International Journal of Education and Human Developments, Vol. 11 No. 1; May 2026
ISSN 2415-1270 (Online), ISSN 2415-1424 (Print)
Published by Research Institute for Progression of Knowledge

Mental Health Stigma among Chinese College Students: Development and Validation of a Theory-Based Mental Health Stigma Measure

Chandra C. Díaz

Po Hu

Douglas R. Tillman

David Hof

University of Nebraska at Kearney
1615 W. 24th St.
Kearney, NE, 68849
USA

Abstract

Research in counseling and higher education recognizes the importance of mental health stigma of university students. However, assessing the attitudes and beliefs toward university students' mental health stigma further requires a rigorous instrument to measure such issues. Therefore, this study developed and tested a theory-based measurement of the University Student Mental Health Stigma Survey (USMHSS). The present study included 810 participants, of which, 606 completed the survey from more than 30 university majors from five geographic areas of China. The results of confirmatory factor analysis (CFA) and structural equation modeling (SEM) found a second-order measurement model and supported the theoretical structure of the mental health stigma model as it applied to this instrument. With the findings from this study, we suggested the modification to the USMHSS. Implications for research and practice and directions for future study are also discussed.

Keywords: mental health stigma, mental health counseling, construct validation, measurement development

The stigma around mental illness in China has been studied in recent years. Yang et al. (2020) discovered higher levels of stigma towards depression correlated with gender, employment status, and family functioning. The constructs of mental health literacy and stigma were explored from a sociodemographic position discovering that the portion of the Chinese population most likely to express stigmatization of mental illness were women, over the age of 50, with lower levels of education and an expressed occupation of retired or homemakers (Lo et al., 2021). Huang, Yang, and Pescosolido (2019) sought to further understand mental health literacy in China and found that a negative stigma associated with mental illness contributed to a reluctance to label symptoms of psychological distress as more socially accepted constructs such as stress and work problems. Xu, Li, Zhang and Wang (2018) found that the public, health professionals, and family members of patients reported the prevalence of stigma to be between 40 to 70 percent in China. Exacerbating this further is a lack of understanding of mainstream services due to language differences and traditional beliefs that mental illness is caused by supernatural forces (Yeung et al. 2017). Yin el al. (2020) found that participants were unfamiliar with the prevention, cause, and treatment of mental illness. Participants believed that others would hold a negative attitude towards individuals with mental health issues. Kudva et al. (2020) found that while there has been significant progress recently in China to combat the stigma of mental illness, there is still much to be done.

The research on instruments that specifically measure the stigma of mental illness, appears to be minimal. Link, Yang, Phelan, and Collins (2004) did a comprehensive review of measuring mental health stigma and while they identified factors and ways that stigma could be measured, instruments measuring the stigma of mental illness were not identified. Ritsher, Otilingam, and Grajales (2003) developed the Internalized Stigma of Mental Illness Inventory (ISMI) to be used with individuals with psychiatric disorders. While similar, the instrument developed in this current study builds on previous work from Diaz, Hu, Tillman and Hof (2019) and Hof et al. (2013) and is designed to measure the stigmatization of mental illness with college-aged participants. The focus of this study was to further develop a valid and reliable instrument to measure mental health stigma, with a Chinese population of college-aged participants.

Method

Sample and Data Collection Procedure

The participants were students from five comprehensive public universities in five major geographic sites across China. An online survey was administered to 810 randomly selected undergraduate and graduate students, based on convenience sampling during the COVID-19 global pandemic under the "Dynamic Zero COVID" policy in China. At the time of data collection, there was a reportedly low number of confirmed COVID infections and deaths in China compared to the rest of the world. As a result, 606 students completed the survey out of 639 respondents, resulting in a 79% response rate and 95% completion rate for the survey. The majority of respondents (89.4%) were 18-22 year-old students, 9.3% of them were 23 years and older, and 1.3% were 17 years and younger. Four hundred sixty-two (76%) respondents are female, one hundred thirty-nine (23%) are male, and five (1%) chose not specify their gender. Five hundred and six students (84%) identified as ethnic Chinese Han majority, and ninety (16%) identified as other Chinese ethnicities. Five hundred sixty-seven respondents (94%) were undergraduate students, thirty-nine (6%) were graduate students, and the participants were enrolled in over thirty academic majors.

Table 1

Demographics of participants, Descriptive Stats

Characteristics	[illegible]	ercent
Age		
17 and under	[illegible]	.3%
18	5	0.7%
19	48	4.4%
20	62	6.7%
21	24	0.5%
22	3	.1%
23 and above	6	.3%
Gender		
Female	62	6.2%
Male	39	2.9%
Would rather not specify	[illegible]	.9%
Ethnicity		
Han Majority	06	3.5%
Other Minorities	00	6.5%
Education Level		
Undergraduate	67	7.1%
Graduate	39	2.9%
Academic Major	0+	0+

Measures

Respondents were voluntarily recruited to participate in the online survey at five university research sites mentioned earlier. Student participants answered a self-developed cross-sectional 23-question University Student Mental Health Stigma Survey (USMHSS). The questionnaire was first developed in English in the United States and translated into Chinese by two English-and-Mandarin-Chinese-speaking mental health professionals. The Chinese version of the questionnaire was back-translated into English for further verification and was reviewed for a high level of accuracy. The bilingual mental health professionals and researchers approved the survey, indicating the instrument was complete. Chinese and English texts were both presented in the survey simultaneously during data collection to ensure high construct validity and reliability.

The questionnaire was developed and thrice modified from an earlier version of a similar college student mental health survey (Diaz, Hu, Tillman, & Hof, 2019). Survey items included questions regarding demographics and attitudes and beliefs toward mental health issues. Participants were then asked to indicate their answers to questions such as "I am willing to seek mental health/psychological counseling if needed" using a five-point-Likert scale from "not willing" to "very willing." Participants also answered questions indicating the level of their agreement on a five-point-Likert scale from "strongly disagree" to "strongly agree" to statements such as "I feel comfortable sitting next to a classmate in class who has a mental health issue". Students also answered several "Yes" or "No" to questions, such as “I have a relative who has a mental health issue."

Procedures and Analyses

The objective of the analyses was to develop a new instrument to measure university students' attitudes and beliefs toward mental health issues. Questions were modified to represent the most effective set of constructs measuring attitudes and beliefs toward mental health issues after a series of exploratory analyses (Diaz, Hu, Tillman, & Hof, 2019) and tested the measurement model using structural equation modeling (SEM) for the confirmatory analyses.

1. **Descriptive Statistics**

We conducted a reliability test showing the measure's Cronbach's αlpha = .90, indicating a high level of reliability and the questions' internal consistency in the survey. In addition, the questions passed the test of normality. Several answers showed a few outliers; however, the variables did not produce substantive differences in analyses with large sample sizes, as in this case (N=606). A further visual examination of the shape of the data distribution showed that almost all the variables did not deviate from the normal distribution curve.

2. **Confirmatory Factor Analysis and Structural Equation Modeling**

Because the USMHSS was initially developed from mental health counseling theories with a solid theoretical framework, confirmatory factor analysis (CFA) is an ideal method intended to examine the instrument by testing the fit of and factor loadings in its models and ultimately improve the instrument (Hurley et al., 1997). Therefore, in order to further refine and validate the USMHSS instrument, we used IBM SPSS Amos 29 with maximum likelihood estimation (ML) to conduct a series of confirmatory factor analyses (CFA) and structural equation modeling (SEM) for models with three different factor structures: a one-factor model, a first-order model including all the original items, and a second-order model after adopting modification indices.

Typically, the Chi-square Goodness of Fit test is used to evaluate whether a model departs significantly from one that fits precisely to the data (Kline, 2012). However, existing studies suggested that the overall chi-square statistic is influenced by the sample size, leading models with good fit to be rejected when the sample sizes are large (Gerbing & Anderson, 1992). We sought additional criteria for model fit statistics (a large sample size, N=606). The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are the most frequently reported comparative and incremental fit indices. Both represent the relative improvement in the fit statistics of the tested model compared to the null, independent, or baseline model. The values of CFI and TLI vary from 0 to 1, with a value of 1 indicating a perfect fit for the sample population, a value of .90 or higher being considered at least acceptable fit (Whittaker & Khojasteh, 2016), and a value of .95 or higher is considered a close or superior fit (Byrne, 2013). The Goodness of Fit Index (GFI) is the proportion of variance accounted for by the estimated population covariance. The value of GFI also varies from 0 to 1, with a value of .90 or higher indicating an acceptable fit and a value of .95 or higher a good fit (Pituch & Stevens, 2016).

The Root Mean Square Error of Approximation (RMSEA) is a parsimony-adjusted (absolute fit) index that estimates the discrepancy per degree of freedom between the original and reproduced covariance matrices. Browne & Cudeck (1993) and Kline (2012) recommended a value of RMSEA between .00 to .05 as a close fit and values up to .08 as reasonable or acceptable. The 90% confidence interval should be within these values as well. The Standardized Root Mean Square Residual (SRMR) represents the square root of the difference between the residuals of the original sample covariance matrix and the hypothesized model. A cutoff SRMR value of .08 was recommended by Hu & Bentler (1999), with smaller value of SRMR generally indicating a better fit. Finally, since we wanted to compare the fit of the original and modified models, we chose Akaike Information Criterion (AIC) to compare the tested models. A lower value of AIC generally indicates a more preferred model fit (Akaike, 1974).

Table 2
Results of Confirmatory Factor Analyses

Model	^{2}	lf	χ^2/df	FI	LI	FI	MSEA	RMR	IC
One-factor Model	33.007	1	.569	927	905	945	077	0617	87
First-order Model	2.142	2	.567	978	970	975	051	0519	28
Second-order Model	5.167	4	.299	986	979	981	046	0331	7

Note: All chi-square values are significant at p<.001. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; GFI = Goodness of Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion.
n = 606.

Results

Using the sample of 606 university students from China, we tested and compared the fit statistics of three different factor structures. The first was a one-factor model, in which all 23 items indicated one significant attitude and belief factor. The second was a first-order factor model in which items were loaded onto their corresponding factors (i.e., comfort, resistance, and environment), and the factors were allowed to correlate with each other. Finally, the third was a second-order factor model in which items were loaded onto their respective factors, and all three factors were loaded on a second-order latent attitude and belief factor. Such a method of testing different models with different factor structures was consistent with previous research using the confirmatory factor analysis and structural equation modeling method (Gerbing & Anderson, 1984).

According to Hair et al. (2010), standardized factor loading estimates in the range of .30 and .40 are considered to meet the minimum level. Factor loadings greater than .50 are considered practically significant. Loadings above .70 are considered indicative of a well-defined factor structure (Hair et al., 2010). In the original survey, item “I would feel comfortable dating someone who has a mental health issue” and item “Yes or No, I have a relative who has a mental health issue” had poor loadings on respective comfort factor (.26) and environment factor (-.04). Those two items according to the survey data were not strong indicators of their respective factors and should be removed. The factor loading of .38 for an original survey item “Mental health issues prevent students from performing their normal academic responsibilities” on resistance factor fell into a range of minimum acceptable level. We decided to retain the item and the reduced 10-item model became our first-order model. We also decided to drop the “Prevent Academic” item after adopting modification indices recommended by Amos, which resulted in our final 9-item second-order model. The standardized factor loadings for the reduced items varied from .54 to .89, indicating these items were practically significant and indicative of a well-defined structure with their corresponding factors. The standardized factor loadings of the second-order factor model are presented in Table 2. After the removal of the non-indicative items, the reduced 9-item instrument (4 items load on comfort, 2 items on resistance, and 3 items on environment) represented a good fit of the tested model to the sample population and a well-defined structure of three independent comfort, resistance, and environment factors in which measure university students’ attitudes and beliefs toward mental health issues.

Table 3

University Student Mental Health Stigma Survey Factor Loadings

Item	Comfort	Resistance	Environment
Willing Close Friend	89		
Willing Acquaintance	85		
Willing Myself	76		
Comfortable Sitting	54		
Uncomfortable Around		87	
Difficult Trust		67	
University Counseling			82
Professional Services			80
City Proactive			71

Note: All the factor loadings are significant at p <.001. Standardized factor loading estimates in the range of .30 and .40 are considered to meet the minimum level. Factor loadings greater than .50 are considered practically significant. Loadings above .70 are considered indicative of a well-defined factor structure (Hair et al., 2010).

The fit statistics from structural equation modeling (SEM) generated through Amos are shown in Table 3. In our findings, each model showed a statistically significant chi-square, which was expected because the chi-square test results are sensitive to the sample size (in our case, a large sample). As a result, we looked for other criterions mentioned earlier for model fit. For the overall one-factor model, the values of CFI (.927) and TLI (.905) were slightly above the acceptable level of .900. The value of GFI was .945, close to a good fit (.095). The value of RMSEA was .077 and SRMR was .0617 and fell into the recommended acceptable range (<.08), which indicated a minimum acceptable fit for the one-factor model. For the first-order model, the value of $\chi 2/df$ fell to 2.567. The values of CFI and TLI were significantly improved to .978 and .970, both indicating a good fit. At the same time, the value of GFI was improved to .975, indicating a good fit (>.950). However, the value of RMSEA was .051 and SRMR was .0519, very close to a good fit (<.05) but still not ideal. It indicated a potentially good fit for the first-order model. For the second-order model, although the chi-square statistic was still significant (p <.001), the value of $\chi 2/df$ fell to 2.299, a value close to a good fit ($\chi 2/df<2$) (Kline, 1998). Furthermore, CFI, TLI, and GFI values were further improved to .986, .979, and .981, all above the .95 reference level for a good or superior fit. Moreover, the RMSEA value was further reduced to .046 and the SRMR value fell to .0331, both met the recommended criteria for a good fit (<.05). The RMSEA values at the 90% confidence intervals (Lo 90=.030 and Hi 90=.063) were all registered lower than the cutoff value of .08, which also indicating a good fit. The results illustrated that the second-order model is the best-fitting, preferable model. When comparing different test models, The AIC values fell more than half from 287 at the one-factor model to 128 at the first-order model, and significantly fell further to 97 at the second-order model. These results proved that the second-order model showed excellent fit between the tested model and sample, and it also showed significantly better improvements than the one-factor and the first-order models.

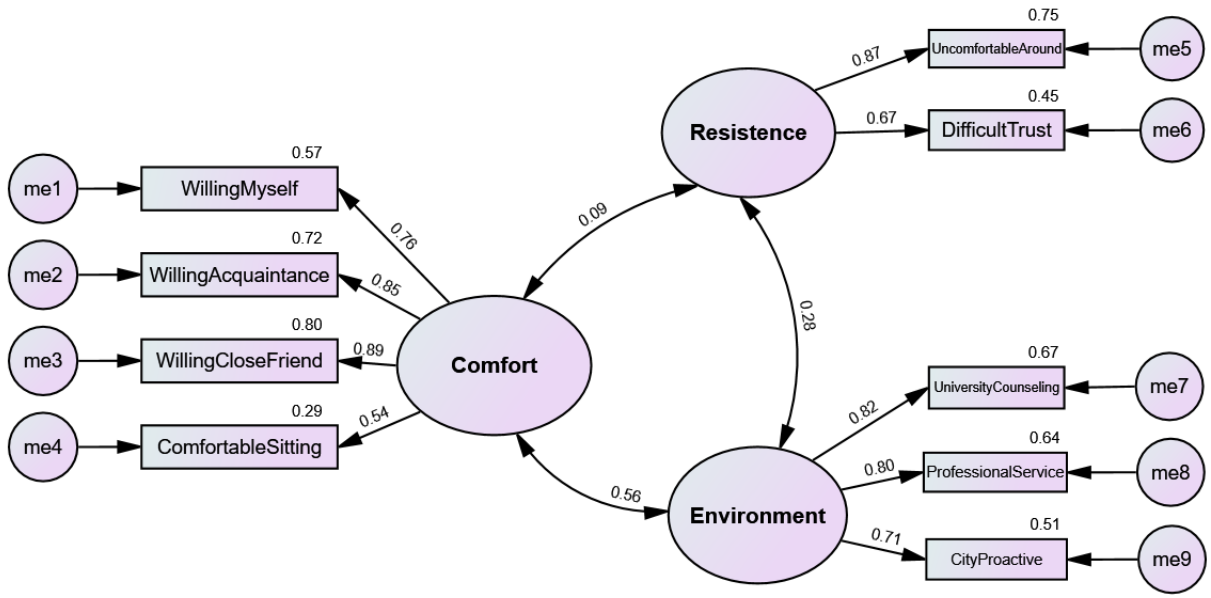

Figure 1. Results of the Second-Order Model Confirmatory Factor Analysis

In further examination, the values of composite reliability (CR) (Fornell & Larcker, 1981) for the three factors were high (Comfort .851, Environment .821, Resistance .750). The values of maximal reliability, MaxR(H), were also high (Comfort .892, Environment .829, Resistance .795). They indicate high internal consistency of the instrument.

The average variance extracted (AVE) was calculated to assess convergent validity of the factors. The maximum shared squared variance (MSV) was calculated to assess discriminant validity. The cut-off values recommended by Hu and Bentler (1999) were adopted. The findings of AVE were larger than .50 and they were all less than the values of CR and greater than the values of MSV, in which they met the cut-off criterion recommended by Hu and Bentler (1999). Such results showed good evidence of both convergent validity and discriminant validity. Table 4 presents the results of the reliability, validity, and factor correlations for the final model tested.

Table 4
Reliability, Validity, and Factor Correlations for the Final Model

	CR	AVE	MSV	MaxR(H)	Comfort	Environment	Resistance
Comfort	.851	.595	.310	.892	.771		
Environment	.821	.606	.310	.829	.557	.778	
Resistance	.750	.603	.080	.795	.091	.283	.777

Note: CR = Composite Reliability; AVE = Average Variance Extracted – Convergent Validity; MSV = Maximum Shared Squared Variance - Discriminant Validity; MaxR(H) = Maximal Reliability

In conclusion, the second-order reduced 9-item model had excellent fit statistics to the sample population. It's a preferred model and shows the best fit compared with the one-factor and the first-order model. The instrument with three independent factors showed high reliability and evidence of construct validity, including convergent validity and discriminant validity.

Discussion

Validity and Utility of Instrument

The survey instrument used in this study was a thrice modified instrument from Hof, Bishop, Dinsmore, Chasek, & Tillman (2013). While there is a strong case for wide-reaching impacts for usage in a variety of settings or locations, the number of participants currently totals 805. This includes 48 college student participants from the original study conducted at a midwestern higher education institution, 151 from the second study using a revised instrument conducted at a different Midwest higher education and recently with 606 participants who represent a variety of higher education institutions in China.

The survey was designed with three parts. The first part included demographic questions and the second part included statements measuring factors that potentially affect the participant's willingness "to actively seek mental health counseling or recommend mental health counseling to a peer" (Diaz et al., 2019, p. 16). The first survey modification took the Hof et al. instrument developed for mental health graduate students and revised it for pre-professional teachers. The second modification included specific references to a U.S. state and included changes to two demographic questions. In the third modification, language specific to the new higher education institution and two more questions were added. A new section was added because the survey would be sent out during the COVID-19 pandemic, the third section included COVID-19 questions to capture thoughts by Chinese college-aged participants living in a country that had the longest shutdown. The goals of this research were to continue to collect data that was a larger sample size than in previous studies, had unique demographics, and to further develop a valid and reliable instrument to measure mental health stigma.

We do believe there is a strong utility for the use of this survey instrument that will measure university students' attitudes and beliefs toward mental health issues. Given the validity and reliability of the instrument, there are wide-reaching implications for its use. We are confident in encouraging the use of this instrument in any country, with any culture, or languages.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155-173.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Sage.

Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.

Diaz, C., Hu, P., Tillman, D.R., & Hof, D.D (2019). Concepts Influencing Attitudes Toward Mental Health Stigma in a Teacher Education Program, *International Journal of Psychology: Biopsychosocial Approach.*

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research, 18*(1), 39-50.

Gerbing, D. W., & Anderson, J. C. (1992). Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods & Research*, *21*(2), 132-160.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate Data Analysis (7th Edition). NJ: Prentice Hall.

Hof, K., Bishop, M., Hof, D., Dinsmore, J. A., Chasek, C., & Tillman, D. R. (2013). Mental health stigma: Impact and interventions, In G. R. Walz, J. C. Bleuer, R. K. Yep (Eds.), *VISTAS 2013*, *1*. Alexandria, VA:American Counseling Association.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Huang, D. Yang, L. H., & Pescosolido, B. A. (2019). Understanding the public's profile of mental health literacy in China: a nationwide study. *BMC Psychiatry*, *19*(1), 20–20. https://doi.org/10.1186/s12888-018-1980-8

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of organizational behavior*, 667-683.

Kline, R. B. (1998). Software review: Software programs for structural equation modeling: Amos, EQS, and LISREL. *Journal of psychoeducational assessment*, *16*(4), 343-364.

Kline, R. B. (2012). Assumptions in structural equation modeling. *Handbook of structural equation modeling*, *111*, 125.

Kudva, K. G., El Hayaek, S., Gupta, A. K., Kurokawa, S., Bangshan, L., Armas-Villavicencio, M.V.C., Oishi, K., Mishra, S., & Tiensuntisook, S. (2020). Stigma of mental illness: Perspective from eight Asian nations. *Asia Pac Psychiatry.* https://doi.org/10.1111/appy.12380

Link, B. G., Yang, L. H., Phelan, J. C., & Collins, P. Y. (2004). Measuring mental illness stigma. *Schizophrenia bulletin*, *30*(3), 511-541.

Lo, L. L. H., Suen, Y. N., Chan, S. K. W., Sum, M. Y., Charlton, C., Hui, C. L. M., Lee, E. H. M., Chang, W. C. & Chen, E. Y. H. (2021). Sociodemographic correlates of public stigma about mental illness: a population study on Hong Kong's Chinese population. *BMC psychiatry*, *21*(1), 1-8. https://doi.org/10.1186/s12888-021-03301-3

Pituch, K. A., & Stevens, J. P. (2016). *Applied Multivariate Statistics for the Social Sciences.* (6th edition). New York and London.

Ritsher, J., Otilingam, P.G., & Grajales, M. (2003). Internalized stigma of mental illness: Psychometric properties of a new measure. *Psychiatry Res, 121*(1), 31-49.

Whittaker, T. A., & Khojasteh, J. (2016). Detecting Appropriate Trajectories of Growth in Latent Growth Models: The Performance of Information-Based Criteria. *The Journal of Experimental Education*, *85*(2), 215–230. https://doi.org/10.1080/00220973.2015.1123669

Xu, X., Li, X. M., Zhang, J., & Wang, W. (2018. Mental health-related stigma in China. *Issues in Mental Health Nursing*, *30*(2), 126-134.

Yang, F., Yang, B. X., Stone, T. E., Wang, X. Q., Zhou, Y., Zhang, J., & Jiao, S. F. (2020). Stigma towards depression in a community-based sample in China. *Comprehensive Psychiatry*, *97*. https://doi.org/10.1016/j.comppsych.2019.152152

Yeung, E. Y. W., Irvine, F., Ng, S. M., & Tsang, K. M. S. (2017). How people from Chinese backgrounds make sense of and respond to the experiences of mental distress: Thematic analysis. *Journal of Psychiatric and Mental Health Nursing*, *24*(8), 589–599. https://doi.org/10.1111/jpm.124066

Yin, H., Wardenaar, K. J., Xu, G., Tian, H., & Schoevers, R. A. (2020). Mental health stigma

and mental health knowledge in Chinese population: A cross-sectional study. *BMC Psychiatry, 20*. https://doi.org/10.1186/s12888-020-02705-x