Article

Journal of Psychoeducational Assessment
1–13

DOI: 10.1177/0734282918789561
journals.sagepub.com/home/jpa
SAGE

# The Chinese Version of the EPOCH Measure of Adolescent Well-Being: Testing Cross-Cultural Measurement Invariance

**Margaret L. Kern[1], Guang Zeng[2], Hanchao Hou[1], and Kaiping Peng[2]**

**Abstract**

Recent decades have brought growing interest in understanding and measuring psychological well-being. Although multiple measures of well-being exist, most were developed with Western populations. The current study tested the factor structure of a Chinese translation of the engagement, perseverance, optimism, connectedness and happiness (EPOCH) Measure of Adolescent Well-Being with 3,629 Chinese students (1,980 males, 1,649 females), and tested measurement invariance. The five-factor structure of the model was supported, and the model was invariant across age and gender. Combined with data from 2,041 American and 1,057 Australian adolescents, measurement invariance across cultures was supported for factor loadings but not intercepts or residuals. Results suggest that the factor structure is adequate across cultures, but the mean scores should not be directly compared. The findings support the EPOCH measure as an adequate scale, raise questions about different modeling decisions, and inform culturally sensitive approaches to comparing positive psychological variables across cultures.

**Keywords**

Well-being measurement, adolescents, measurement invariance, confirmatory factor analysis, bifactor model, cross-cultural comparisons

## Introduction

Positive psychology, as a scientific field that studies well-being, has developed rapidly over the past few decades (Donaldson, Dollwet, & Rao, 2015; Rusk & Waters, 2013). Multiple theories and corresponding measures of well-being exist (e.g., Butler & Kern, 2016; Diener et al., 2010; Huppert & So, 2013; Rusk & Waters, 2015; Ryff, 1995). Across the various definitions and models, studies clearly show that many desirable variables correlate with high levels of well-being. However, most existing models and measures were tested with adults (Rose et al., 2017), and most measures were developed with Western samples. The extent to which well-being measures—and related associations—generalize cross-culturally is questionable.

[1]The University of Melbourne, Victoria, Australia
[2]Tsinghua University, Beijing, China

**Corresponding Author:**
Guang Zeng, Department of Psychology, Tsinghua University, Haidian District, Beijing 100084, People's Republic of China.
Email: cengg14@mails.tsinghua.edu.cn

The current study examines a measure of well-being that was specifically developed for adolescents—the engagement, perseverance, optimism, connectedness and happiness (EPOCH) Measure of Adolescent Well-Being (Kern, Benson, Steinberg, & Steinberg, 2016)—which assesses five positive psychological characteristics: *Engagement, Perseverance, Optimism, Connectedness*, and *Happiness*. A series of studies with Australian and American students provided initial information about the reliability of scores and the validity of inference made from the measure. In this article, we consider the extent to which the measure generalizes to Chinese students, testing the factor structure and measure invariance across gender, age, and culture.

## *Well-Being Definitions and Measurement*

Definitions of well-being have been considered and debated across the millennia (Froh, 2004), and appear throughout different disciplines (e.g., philosophy, theology, psychology, medicine, economics, law). In general, well-being refers to feeling good and functioning well across multiple life domains (Forgeard, Jayawickreme, Kern, & Seligman, 2011; Huppert & So, 2013; McQuaid & Kern, 2017). The specific components that are included within any given definition of well-being depend on the theorist. For instance, Keyes' (2002) definition includes emotional, psychological, and social domains. Diener (1984) suggested affective and evaluative components, and other theorists add eudaimonic elements, such as meaning and purpose, self-acceptance, engagement, and optimism (e.g., Ryff, 1995; Seligman, 2011).

Foci on children and adolescents have occurred primarily within developmental psychology and related areas (Kern, Park, Peterson, & Romer, 2017). Scales, Benson, Leffert, and Blyth (2000) suggested that adolescent well-being is indicated by seven thriving behaviors (school success, leadership, helping others, maintenance of physical health, delay of gratification, valuing diversity, and overcoming adversity). The Child and Adolescent Wellness Scale (CAWS; Copeland, Nelson, & Trauberger, 2010) includes 10 domains (adaptability, connection, conscientiousness, emotion regulation, empathy, initiative, mindfulness, optimism, self-efficacy, and social competence). Others identify positive characteristics, such as character, optimism, connectedness, competence, and perseverance, as important components of well-being (e.g., Lerner, Fisher, & Weinberg, 2000; Roth & Brooks-Gunn, 2003).

We focus here on the EPOCH model (Kern et al., 2016), which includes five positive characteristics. Engagement refers to the capacity to become absorbed in and focused on what one is doing. Perseverance means the ability to work hard and pursue one's goals to the end even when facing obstacles. Optimism is conceptualized as hopefulness and confidence about the future, a tendency to take a favorable view of things, and as an explanatory style marked by evaluating negative events as temporary, external, and specific to situation. Connectedness refers to the sense of having satisfying relationships with others, believing that one is cared for, loved, esteemed, and valued and providing friendships or support to others. Happiness is characterized as being generally happy, fun loving, and content.

Based on this theoretical model, Kern and colleagues (2016) developed a 20-item measure to capture these domains. The authors conducted a series of 10 studies with 4,480 adolescents (age = 10-18 years) from the United States and Australia. The measure had high internal consistency (range $\alpha$ = .85-.95 for overall well-being), test–retest stability in the short term, with lower stability over extended time periods (ranging from $r$ = .71 over 3 weeks to $r$ = .39 over 3 years). It demonstrated measurement invariance across age, gender, and culture (United States vs. Australia). There was also initial evidence of convergent and discriminant validity (see Kern et al., 2016). For instance, meta-analytically combining the samples, Engagement was moderately correlated with measures of school engagement ($r$ = .40) and meaning and purpose ($r$ = .38). Perseverance was strongly correlated with grit ($r$ = .78) and school engagement ($r$ = .58). Optimism was strongly correlated with life satisfaction ($r$ = .64), meaning and purpose ($r$ = .57), and positive affect ($r$ = .58). Connectedness was strongly correlated with measures of parent, teacher, and peer connection ($r$ = .41-.63). Happiness was robustly correlated with life satisfaction ($r$ = .83) and positive affect ($r$ = .59). The domains were moderately correlated with less depression and anxiety, with stronger associations for Optimism, Connectedness, and Happiness, and weaker correlations for Engagement and Perseverance.

Rose and colleagues (2017) conducted a systematic review of adolescent well-being measures. Of 12 identified measures, only four were specifically developed for adolescents, one of which was the EPOCH measure. The others were the CAWS, the patient generated index (PGI; Verma, Dubey, & Gupta, 1983), and the Social and Emotional Health Survey (SEHS; Furlong, You, Renshaw, Smith, & O'Malley, 2014). The EPOCH and PGI measures include both feeling and functioning domains and are each 20 items long. The PGI includes physical, anxiety, mood, and self/other domains, but is not publicly available. The CAWS and SEHS focus primarily on functioning, and are longer measures (100 and 36 items, respectively). Of the available measures, EPOCH is a brief measure that was developed specifically with adolescents. It reflects domains with predictive ability, and, thus, has practical appeal to educators. It is freely available for research purposes and has demonstrated adequate psychometric properties in its development across two cultures. Thus, we chose to consider whether the measure can be extended to a very different culture—China.

## *Invariance Across Cultures*

Over the past two decades, cross-cultural researchers have advocated for the need to establish measurement equivalence for psychometric measurements across culturally diverse groups (Van de Vijver & Leung, 1997). Comparisons between cultural groups are appropriate only after confirming measurement invariance (Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014). Different degrees of invariance are possible, with implications for what comparisons can be made (Reise, Widaman, & Pugh, 1993). At the weakest level, configural invariance (or nonmetric invariance) specifies a similar factor pattern across groups (i.e., the same items loading on the same factors), without imposing additional constraints. In this case, we specify the five EPOCH factors with the same four items loading on each factor, but do not constrain the factor loadings to equality. This would suggest that the same model occurs across cultures, but the items that make up that factor are weighted differently. Comparison of scores across the two groups is questionable and potentially misleading (Reise et al., 1993).

For cross-group comparisons to be made, at least some level of metric invariance is needed (Widaman & Reise, 1997). Metric invariance implies that populations from different cultural groups understand the concept of the items identically and, therefore, respond to the measure in the same way. Meredith (1993) suggested weak, strong, and strict forms, which are tested by systematically constraining parameters to equality. As additional constraints are added, invariance is harder to achieve, but stronger statements about cross-group differences can be made. Weak invariance constrains the factor loadings ($\lambda$) to equality. Strong invariance adds constraints to the intercepts ($\tau$). This would mean that norm values developed in one culture would extend to the other culture. Strict invariance adds constraints to the residuals ($\varepsilon$). Although strict invariance allows stronger comparisons to be made, there are numerous reasons why strict invariance will not hold; for substantive comparisons, equality across the factor loading and intercepts is most crucial (Widaman & Reise, 1997).

In the current study, we translated the EPOCH measure to the Chinese and tested it with a large sample of Chinese students, testing the factor structure and invariance across age and gender. We then combined the Chinese data with U.S. and Australian samples used in the original paper (Kern et al., 2016), and examined measurement invariance across the three samples. In addition, we further explored the factor structure of the measure, comparing one-factor, five-factor, higher order, and bifactor models.

# Method

## Participants

Students from three schools in different cities in China were asked to complete a survey during class in a single week at the beginning of the semester. Parents were informed of the study and could opt out of allowing their child to participate, and students consented to their own participation. Responses were received from 3,629 students,[1] ranging in age from 6 to 19 years. As the original measure was designed and tested with participants who were 9 years and above, nine students aged 6 to 8 years were excluded from the current analysis, leaving a final sample size of 3,620 students (1,975 males, 1,645 females). Students ranged in age from 9 to 19 years ($M$ = 14.11 years, $SD$ = 2.91 years), and represented a wide variety of demographic characteristics (specific demographics for individual students were unavailable). The sample included 1,219 students from a primary school, ranging in age from 9 to 13 years ($M$ = 10.20 years, $SD$ = 0.90 years), and 2,401 students from two secondary schools, ranging in age from 13 to 19 years ($M$ = 16.10 years, $SD$ = 0.80 years).

## Measures

Students completed an online survey that included the EPOCH measure (Kern et al., 2016). The 20-item scale consists of five subscales, with four items per domain. Items are scored on a 1 to 5 scale (*not at all like me* = 1, *very much like me* = 5) and averaged together to form each domain. Overall well-being is calculated as the average of the five domains. The survey also included questions on age, gender, physical health, school engagement, academic performance, anxiety, depression, and overall happiness[2]; the current study focuses on age, gender, and the 20 EPOCH items.

The survey was translated to Chinese, following Brislin's (1980) translation/back-translation procedure. First, two Chinese bilinguals (one who had lived in the United States for 6 years and the other who had lived in the United Kingdom for more than 6 years) translated the items into Chinese. Discrepancies were discussed and resolved, resulting in a Chinese version of the survey. A Chinese translation professional then translated the Chinese questions into English. Three PhD students from Tsinghua University and the primary author of the original EPOCH measure examined the English version and the back-translated English version, agreeing that the translated version captured the language and meaning of the original measure. Online Appendix 1 provides the final Chinese measure, with questions ordered as presented to the students.

## U.S. and Australian Comparison Samples

To consider cultural measurement invariance, we included data from six of the samples from the original validation study (Kern et al., 2016).[3] Samples included (a) 1,515 adolescents of mixed socioeconomic status (SES) backgrounds from cities across the United States; (b) 730 male students from a private independent all-boys Anglican school in Adelaide, Australia; (c) 290 low SES inner city adolescents from Philadelphia, Pennsylvania; (d) 327 middle to upper middle SES adolescent girls from a Catholic all-girls school in Victoria, Australia; (e) 146 U.S. hospitalized adolescents who were receiving treatment for various medical conditions; and (f) 90 low SES students from a public school in Atlantic City, New Jersey. In total, the comparison samples included 2,041 participants (944 male, 876 female, 221 unknown) from the United States and 1,057 participants (730 male, 327 female) from Australia.

## Data Analyses

Analyses were conducted in R (version 3.4.2), using the psych (Revelle, 2016) and lavaan (Rosseel, 2012) packages. First, in the Chinese sample, we calculated descriptive statistics and interitem correlations across the 20 items, and examined the items for univariate and multivariate normality. We then used confirmatory factor analysis (CFA) to test the five-factor EPOCH structure, comparing the five-factor model (five correlated factors, each with four items) to a one-factor model (all 20 items loading on a single factor) and a higher order factor (five factors, each with four items, loading on an overarching latent factor). (See Online Appendix 2 for a visual illustration of the three models.)

Due to the univariate analyses suggesting that some items violated the univariate normality assumption, we used the Satorra–Bentler scaled test statistic (Satorra & Bentler, 1994), which rescales $\chi^2$, based on the maximum likelihood (ML) estimation to reflect kurtosis, with robust standard errors (Rosseel, 2012). We report the robust ML estimates. We evaluated model fit using the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), the Tucker–Lewis index (TLI), and the comparative fit index (CFI). Hu and Bentler (1999) suggest that adequate fit is indicated by RMSEA $\leq$ 0.06 combined with SRMR $\leq$ 0.09, and a minimum TLI and CFI of 0.90.

The well-being items are on a 5-point Likert-type scale. As is common across the well-being literature, we treat these as continuous variables, but arguably they are ordinal. As a supplemental analysis, we reran the models using robust diagonally weighted least squares (DWLS), which uses the polychoric correlation matrix to estimate model parameters (using the weighted least squares means and variance [WLSMV] estimator in lavaan) and report robust DWLS estimates.

We then tested invariance across gender (male vs. female) and age (age 9-12 and age 13-19) in the Chinese samples, and invariance across culture (China, United States, Australia) with the combined sample, using the process outlined by Hirschfeld and von Brachel (2014). Conceptually, a group variable is added to the CFA model, and parameters are fixed to equality, constraining the factor loadings (for weak invariance), intercepts (for strong invariance), and residuals (for strict invariance). Practically, the semTools package (Jorgensen, Pornprasertmanit, Miller, Schoemann, & Rosseel, 2016) allows the three levels of invariance to be simultaneously tested within the lavaan model. The more constrained models are compared with the configural model. In determining invariance, we not only focused on change in CFI, where $\Delta$CFI < 0.01 implies that the invariance assumption holds (Cheung & Rensvold, 2002), but also report $\Delta\chi^2$ for completeness.

Finally, although we primarily compare a one-factor, five-factor, and higher order factor models, alternative models are possible. To further inform understanding of the EPOCH model, we tested a bifactor model—in which items load on a general factor as well as their specific domains (see Online Appendix 2)—in the Chinese and Western samples. In the bifactor model, internal consistency is affected by both specific and general variance (Rodriguez, Reise, & Haviland, 2016a); thus, to aid the interpretation of the total and subscale scores, we calculated several additional indices, using Dueber's (2017) Microsoft Excel–based calculator.[4] Coefficient omega (ω) provides an indication of the amount of variance in the overall score that can be attributed to all sources of common variance that are included in the model (in this case, the 20 items loading on a general factor and the five specific factors), omega hierarchical (ωH) indicates the variance attributed to the single general factor (which treats variability due to the specific factors as measurement error), and omega subscale (ωS) indicates the variance attributed to specific factors (Reise, Bonifay, & Haviland, 2013; Revelle & Zinbarg, 2009; Rodriguez et al., 2016c). Comparison of ωH with ω provides an indication of how much the total variance can be attributed to the general factor, and omega hierarchical subscale (ωHS) indicates the unique variance associated with the specific factors, after partialling out the general factor. High (>.80) ωH values

**Table 1.** Summary of Fit Indices From Confirmatory Factor Analyses, Comparing One-Factor, Five-Factor, and Higher Order Models, Using Robust ML and Robust DWLS Estimation.

| | Robust ML estimation | | | Robust DWLS estimation | | |
|---|---|---|---|---|---|---|
| | One factor | Five factors | Higher order | One factor | Five factors | Higher order |
| $\chi^2$ | 4,581*** | 1,109*** | 1,628*** | 10,073*** | 2,638*** | 4,943*** |
| *df* | 170 | 160 | 165 | 170 | 160 | 165 |
| CFI | 0.814 | 0.960 | 0.938 | 0.873 | 0.968 | 0.939 |
| TLI | 0.787 | 0.952 | 0.929 | 0.858 | 0.962 | 0.930 |
| RMSEA | 0.101 | 0.048 | 0.059 | 0.127 | 0.065 | 0.089 |
| Lower CI | 0.098 | 0.045 | 0.056 | 0.125 | 0.063 | 0.087 |
| Upper CI | 0.103 | 0.051 | 0.061 | 0.129 | 0.068 | 0.092 |
| SRMR | 0.067 | 0.034 | 0.049 | 0.076 | 0.034 | 0.050 |

*Note.* $N = 3{,}620$. Estimates are scaled using the Satorra–Bentler scaled test statistic and robust standard errors. ML = maximum likelihood; DWLS = diagonally weighted least squares; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval (90%); SRMR = standardized root mean square residual.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

indicate that the majority of the explained variance is attributable to a common source (i.e., the general factor), rather than the specific factors, whereas lower ωH values combined with higher ωHS values support the multidimensional nature of the measure (Rodriguez et al., 2016a, 2016c)

## Results

### Testing the EPOCH Model

Individual item descriptives and correlations for the Chinese sample are provided in Online Appendix 3. Table 1 summarizes fit indices for the CFA models comparing a one-factor model, the expected five-factor model, and a higher order factor model (see Online Appendix 3 for item factor loadings). Using robust ML estimation, the five-factor model was superior to the one-factor and higher order factor models and fit the data well (RMSEA = 0.048, 90% confidence interval [CI] = [0.045, 0.051], SRMR = 0.053, CFI = 0.960, TLI = 0.952), replicating Kern et al. (2016). Treating items as ordinal, fit indices were almost equivalent (Table 1), with the robust DWLS model showing slightly worse fit than the ML model. Latent factor covariances were equivalent, whereas factor loadings slightly increased (see Online Appendix 4).

### Measurement Invariance

We next considered invariance for the five-factor model across gender (male vs. female) and age (preteen vs. teen). As summarized in Table 2, results suggest that strict invariance held across age and gender. Combining the Chinese sample with the U.S. and Australian samples, invariance held for configural (i.e., the same five-factor structure holds across samples) and weak (i.e., equivalence of factor loadings) invariance, but not strong (equivalence of intercepts) or strict invariance (equivalents of residuals).

### Bifactor Model

Finally, as a supplemental analysis, we tested a bifactor model. Using robust ML estimation, the model demonstrated good fit in the Chinese, $n = 3{,}620$, $\chi^2(150) = 1{,}186$, $p < .001$, CFI = 0.956,

**Table 2.** Structural Invariance, Comparing Age (Preteen and Teen), Gender (Males and Females), and Country (China, United States, Australia).

| Model | $\chi^2$ | *df* | CFI | ΔCFI | RMSEA |
|---|---|---|---|---|---|
| Gender | | | | | |
| Configural invariance | 1,771.2 | 320 | 0.960 | | 0.041 |
| Weak invariance | 1,806.4 | 335 | 0.959 | 0.001 | 0.040 |
| Strong invariance | 1,847.8 | 350 | 0.958 | 0.001 | 0.040 |
| Strict | 2,079.7 | 370 | 0.953 | 0.005 | 0.041 |
| Age | | | | | |
| Configural invariance | 1,796.8 | 320 | 0.958 | | 0.041 |
| Weak invariance | 1,886.3 | 335 | 0.955 | 0.002 | 0.041 |
| Strong invariance | 2,156.5 | 350 | 0.946 | 0.009 | 0.044 |
| Strict | 2,352.1 | 370 | 0.943 | 0.004 | 0.044 |
| Country | | | | | |
| Configural invariance | 2,983.5 | 480 | 0.960 | | 0.042 |
| Weak invariance | 3,476.0 | 510 | 0.952 | 0.008 | 0.045 |
| Strong invariance | 6,446.2 | 540 | 0.901 | 0.051 | 0.062 |
| Strict | 9,467.9 | 580 | 0.851 | 0.050 | 0.074 |

*Note.* Models test the five-factor model, estimated using robust maximum likelihood estimation. As $\chi^2$ is affected by sample size, invariance decisions were made based on the CFI, where ΔCFI < 0.01 implies the invariance assumption holds (Cheung & Rensvold, 2002; Hirschfeld & von Brachel, 2014), combined with adequate model fit (as indicated by CFI and RMSEA). CFI = comparative fit index; RMSEA = root mean square error of approximation.

TLI = 0.945, RMSEA = 0.052 [0.049, 0.054], SRMR = 0.041, and Western, $n = 2,859$, $\chi^2(150) = 900$, $p < .001$, CFI = 0.963, TLI = 0.954, RMSEA = 0.046 [0.044, 0.049], SRMR = 0.033, samples, although fit indices were slightly worse than the basic five-factor model.

Table 3 summarizes the omega indicators. In both samples, overall reliability of the multidimensional model was high (ω=.94). Comparing ωH with ω, most of the reliable variance in total scores was attributable to the general factor, as opposed to the subscale factors, with the general factor accounting for slightly more variance in the Chinese sample (ωH = .87) than in the Western sample (ωH = .84). This was further apparent with the subscale reliabilities, which generally demonstrated acceptable levels of reliability (ωS = .77-.90) on their own, but were reduced considerably after partitioning out the variability attributable to the general factor. Engagement and Perseverance captured a greater amount of unique variance (Chinese sample: $\omega SH_{Engagement} = .31$, $\omega SH_{Perseverance} = .31$; Western sample: $\omega SH_{Engagement} = .50$, $\omega SH_{Perseverance} = .40$), whereas Optimism was almost completely captured by the general factor.

Factor loadings were relatively consistent across cultures (see Table 3), although some variance occurred. Using an arbitrary difference of .15, E2, O4, and H3 had stronger loadings on the specific factors in the Western samples, whereas O3 and H4 had stronger loadings on their specific factors in the Chinese sample. E3 had stronger loadings on the general factor for the Chinese sample, whereas O3 had stronger loadings on the general factor for the Western sample. These differences might point to cultural differences in how reliably specific items reflect the intended construct.

## Discussion

The past few decades have brought growing interest globally in assessing human well-being (Organisation for Economic Co-operation and Development [OECD], 2013). Well-being measures can be used to (a) assess change over time, (b) review and evaluate policy decisions, (c)

**Table 3.** Supplemental Analysis Testing a Bifactor Model (All Items Loading on a General Factor and Their Respective Factors in the Chinese and Western [United States and Australia Combined] Samples), Using Robust Maximum Likelihood Estimation.

| | Chinese sample | | | | | | Western samples | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | E | P | O | C | H | Gen | E | P | O | C | H | Gen |
| E1 | .47 | | | | | .48 | .56 | | | | | .44 |
| E2 | .25 | | | | | .58 | .53 | | | | | .47 |
| E3 | .58 | | | | | .48 | .60 | | | | | .32 |
| E4 | .40 | | | | | .54 | .49 | | | | | .40 |
| P1 | | .51 | | | | .58 | | .54 | | | | .48 |
| P2 | | .41 | | | | .50 | | .52 | | | | .44 |
| P3 | | .45 | | | | .56 | | .49 | | | | .55 |
| P4 | | .37 | | | | .55 | | .45 | | | | .52 |
| O1 | | | .12 | | | .73 | | | .12 | | | .69 |
| O2 | | | .30 | | | .47 | | | .30 | | | .59 |
| O3 | | | .52 | | | .56 | | | .22 | | | .75 |
| O4 | | | .07 | | | .69 | | | .28 | | | .70 |
| C1 | | | | .30 | | .51 | | | | .44 | | .58 |
| C2 | | | | .35 | | .62 | | | | .49 | | .58 |
| C3 | | | | .33 | | .59 | | | | .43 | | .58 |
| C4 | | | | .44 | | .55 | | | | .34 | | .50 |
| H1 | | | | | .52 | .71 | | | | | .51 | .70 |
| H2 | | | | | .43 | .72 | | | | | .35 | .66 |
| H3 | | | | | .14 | .75 | | | | | .31 | .73 |
| H4 | | | | | .50 | .72 | | | | | .28 | .68 |
| ω | | | | | | .941 | | | | | | .938 |
| ωS | .772 | .791 | .768 | .766 | .903 | | .774 | .799 | .814 | .798 | .865 | |
| ωH | | | | | | .870 | | | | | | .843 |
| ωHS | .308 | .306 | .111 | .216 | .207 | | .495 | .401 | .084 | .290 | .183 | |

*Note*. Numbers indicate the latent factor loadings. Chinese sample $n$ = 3,620; Western (combined United States and Australian) samples $n$ = 2,859. E = Engagement, P = Perseverance, O = Optimism, C = Connectedness, H = Happiness, Gen = general factor, ω = omega, ωS = omega subscale, ωH = omega hierarchical, ωHS = omega hierarchical subscale. See Online Appendix 2 for item labels.

enable international comparisons, (d) assess subgroup differences, (e) identify future areas of need or opportunity, (f) evaluate the impact of policy proposals, (g) shape the content and delivery of policy, and (h) inform targeting of new policies according to population subgroups (Michaelson, Abdallah, Steuer, Thompson, & Marks, 2009). Numerous theories and corresponding measures have been developed and used to compare individual and societal well-being. However, for comparisons to be made, establishing the extent to which measurement invariance holds across cultures is critical (Davidov et al., 2014). Although such validation work has occurred with several adult well-being measures, consideration of measures specifically designed for adolescents is needed (Rose et al., 2017). The current study tested a Chinese version of the EPOCH Measure of Adolescent Well-Being (Kern et al., 2016), and compared responses with American and Australian respondents. Analyses supported the structure of the measure but suggest that mean values may not be directly comparable.

The EPOCH model is a theoretical model that arose from Seligman's (2011) multidimensional PERMA theory. Seligman proposed that psychological flourishing is comprised of five independent outcomes, which many people value and pursue: Positive emotions, Engagement, positive Relationships with others, a sense of Meaning and purpose, and Accomplishment.

When applying the model to youth, some of these domains are potentially affected more by developmental factors than others. For example, older youth often have more opportunities for accomplishment than young teenagers. Meaning in life looks different for a 12-year-old versus an 18-year-old. Thus, the authors chose characteristics that theoretically relate to the PERMA factors, and which longitudinal studies find are predictive of the PERMA factors, but are not dependent on developmental stage.

Based on this theoretical framework, the original English measure was developed and tested with thousands of adolescent students in Australia and the United States, providing initial evidence that items were internally reliable, responses were consistent over time, and the scores represent the intended constructs (Kern et al., 2016). In the current study, a translated version of the measure was tested with more than 3,600 students, aged 9 through 18 years, who represented a broad range of demographic characteristics. The five-factor model demonstrated good fit and was invariant across age and gender. Invariance of the factor structure held across cultures, which suggests that associations with other variables can be tested and compared across cultures. Internal and cross-time estimates of reliability, convergent and discriminant validity, and predictive ability of the factors might be examined to inform the practical usefulness of the measure and the boundaries of what inferences can be made when using the measure.

However, variance in the intercepts indicates that mean scores are not directly comparable across cultures. Indeed, in a simulation study, Steinmetz (2013) found that even one variant intercept can substantially affect composite mean differences. Furthermore, Steinmetz found that as sample size increases, it becomes increasingly likely that means will significantly differ. Applied to the EPOCH measure, if comparisons are made across cultures based on the average of the four manifest items, differences are likely to be found, whether or not differences actually exist. This implies that if composite scores are used (i.e., taking the average of the four items for each factor), then culture-specific normed values need to be established. To compare means across culture, latent variable models that directly test the factor structure and invariance, and correct for measurement error, should be used (Steinmetz, 2013).

Nonequivalent intercepts can be caused by cross-cultural differences in acquiescence response style, social desirability biases, reference group effects, or various other reasons. The samples also were not nationally representative (especially the Australian sample, which involved boys from an elite private school in Australia). Future studies might examine the extent to which these differences are a function of the measure, a function of the samples included, or actual cultural differences.

Across the psychological and well-being literature, self-reported Likert-type items are typically treated as continuous variables, despite their ordinal nature. Using robust ML and robust DWLS, we found little difference in the ordinal versus continuous models, although the ordinal models had better model fit and stronger factor loadings. Simulation studies may be beneficial for unpacking the possible implications of assuming different variable structures in the analysis.

Replicating Kern et al. (2016), our main analyses compared three structural models: a one-factor model, a five-factor model, and a higher order factor model. As a supplemental analysis, we tested a bifactor model. Although the five-factor model provided slightly better fit in both the Chinese and Western samples, the bifactor model also adequately fit the data, raising several theoretical and practical issues. In the estimated bifactor model, most of the common variance could be explained by the general factor. As the items aim to capture positive characteristics, a general factor might reflect a general positivity factor. Indeed, the Optimism and Happiness items primarily loaded on the general factor, whereas the other factors had more unique variance. It might also reflect a halo effect, method bias, or another underlying factor. The high amount of variance captured by the general factor suggests that what appears to be reliable subscales (based on $\alpha$ or $\omega$ alone) is mostly due to individual differences on the general factor, and interpreting the specific factors as unique from the general factor may be misguided (Rodriguez et al., 2016a). Notably, this was consistent across cultures.

This finding needs to be replicated in other samples but raises questions about how overall well-being is measured and tested, as well as the added value of different subscales. For instance, Goodman et al. (2018) compared Seligman's (2011) PERMA model with Diener's (1984) subjective well-being model, suggesting that statistically there was no added benefit to including additional domains beyond positive emotion and life satisfaction. Still, others have argued that from a practical standpoint, it is easier to intervene on specific domains of well-being than on an overall sense of well-being (e.g., Kern, Waters, Adler, & White, 2015; McQuaid & Kern, 2017).

As the unique variance for each subscale and individual item loadings varied by domain and culture, it is possible that a different underlying factor structure may be more appropriate—in general or across cultures—after the general factor is accounted for. Well-being is an abstract construct, and its definition and structure are subject to considerable debate across multiple disciplines. Although various theoretical models have been proposed, the extent to which well-being is a uni- or multidimensional construct is subject to opinion and debate (see, for instance, Forgeard et al., 2011; OECD, 2013). Testing different subconstructs within the bifactor framework might inform which primarily reflect a single construct (such as Optimism and Happiness in this study), versus which reflect more distal constructs (such as Engagement and Perseverance).

The practical implications of these results are unclear. The modeling decisions made clearly can affect the interpretation of the results. Across cultures, the five-factor model demonstrated superior fit to the other models, suggesting that the measure can be used across Chinese and Western cultures to capture Engagement, Perseverance, Optimism, Connectedness, and Happiness. But the unique added value of these subscales—or any other positive characteristic subscales—is unclear. Future studies might further examine the presence and implications of the bifactor model for well-being measures, both within and across cultures, as well as consider the practical implications of different modeling decisions.

### *Limitations and Additional Considerations*

Several limitations and issues need to be acknowledged. First, the computer software used to collect data in the current study only recorded complete responses. It is unknown how many respondents began the questionnaire but did not finish. Thus, the analyses essentially use a casewise deletion approach, despite the many known drawbacks, with no way of estimating the extent to which missingness affected the results. In contrast, for the U.S. and Australian samples, both full and partial responses were recorded. Missing data were imputed using ML estimates. The differences in data collection methods, in which the amount of missingness can be tested and imputed (for the U.S. and Australian samples) versus unknown missingness (in the Chinese sample), may have produced additional differences among the three samples.

Second, although the Chinese schools represented a range of backgrounds, demographic information beyond age and gender was unavailable. Thus, the extent to which the factor structure and invariance tests hold across other Chinese samples, as well as other Eastern and Western cultures, needs to be determined. Third, responses across the three samples may be nested within class or school. Age and school effects may be confounded with responses. Future studies might further investigate the impact of nesting on responses to well-being data.

## Conclusion

Most parents, when asked, indicate that they want their children to be happy (Seligman, Ernst, Gillham, Reivich, & Linkins, 2009) but the high prevalence of mental disorder globally suggests that there is considerable room for improvement. To move beyond treating psychopathology in a reactive manner, supporting optimal functioning in young people is vital. Measurement plays an important role in understanding and building well-being. The well-being of young people is often either approximated based on performance on standardized academic tests such as the Programme for International Student Assessment (PISA) or determined based on measures of depression, anxiety, and other psychological disorders. Measures that directly estimate the presence of well-being have mostly been developed for adults or have been tested on limited samples (Rose et al., 2017). Measures that adequately capture young people's perspectives on their own well-being, which are comparable across different cultures, are needed.

The current study provides a step in this direction. The EPOCH measure was developed specifically for young people, with a variety of participants from the United States and Australia, finding support for the psychometric properties of the measure. The measure has subsequently been tested in a growing number of samples, including Turkish students (Demircia & Ekşib, 2015). The current study adds a Chinese version of the measure, finding that it is an acceptable tool for Chinese respondents, and providing additional support for the scale's structure. However, it also raises questions around how different modeling decisions might affect the interpretation and use of well-being measures. The study adds to the knowledge base tools that can be used to investigate similarities and differences in positive psychological variables across various cultural backgrounds.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Notes

1. The data collection system required that participants answer every item, and only recorded complete responses. It is possible that some students did not finish the questionnaire, but we were unable to test the extent to which missing data may have occurred.
2. Five questions assessing physical vitality, four items assessing school engagement, and a single item on academic performance came from the Healthy Pathways Child-Report Scales (Bevans, Riley, & Forrest, 2010). Ten items assessing symptoms of anxiety/worry and depression/sadness came from the patient-reported outcome measurement information system (PROMIS) pediatric eight-item short forms (Varni et al., 2014). Overall happiness was measured with a single item ("Taking all things together, how happy would you say you are?").
3. We include Samples 1, 2, and 7 to 10. Samples 3 to 6 were used to help develop the measure, completing a 25-item version of the measure, with some items that differed from the final 20-item version, and, thus, are excluded.
4. The calculator allows the user to input the unidimensional and general factor loadings, and then calculates relevant indices based on formulas provided by Rodriguez, Reise, and Haviland (2016c) and noted in the corrected version of Rodriguez et al. (2016a; see Rodriguez et al., 2016b, for corrected formulas). Additional statistical indices could be calculated, informing factor weightings and how the items should subsequently be specified in a structural equation model (cf. Rodriguez et al., 2016a, 2016c). As a complete consideration of the bifactor model was beyond our scope, we consider a limited number of metrics here to aid the interpretation of the total and subscale scores.

## ORCID iDs

Margaret L. Kern https://orcid.org/0000-0003-4300-598X
Hanchao Hou https://orcid.org/0000-0002-5019-3010

## Supplemental material

Supplemental material for this article is available online.

## References

Bevans, K. B., Riley, A. W., & Forrest, C. B. (2010). Development of the healthy pathways child report scales. *Quality of Life Research*, *19*, 1195-1214.

Brislin, R. W. (1980). Translation and content analysis of oral and written materials. In H. C. Triandis & W. W. Lambert (Eds.), *Handbook of cross-cultural psychology* (Vol. 2, pp. 349-444). Boston, MA: Allyn & Bacon.

Butler, J., & Kern, M. L. (2016). The PERMA-Profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing*, *6*, 1-48.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

Copeland, E. P., Nelson, R. B., & Trauberger, M. C. (2010). Wellness dimensions relate to happiness in children and adolescents. *Advances in School Mental Health Promotion*, *3*, 25-37.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*, 55-75.

Demircia, İ., & Ekşib, F. (2015). Five-dimensional model of adolescent well-being: Validity and reliability of the Turkish version of the EPOCH measure. *Journal of Youth Research*, *3*, 21-30.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*, 542-575.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*, 143-156.

Donaldson, S. I., Dollwet, M., & Rao, M. A. (2015). Happiness, excellence, and optimal human functioning revisited: Examining the peer-reviewed literature linked to positive psychology. *The Journal of Positive Psychology*, *10*, 185-195.

Dueber, D. M. (2017). *Bifactor indices calculator: A Microsoft excel-based tool to calculate various indices relevant to bifactor CFA models*. doi:10.13023/edp.tool.01

Forgeard, M. J. C., Jayawickreme, E., Kern, M. L., & Seligman, M. E. P. (2011). Doing the right thing: Measuring well-being for public policy. *International Journal of Wellbeing*, *1*, 79-106.

Froh, J. J. (2004). The history of positive psychology: Truth be told. *The Psychologist*, *16*, 18-20.

Furlong, M. J., You, S., Renshaw, T. L., Smith, D. C., & O'Malley, M. D. (2014). Preliminary development and validation of the Social and Emotional Health Survey for secondary school students. *Social Indicators Research*, *117*, 1011-1032.

Goodman, F. R., Disabato, D. J., Kashdan, T. B., & Barry Kaufman, S. (2018). Measuring well-being: A comparison of subjective well-being and PERMA. *The Journal of Positive Psychology*, *13*(4), 321-332.

Hirschfeld, G., & von Brachel, R. (2014). Improving Multi-group confirmatory factor analysis in R: A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research & Evaluation*, *19*(7). Retrieved from http://pareonline.net/getvn.asp?v=19&n=7

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Huppert, F. A., & So, T. T. C. (2013). Flourishing across Europe: Application of a new conceptual framework for defining well-being. *Social Indicators Research*, *110*, 837-861.

Jorgensen, T. D., Pornprasertmanit, S., Miller, P., Schoemann, A., & Rosseel, Y. (2016). *Useful tools for structural equation modeling: Package "semTools"* (version 0.4-14). Retrieved from https://CRAN.R-project.org/package=semTools

Kern, M. L., Benson, L., Steinberg, E. A., & Steinberg, L. (2016). The EPOCH measure of adolescent well-being. *Psychological Assessment*, *28*, 586-597.

Kern, M. L., Park, N., Peterson, C., & Romer, D. (2017). The positive perspective on youth development. In D. Romer (Ed.), *Treating and preventing adolescent mental disorders: What we know and what we don't know* (Vol. 2, pp. 543-567). New York, NY: Oxford University Press.

Kern, M. L., Waters, L., Adler, A., & White, M. (2015). A multifaceted approach to measuring wellbeing in students: Application of the PERMA framework. *The Journal of Positive Psychology*, *10*, 262-271.

Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior*, *43*, 207-222.

Lerner, R. M., Fisher, C. B., & Weinberg, R. A. (2000). Toward a science for and of the people: Promoting civil society through the application of developmental science. *Child Development*, *71*, 11-20.

McQuaid, M., & Kern, M. L. (2017). *Your wellbeing blueprint: Feeling good and doing well at work*. Melbourne, Victoria, Australia: McQuaid Ltd.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525-543.

Michaelson, J., Abdallah, S., Steuer, N., Thompson, S., & Marks, N. (2009). *National accounts of well-being: Bringing real wealth onto the balance sheet*. London, England: New Economics Foundation.

Organisation for Economic Co-Operation and Development. (2013). *OECD guidelines on measuring subjective well-being*. OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264191655-en

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*, 129-140.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*, 552-56652.

Revelle, W. (2016). psych: procedures for psychological, psychometric, and personality research. Retrived from https://cran.r-project.org/web/packages/psych/index.html

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*, 145-154.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*, 223-237.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Correction to: Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*, 444.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016c). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*, 137-150.

Rose, T., Joe, S., Williams, A., Harris, R., Betz, G., & Stewart-Brown, S. (2017). Measuring mental well-being among adolescents: A systematic review of instruments. *Journal of Child and Family Studies*, *26*, 2349-2362.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1-36.

Roth, J., & Brooks-Gunn, J. (2003). Youth development programs: Risk, prevention, and policy. *Journal of Adolescent Health*, *32*, 170-182.

Rusk, R. D., & Waters, L. E. (2013). Tracing the size, reach, impact, and breadth of positive psychology. *The Journal of Positive Psychology*, *8*, 207-221.

Rusk, R. D., & Waters, L. E. (2015). A psycho-social system approach to well-being: Empirically deriving the five domains of positive functioning. *The Journal of Positive Psychology*, *10*, 141-152.

Ryff, C. D. (1995). Psychological well-being in adult life. *Current Directions in Psychological Science*, *4*, 99-104.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Scales, P. C., Benson, P. L., Leffert, N., & Blyth, D. A. (2000). Contributions of developmental assets to the prediction of thriving among adolescents. *Applied Developmental Science*, *4*, 27-46.

Seligman, M. E. P. (2011). *Flourish: A visionary new understand of happiness and well-being*. New York: Free Press.

Seligman, M. E. P., Ernst, R. M., Gillham, J., Reivich, K., & Linkins, M. (2009). Positive education: Positive psychology and classroom interventions. *Oxford Review of Education*, *35*, 293-311.

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *9*, 1-12.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Varni, J. W., Magnus, B., Stucky, B. D., Liu, Y., Quinn, H., Thissen, D., . . . DeWalt, D. A. (2014). Psychometric properties of the PROMIS (R) Pediatric Scales: Precision, stability, and comparison of different scoring and administration options. *Quality of Life Research*, *23*, 1233-1243.

Verma, S. K., Dubey, B. L., & Gupta, D. (1983). P.G.I general wellbeing scale. *Indian Journal of Clinical Psychology*, *10*, 299-304.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychiatric Association.