6

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 37–41. https://doi.org/10.1080/10705519909540118

Ishihara, M., Harel, D., Levis, B., Levis, A. W., Riehm, K. E., Saadat, N., ... Thombs, B. D. (2019). Shortening self-report mental health symptoms measures through optimal test assembly methods: Development and validation of the Patient Health Questionnaire-Depression-4. *Depression and Anxiety*, *36*, 82–92. https://doi.org/10.1002/da.22841

Jonsson, U., Bohman, H., von Knorring, L., Olsson, G., Paaren, A., & von Knorring, A. L. (2011). Mental health outcome of long-term and episodic adolescent depression: 15-year follow-up of a community sample. *Journal of Affective Disorders*, *130*, 395–404. https://doi.org/10.1016/j.jad.2010.10.046

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2010). The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry*, *32*, 345–349. https://doi.org/10.1016/j.genhosppsych.2010.03.006

Lowe, B., Decker, O., Muller, S., Brahler, E., Schellberg, D., & Herzog, W. (2008). Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. *Medical Care*, *46*, 266–274.

Mak, Y. W., Loke, A. Y., Chiang, C. L. V., Man, N., Ha, J., Leung, S. L., Chan, K. K., & Chow, S. Y. (2014). Anxiety and depressive symptoms among adolescents in Hong Kong: exploring risk and protective factors. Centre for Adolescent Mental Health Prevention and Intervention, Christian Family Service Centre. *Hong Kong*.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Naicker, K., Galambos, N. L., Zeng, Y., Senthilselvan, A., & Colman, I. (2013). Social, demographic, and health outcomes in the 10 years following adolescent depression. *Journal of Adolescent Health*, *52*(5), 533–538. https://doi.org/10.1016/j.jadohealth.2012.12.016

Pearlin, L. I., Menaghan, E. G., Lieberman, M. A., & Mullan, J. T. (1981). The stress process. *Journal of Health & Social Behavior*, *22*, 337–356.

Peterson, J,,. J., Paulitsch, M. A., Hartig, J., Mergenthal, K., Gerlach, F. M., & Gensichen, J. (2015). Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany. *Journal of Affective Disorders*, *170*, 138–142. http://doi.org/10.1016/j.jad.2014.08.053

Quiroga, C. V., Janosz, M., Bisset, S., & Morin, A. J. (2013). Early adolescent depression symptoms and school dropout: Mediating processes involving self-reported academic competence and achievement. *Journal of Educational Psychology*, *105*, 552–560. https://doi.org/10.1037/a0031524

Richardson, L. P., McCauley, E., Grossman, D. C., McCarty, C. A., Richards, J., Russo, J. E., ... Katon, W. (2010). Evaluation of the Patient Health Questionnaire-9 items for detecting major depression among adolescents. *Pediatrics*, *126*, 1117–1123.

Rohde, P., Lewinsohn, P. M., Klein, D. N., Seeley, J. R., & Gau, J. M. (2013). Key characteristics of major depressive disorder occurring in childhood, adolescence, emerging adulthood, and adulthood. *Clinical Psychology Science*, *1*, 41–53. https://doi.org/10.1177/2167702612457599

Rosenberg, M. (1986). *Conceiving the self*. Malabar, FL: Krieger.

Schuler, M., Strohmayer, M., Mühlig, S., Schwaighofer, B., Wittmann, M., Faller, H., & Schultz, K. (2018). Assessment of depression before and after inpatient rehabilitation in COPD patients: Psychometric properties of the German version of the Patient Health Questionnaire (PHQ-9/PHQ-2). *Journal of Affective Disorders*, *232*, 268–275. http://doi.org/10.1016/j.jad.2018.02.037

Shek, D. T. L. (2006). Assessment of perceived parental psychological control in Chinese adolescents in Hong Kong. *Research on Social Work Practice*, *16*, 382–391. https://doi.org/10.1177/1049731506286231

Song, H., Thompson, R. A., & Ferrer, E. (2009). Attachment and self-evaluation in Chinese adolescents: Age and gender differences. *Journal of Adolescence*, *32*, 1267–1286. https://doi.org/10.1016/j.adolescence.2009.01.001

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder—The GAD-7. *JAMA Internal Medicine*, *166*, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Tsai, F. J., Huang, Y. H., Liu, H. C., Huang, K. Y., Huang, Y. H., & Liu, S. I. (2014). Patient Health Questionnaire for school-based depression screening among Chinese adolescents. *Pediatrics*, *133*, e402–e409. http://doi.org/10.1542/peds.2013-0204

Yu, X., Tam, W. W. S., Wong, P. T. K., Lam, T. H., & Stewart, S. M. (2012). The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. *Comprehensive Psychiatry*, *53*, 95–102. http://doi.org/10.10106/j.comppsych.2010.11.002

Zhang, Y. L., Liang, W., Chen, Z. M., Zhang, H. M., Zhang, J. H., Weng, X. Q., ... Zhang, Y. L. (2013). Validity and reliability of Patient Health Questionnaire-9 and Patient Health Questionnaire-2 to screen for depression among college students in China. *Asia-Pacific Psychiatry*, *5*, 268–275. http://doi.org/10.1111.appy.12103