



Theses and Dissertations

2022-04-18

Psychometric Properties of the Survey on Flourishing (SURF) in a Nationally Representative Adolescent Sample

Gus C. Salazar
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Family, Life Course, and Society Commons](#)

BYU ScholarsArchive Citation

Salazar, Gus C., "Psychometric Properties of the Survey on Flourishing (SURF) in a Nationally Representative Adolescent Sample" (2022). *Theses and Dissertations*. 9477.
<https://scholarsarchive.byu.edu/etd/9477>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Psychometric Properties of the Survey on Flourishing (SURF)
in a Nationally Representative Adolescent Sample

Gus C. Salazar

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Jared Warren, Chair
Gary Burlingame
Sam Hardy

Department of Psychology
Brigham Young University

Copyright © 2022 Gus C. Salazar

All Rights Reserved

ABSTRACT

Psychometric Properties of the Survey on Flourishing (SURF) in a Nationally Representative Adolescent Sample

Gus C. Salazar
Department of Psychology, BYU
Master of Science

Background: Adolescent subjective well-being is a topic that has gained significant focus over recent years. This focus is reflected in the formation of measurement tools and interventions used to better understand and improve adolescent mental health. While these are important steps, there still exist problems related to adolescent subjective well-being measurement. Notably, current measurement tools are limited in their content, applicability to various populations, and in their accessibility. **Aims:** The purpose of this paper is to examine the psychometric properties of the Survey on Flourishing (SURF) when used with a nationally representative adolescent sample. The SURF aims to address some of the issues with current adolescent subjective well-being measurement tools. **Method:** A nationally representative sample of 334 participants participated in the present study. We examined the reliability and validity of the SURF by examining its internal consistency, convergent validity, and discriminant validity. We also examined the factor structure of the SURF using a confirmatory factor analysis (CFA). **Results:** The SURF demonstrated high internal consistency ($\alpha = .92$), strong positive correlation with convergent measures, and a weak negative correlation with a discriminant measure. A one-factor model best fits the observed data. **Conclusion:** The SURF demonstrated good psychometric properties and addresses several of the problems that exist in current measures. The SURF is a useful and effective measure of adolescent subjective well-being.

Keywords: adolescent subjective well-being, well-being, youth, flourishing, measurement

ACKNOWLEDGEMENTS

During the past two years I have put many hours into this project, but I could not have done it without the support of many. First, I'd like to thank my mentor Jared Warren for contributing his expertise in human development and flourishing and also giving me encouragement and guidance, without which this project would have been much more difficult. I'd also like to thank my thesis committee members Gary Burlingame and Sam Hardy for challenging me and providing me with thoughtful and critical feedback. I'd also like to thank Scott Braithwaite and Joe Olsen for sharing with me their knowledge and expertise in statistical analysis. Most of all, I'd like to thank my wife Emmy standing by my side and supporting me as I put countless hours into this manuscript, especially when it took time spent away from home.

TABLE OF CONTENTS

INTRODUCTION	i
ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES.....	vi
LIST OF FIGURES	vii
Psychometric Properties of the Survey on Flourishing (SURF).....	1
in a Nationally Representative Adolescent Sample.....	1
What is Subjective Well-Being?.....	2
Models of Subjective Well-being	3
Measures of Adolescent Well-Being	4
Limitations to Current Measures	7
The Survey on Flourishing.....	10
Method.....	11
Participants.....	11
Procedures.....	12
Measures	13
Survey on Flourishing (SURF).	13
Positive and Negative Emotion Schedule, Short Form.....	15
Satisfaction with Life Scale	15
Data Analyses	16
Results.....	18

Data Preparation.....	18
Internal Consistency.....	19
Factor Structure.....	19
Model 1	20
Model 2	23
Model 3	25
Model 4	27
Convergent and Discriminant Validity	30
Discussion.....	31
SURF Psychometric Properties and Structure	31
SURF Performance in Adults Versus Adolescents.....	33
Other Contributions to Adolescent Well-Being Measurement.....	33
Conclusion	35
Limitations	35
Future Directions	36
References.....	38

LIST OF TABLES

Table 1: *SURF Items* 14

Table 2: *Demographic Summary and Normative Data From Each Sample* 19

Table 3: *Surf Items, Factor Descriptors, and Loadings for the Modified One-Factor Model* 27

Table 4: *Fit Statistics for the Models Tested* 30

Table 5: *Cronbach's Alpha and Correlations between SURF and Other Measures*..... 31

LIST OF FIGURES

Figure 1: <i>Diagram of Model 1 Structure and Factor Loadings</i>	22
Figure 2: <i>Diagram of Model 2 Structure and Factor Loadings</i>	24
Figure 3: <i>Diagram of Model 3 Structure and Factor Loadings</i>	26
Figure 4: <i>Diagram of Model 4 Structure and Factor Loadings</i>	29

Psychometric Properties of the Survey on Flourishing (SURF) in a Nationally Representative Adolescent Sample

Flourishing is a state of well-being that occurs when individuals have the psychological, social, and physical resources they need to meet psychological, social, or physical challenges (Dodge et al., 2012). It goes beyond just focusing on the absence of pathology, also examining positive outcomes such as emotional regulation, supportive relationships, meaning and purpose, and life satisfaction. Currently, much of the research on flourishing has been conducted using adult populations rather than youth (Moore & Keyes, 2003). However, in recent years there has been a push to better understand adolescent well-being.

Well-being research shows that flourishing is associated with several positive life outcomes and circumstances, including supportive social networks and relationships, positive work life, higher levels of physical and mental health, and improved school performance (Crum et al., 2013; Diener & Biswas-Diener, 2011; Diener & Ryan, 2009; Howell, 2009; Ong, 2010; Pérez-Garín et al., 2015; Weich et al., 2011). Given the prevalence of psychosocial stressors present during adolescence and the nature of adolescence as a critical period for social and emotional growth, it becomes more important to accurately understand and measure adolescent flourishing.

Improving adolescent flourishing also has broader societal implications. Historically, developmental science, psychology, education, and other fields have underestimated adolescents, tending to focus on the problems they face (e.g., learning difficulties, mental illness, low motivation, substance use, etc.) rather than the strengths they possess (Damon, 2004; Lerner et al., 2009). However, positive youth development research and other similar areas of research identify adolescents as having unique resources which they can use to meaningfully contribute to

their community (Damon, 2004; Shek et al., 2019). Although working to improve adolescent well-being and enabling them to use their strengths to contribute meaningfully to society is an important task, it may be difficult to accomplish if we cannot measure it. Thus, the purpose of the present study is to examine the psychometric properties of the Survey on Flourishing (SURF), a measure of subjective well-being, using a nationally representative adolescent sample.

What is Subjective Well-Being?

Subjective well-being is a broad, multifaceted construct, and has historically been difficult to define. For many decades well-being was determined to be the absence of physical or mental malfunction. However, more recent research indicates that well-being is not just the absence of problems, but it includes assets, strengths, values, and other positive characteristics (Magyar & Keyes, 2019; Keyes, 1998). Diener, in a classic paper on subjective well-being, defined subjective well-being as a combination of positive emotion and life satisfaction (1984). Currently, definitions of subjective well-being most commonly include two components: *emotional well-being*, which includes the presence of positive emotion and life satisfaction, and *positive functioning*, which includes social functioning (e.g., social integration and contribution) and psychological functioning (e.g., autonomy and personal growth; Magyar & Keyes, 2019).

These components of well-being also apply to adolescents. Researchers have identified certain developmental tasks which may indicate whether a child is doing well. Some of these tasks which are critical in adolescence include academic achievement, forming close peer relationships, learning to follow rules, participating in extracurricular activities, and forming a sense of self-identity (Masten & Coatsworth, 1998). These tasks which are critical to healthy adolescent development generally align with the social, emotional, and psychological components included in subjective well-being.

Taken together, current research suggests that subjective well-being is more than just the absence of pathology, it is subjectively experienced, and it includes emotional, social, and psychological well-being components (Ryan & Deci, 2001). This definition also applies across developmental periods, although how they might manifest may differ. This definition of subjective well-being has been used to create various models of well-being which we discuss below.

Models of Subjective Well-being

One recent model that has been suggested is Martin Seligman's five-factor PERMA model. The PERMA model conceptualizes subjective well-being through the domains of positive emotion, engagement, supportive relationships, meaning, and achievement (Butler & Kern, 2016; Seligman, 2011). This model uses these five domains to capture emotional well-being (via the positive emotion domain), social functioning (via the supportive relationships domain), and psychological functioning (via the engagement, meaning, and achievement domains). Although this model has mainly been applied to adults, the EPOCH (engagement, perseverance, optimism, connectedness, happiness) is a model which adapted the PERMA to better apply to adolescents.

A second commonly used model is the Ryff model of psychological well-being (Ryff & Keyes, 1995). This model focuses specifically on the psychological well-being component of subjective well-being. It describes psychological well-being as encompassing six dimensions: autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance.

Third, Keyes' model of social well-being is a frequently used model for describing social well-being (1998). According to this model, social well-being comprises five domains: social integration, social contribution, social coherence, social actualization, and social acceptance.

Both the Ryff model and the Keyes model clearly describe components of subjective well-being, and they are brought together in the Mental Health Continuum (MHC), which is a measure of subjective well-being which we discuss in the following section (Keyes, 2006).

Measures of Adolescent Well-Being

Using the conceptualization of subjective well-being and the models described above, researchers have developed measures to assess subjective well-being. However, these measures have mainly relied on adult populations to determine their utility. However, there are still some measures which have either been developed specifically for adolescents, or which have been shown to be effective when used with adolescents. In this section we describe the most commonly used measures of adolescent well-being.

A recent literature review identified seven measures of subjective well-being which are available for use with adolescents, and which contain items which measure both the *emotional well-being* and *positive functioning* components of subjective well-being (Rose et al., 2017). These measures included the MHC-short form, the Ryff scales of psychological well-being, the EPOCH (engagement, perseverance, optimism, connectedness, happiness) measure of adolescent subjective well-being, the Child and Adolescent Wellness Scale (CAWS), the Social and Emotional Health Survey (SEHS), the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS), and the World Health Organization-Five well-being index (WHO-5; Allgaier et al., 2012; Copeland et al., 2010; Fernandes et al., 2010; Furlong et al., 2014; Kern et al., 2016; Keyes, 2006).

The Ryff scales of psychological well-being constitute a 20-item measure which is based upon the Ryff model of psychological well-being discussed above. While Rose and colleagues (2017) reported that this scale includes both *emotional well-being* and *psychological functioning*

components, the creators of this study only aimed to capture the *psychological well-being* aspect of subjective well-being (Ryff, 1989). It measures psychological well-being along the six domains of autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance. Items from this measure were adapted to create the MHC, which is a more comprehensive measure of subjective-well-being.

The Mental Health Continuum—Long Form is a 35-item measure which draws from the Ryff and Keyes models discussed above (Keyes, 2005; Keyes, 2006). The MHC-short form is an adapted version of this measure which includes 14 items— three which assess emotional well-being, five which assess social well-being, and six which assess psychological well-being. In a recent study summarizing research on the latent profile analysis of the MHC, a bi-factor model (like that found for the PERMA Profiler) appeared to demonstrate the best fit with observed data, although the general subjective well-being factor accounted for a substantially greater amount of variance than either of the other latent variables (Reinhardt et al., 2020). A three-factor model (with factors representing MHCM’s three main foundations) also demonstrated good fit, although slightly less than the bi-factor model (Reinhardt et al., 2020).

The EPOCH is a 20-item self-report measure which was developed for adolescents and adapted from the PERMA model described above (Seligman, 2011). The EPOCH measures engagement, perseverance, optimism, connectedness, and happiness. The researchers aimed to develop a measure using domains which influence the PERMA domains in adulthood. They determined that the EPOCH demonstrated adequate psychometric properties, although more research is needed to determine the extent of its utility.

The EPOCH’s counterpart is the PERMA-profiler, which is a 23-item measure designed for adults that is also based on Seligman’s PERMA model (Butler & Kern et al., 2016). A recent

study examining this measure's effectiveness suggested that this model was an accurate and valid well-being measurement tool. They also used confirmatory factor analysis (CFA) to determine that a bi-factor model of subjective well-being best fit the data. The model contained one general factor which accounted for much of the shared variance between all the items, and five secondary factors that represented the PERMA domains (Bartholomaeus et al., 2020). They determined that the general factor explained most of the variance in scores, while the secondary factors explained additional (although relatively weak) amounts of variance.

The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) is also a widely used subjective well-being measurement tool (Clarke et al., 2011). Items in this measure were developed to reflect the domains of positive affect, psychological functioning (autonomy, competence, self-acceptance, personal growth) and interpersonal relationships (Tennant et al., 2007) Similar to the MHCM and the PERMA Profiler, research examining the WEMWBS' performance with adolescents suggested that a bi-factor model best fit the observed data, with a broad factor representing general subjective well-being, and two relatively weak secondary factors representing psychological and social well-being (Shannon et al., 2020).

The CAWS is a 100-item measure that assesses child well-being through ten domains: empathy, connectedness, self-efficacy, initiative, adaptability, social competence, conscientiousness, optimism, emotional regulation, and mindfulness (Copeland et al., 2010). This scale measures domains which are not generally included in adult measures of well-being, including social competence and emotional regulation. While the original validation study noted that this measure demonstrated good reliability, they did not examine the factor structure of the measure.

The SEHS is a 36-item measure designed to assess subjective well-being along the domains of belief in self, belief in others, emotional competence, and engaged living (Furlong et al., 2014). This model is based largely on social-emotional learning theory, which suggests that developing social and emotional competencies are among the most important tasks that will help adolescents live meaningful lives (Buckley et al., 2003). This model also drew from positive psychology research on the importance of supportive relationships on promoting well-being, although it was not explicitly grounded in the models of subjective well-being described above.

In summary, researchers have developed many models of subjective well-being for adolescents. While these measures differ somewhat in their underlying theories and content, researchers generally agree that subjective well-being is a subjectively experienced, multifaceted construct. There is strong overlap among these measures which suggest subjective well-being consists of social (e.g., connection, supportive relationships, etc.), psychological (purpose, achievement, etc.), and emotional (e.g., life satisfaction, positive emotion, etc.) components. Statistical approaches used to examine some of the most commonly referenced models suggest that using a bi-factor approach may be an effective method for measuring subjective well-being. Specifically, after accounting for general subjective well-being or positive emotion, other components may explain additional (although relatively small) differences in people's levels of subjective well-being. It is important to note, however, that the literature did not provide a consensus on how many secondary factors exist or what exactly they represent. In the studies above, the researchers specified the factors based on their theoretical model (e.g., the PERMA-profiler used five secondary factors while the MHC identified three). However, these factors generally seemed to represent aspects of social, psychological, and emotional well-being.

Limitations to Current Measures

Although the development of measures of adolescent subjective well-being represents significant progress, there are some notable limitations that impact their utility. First, measure content must be considered. There is a consensus among well-being researchers that subjective well-being is a broad, multifaceted construct. While three generally agreed on domains include social, emotional, and psychological well-being, there are many important domains which have not yet been tapped into, such as gratitude, transcendence, or mindfulness. Other measures of subjective well-being focus on a particular facet of well-being. Although this approach may be intentional, these measures may be too narrow to capture certain elements important to broader adolescent subjective well-being. Similarly, Seligman (2018) suggests that no single measure can capture the breadth and depth of well-being. Thus, although these measures may provide valuable information, they may be most useful when used in conjunction with measures that examine alternate facets of well-being. The depth and breadth of subjective well-being suggests that there is a need for additional measures that expand on the content of current measures.

Second, the generalizability of these measures depends on the sample which was used to examine their psychometric properties. When examining these measure's validation studies, we determined that only the EPOCH obtained a nationally representative sample from within the United States. Additionally, most of these validation studies relied on samples obtained from outside of the United States. Although this should be noted as a strength for those using these measures with clients from the areas where the measure was validated, it would be inappropriate to expect these measures to perform equally across cultural groups. Thus, because the purpose of this study is to examine the psychometric properties of the SURF with data from a representative sample from the United States, these validation studies suggest there may be a need for additional measures which are supported for use within this population.

One additional concern regarding the samples used in these validation studies relates to the size of the sample. Although a general rule of thumb regarding sample size states that researchers should collect ten responses per test item, several adolescent subjective well-being measures fall short of this recommendation (Clark & Watson, 2016; Morgado et al., 2017). When developing a measure, having a too-small sample size may increase measurement error and lead to inaccurate or biased measurement.

Finally, there are other relatively smaller limitations that may affect these measure's utility. First, some measures, such as the 100-150 item CAWS, are extensively lengthy (Copeland et al., 2010). Shorter measures may be more time-sensitive while still demonstrating good psychometric properties with little item overlap. Second, accessibility determines the extent to which the measures can be used for many practical purposes. Out of these measures, the PGI is not available in the public domain, and the CAWS, SEHS, WEMWBS, and SEHS are free to use with developer permission. The EPOCH, MHC-SF, Ryff scales, and the WHO-5 are free with developer acknowledgement (Rose et al., 2017). Although paid measures may be effective, free measures, such as the EPOCH and the SURF may display similar effectiveness and allow for more widespread use.

In summary, although the development of measures of adolescent subjective well-being is a step forward, these measures have some limitations such as content domain, sample population, sample size, accessibility, and length which impact their utility as adolescent subjective-well-being outcome measures. Because adolescent well-being is becoming a greater priority in society, it is important that accurate measurement tools are available to help individuals understand and improve it. In this paper, we examine the psychometric properties of the Survey on Flourishing adolescent version (SURF) as a novel measure of subjective well-

being that is based on a nationally representative US sample, accessible, and quick to administer. We also examine its reliability, validity, and factor structure.

The Survey on Flourishing

This study aims to use current research on adolescent well-being to examine the psychometric properties of the Survey on Flourishing (SURF) in a nationally representative adolescent sample within the United States. The original SURF questionnaire was designed to obtain a measurement of subjective well-being by including items reflecting both positive functioning and emotional well-being. The SURF was shown to have good reliability and validity when used in an adult population (Linford, Salazar, et al., 2021). We expect that the SURF will demonstrate similar psychometric properties and structure when used in an adolescent population. Thus, this study aimed to examine the utility of the SURF through examining its internal consistency, factor structure, and convergent and discriminant validity.

Regarding the reliability of the SURF, we expected that the SURF would demonstrate good internal consistency by having a Cronbach's alpha score (average inter-item correlation) between .80 and .90. Having a Cronbach's alpha statistic in this range means that the test displays strong internal consistency, which is one facet of reliability.

Regarding the factor structure of the SURF, we expected the items to load onto a single general factor of adolescent subjective well-being (see Methods). This suggests the SURF measures a unitary construct, which would align with previous research on the measure (Linford, Salazar, et al., 2022).

Regarding the SURF's validity, we expected the SURF to show a strong positive correlation ($r > .70$) with similar measures of well-being such as the PANAS Positive Affect subscale (PANAS-Pos) and the Satisfaction With Life Scale (SWLS), while showing a weak

negative correlation ($r < -0.5$) with discriminant measures such as the PANAS Negative Affect Subscale (PANAS-Neg). These predictions were based on a previous study which showed that the SURF demonstrated similar psychometric properties when used with adults and adolescents (Linford, Salazar et al., 2022). Good convergence with the PANAS Positive Affect subscale and the SWLS would suggest that the SURF is measuring a similar construct, whereas a low correlation with discriminant measures would suggest that the SURF is not measuring constructs that are different from adolescent subjective well-being. Taken together, these measures provide evidence that the SURF is measuring what it purports to measure.

Of note, we planned to examine the test-retest reliability and to calculate the SURF test-retest reliability and Reliable Change Index (RCI), although due to invalid second phase data provided to us by the data collection site, we were unable to conduct these analyses. We discuss this further in the discussion section.

Method

Participants

A total of 380 participants completed the online questionnaire and were compensated through Qualtrics Online Sample. Responses were collected between July and October of 2021. Prior to delivering us the data, Qualtrics screened out 17 participants who failed an attention check item (i.e., “Please answer ‘Strongly Disagree’ to this item”) and 11 participants who completed the study measures faster than two standard deviations from the average. To ensure consistent and valid responses, we created a response validity scale using two matched item pairs; 18 observations of participants whose average response deviation on these items were at least two standard deviations greater than the average were not included in the final analyses. After these participants were removed, a total of 334 participants were included in the analyses.

The participants ranged from 12 to 17 years of age, with a mean age of 14.8 years. Of the participants, 176 (52.7%) were female. Regarding race, 63.17% of the participants identified as White, 14.07% identified as Hispanic or Latino, 13.77% as identified as Black or African American, 4.49% identified as Asian, .6% identified as Native American/American Indian or Alaska Native, .3% identified as Native Hawaiian or Pacific Islander, and 3.39% identified as another race or multiple races, and .3% preferred not to answer. Regarding region, 19.76% of participants lived in the Northeast, 20.66% lived in the Midwest, 18.86% lived in the West, and 40.72% lived in the South. The sample was nationally representative based on region, race, and sex (see Results). Because the study was developed in English, participants who were not able to understand English were excluded from the study. Demographic information is listed in Table 2.

Procedures

The data were collected through an online survey after the study procedures were approved by the Institutional Review Board at Brigham Young University. All individuals contacted were offered the opportunity to participate in the study, though participation was completely voluntary. Inclusion in the study required participants to speak English. Before beginning the study, both parent and child consent forms were completed on the first page of the survey. Upon completion of the consent forms, the participants were either shown a study completion page if they opted out of the study or directed to the first page of the study measures. Participants were asked to complete the measures, lasting approximately 25 minutes, in one sitting. Participants were also sent identical study measures two weeks after the initial measures were completed and were given one week to complete the second session. Participants were compensated for their participation by Qualtrics Online Sample after each session was completed.

Measures

Survey on Flourishing (SURF).

The Survey on Flourishing (SURF) was used as a general measure of subjective well-being (Linford, Salazar, et al., 2022; see Table 1). The SURF is a 20-item Likert scale measure which assesses subjective well-being through tapping into emotional, social, and psychological well-being (. The SURF expands the content included in these areas through including items that examine topics known to promote well-being, but which are not included in commonly used measures, including transcendence (i.e., I often have experiences where I feel connected to something greater than myself.) and vitality (i.e., I usually wake up excited for the day ahead.). The SURF contains 4 negatively worded items. For all questions, respondents rate their level of agreement on a 7-point scale ranging from “strongly disagree” to “strongly agree”. The final score is calculated by taking the total of all items. The SURF requires approximately 5-10 minutes to complete.

In a study examining the psychometric properties of the SURF with adults using multiple samples, the SURF demonstrated high internal consistency ($\alpha = .93 - .96$). It also demonstrated convergent validity by correlating significantly with other measures of subjective well-being, including the PERMA profiler ($r = .82$), the Satisfaction with Life Scale ($r = .74$), and the PANAS Positive Affect subscale ($r = .74$) (Butler & Kern, 2016; Diener et al., 1985; Thompson, 2007; Linford, Salazar, et al., 2022). It also negatively correlated with discriminant measures, such as the Negative Affect subscale of the PANAS ($r = -.61$)

In a previous study involving a small sample of adolescents from a high school in the Mountain West, the SURF also demonstrated high internal consistency ($\alpha = .94$). It also demonstrated convergence with the PERMA Profiler ($r = .79$), the SWLS ($r = .75$), and the

PANAS-Pos ($r = .69$) (Butler & Kern, 2016; Diener et al., 1985; Thompson, 2007; Linford, Bekker, et al., 2022).

Table 1

SURF Items

Item
1. My life is full of joy.
2. Other people genuinely appreciate me and care about me.
3. I often feel like I should be happier than I am. *
4. The things I do in life are valuable and worthwhile.
5. I am very satisfied with the way I am living my life.
6. I usually wake up excited for the day ahead.
7. I feel a strong sense of purpose and meaning in my life.
8. My relationships are supportive and rewarding.
9. Most days, I experience more negative emotions than positive ones. *
10. I feel a genuine sense of connection to other people.
11. I regularly spend time doing things I enjoy
12. I often feel like no one understands me. *
13. I feel happy and peaceful most of the time.
14. There is very little (if anything) I would change about my life.
15. I do many things that contribute to others' well-being.
16. I often feel like it's a struggle to get through the day. *
17. I can achieve all the goals I set for myself.
18. I feel intense gratitude to be alive.
19. I often have experiences where I feel connected to something greater than myself.
20. I would say I'm making very good progress in life.

Note: * = Reverse scored items

Positive and Negative Emotion Schedule, Short Form (PANAS; Watson et al., 1988).

Because affective experience is an essential component of subjective well-being, the PANAS was used to measure positive and negative affect. The positive affect subscale of the PANAS was used as evidence for convergent validity, and the negative affect subscale was used for discriminant validity. The PANAS-SF contains 20 items, each of which is a positive affect word (e.g., “enthusiastic” or “inspired”) or a negative affect word (e.g., “scared” or “hostile”). Respondents then use a five-point Likert scale to report the extent that they are currently experiencing each emotion. Previous research has estimated the test-retest reliability after an 8-week period to be .54 for the positive affect scale and .45 on the negative affect scale (Watson et al., 1988).

Satisfaction with Life Scale (SWLS; Diener, et al., 1985).

The Satisfaction with Life Scale was used to measure overall life satisfaction and was also used to measure convergent validity. The SWLS is a 5-item Likert style scale, which sums the item responses to estimate a total subjective well-being score. It is the most commonly used instrument to measure life satisfaction and has research supporting its reliability and validity in many populations including adolescents (Neto, 1993). The SWLS has demonstrated good test-retest reliability both after a 2-week period ($r = .83$) and after a period of one month ($r = .84$; Alfonso et al., 1996; Pavot, et al., 1991). Those people expected to report low life satisfaction scores (e.g., prison inmates, women experiencing intimate partner violence, and psychiatric patients) demonstrated low scores on the SWLS (Pavot et al., 1991). The SWLS also demonstrated convergent validity; the SWLS correlated significantly with other subjective well-being measures including the Andrews/Withey Scale ($r = .52-.68$) and the Fordyce Global Scale

($r = .55-.82$), as well as interviewer ratings ($r = .43-.66$) and informant reports of well-being ($r = .28-.58$) (Diener et al., 1985; Larsen et al., 1985; Diener et al., 1991; Pavot & Diener, 1993; Linford, Salazar, et al., 2022)

Data Analyses

After the data was collected, the Stata v16.1 statistical package was used to analyze the results. To determine the internal consistency of the SURF, Cronbach's alpha (α) and Pearson bivariate correlations were calculated. To examine the factor structure of the SURF, we used confirmatory factor analysis to compare the fit of four competing models.

The primary model we examined was a one-factor model with the latent variable of "subjective well-being" predicting scores on each item. Our decision to run a one-factor model was grounded in previous research examining the performance of the original SURF in an adult population, which demonstrated that the SURF items loaded onto a single factor identified as subjective well-being (Linford, Salazar, et al., 2022). Additionally, research on similar measures of well-being found that a general factor of subjective well-being explained a large portion of the variance in users' scores (Bartholomaeus et al., 2020; Reinhardt et al., 2020; Shannon et al., 2020). Although prior research also suggests that a bi-factor model may fit observed data well, we tested a one-factor model of subjective well-being because the SURF was not designed with discrete factors in mind. After we conducted our a priori analysis we used modification indexes in an exploratory fashion to identify other options for improving our primary model's fit, although we determined no modifications were necessary.

The second model we examined was a bi-factor model with all items loading onto a general subjective well-being factor, as well as certain items loading onto secondary factors representing social, emotional, and psychological well-being. These secondary factors were

based on conceptual definitions of subjective well-being, and the item assignments were determined by a qualitative examination of each item's content. We included this model after examining previous literature which suggested bi-factor models have shown effective fit with alternate measures of well-being, such as the MHC, PERMA, and the WEMWBS discussed above.

In addition to the two models listed above, we also ran two other models to investigate the impact of negatively worded items on the SURF. Research suggests that having an unequal amount of negatively and positively worded items may cause an unintended, "negatively worded item" factor to emerge during a CFA due to response bias (DiStefano & Motl, 2006; Merritt, 2012). The third model we examined was a two-factor model, which consisted of a broad subjective well-being factor and a negatively worded item factor. This model examined the possibility of whether negatively worded items resulted in a statistical artifact. This type of model has been used in a similar measurement study to examine the impact of negatively worded items on the measurement instrument (Ryff & Keyes, 1995).

The final model we examined was a one-factor model of broad subjective well-being, similar to Model 1. However, with this model we aimed to account for the effect of negatively worded items by allowing the error variances of the four negatively worded items to covary. We compared this model to the two-factor model to explore whether the negatively worded items in the measure comprised an independent factor, or whether they fit better in a one-factor model.

Lastly, the SURF mean scores were correlated with the PANAS positive affect subscale, PANAS negative affect subscale, and the Satisfaction with Life Scale total scores to determine the measure's convergent and discriminant validity.

Results

The aims of this study were to determine reliability and validity of the SURF, and to examine its factor structure. After cleaning the data, we examined the SURF's internal consistency, convergent and discriminant validity, and the factor structure.

Data Preparation

Prior to running our main analysis, we conducted preliminary analyses to determine whether our data met the assumptions of normality for our planned statistical tests. We first identified outliers in the mean scores for the SURF, PANAS scales, and SWLS. We defined outliers as observations that lay outside the bounds of two standard deviation units greater or less than the median score. We then fenced outliers to these outer bounds (median plus or minus two interquartile ranges). Ultimately, we identified and fenced 11 observations to the lower bound of the surf total score, 10 observations to the upper bound of the PANAS Negative Affect subscale, and nine observations to the upper bound of the SWLS.

We then examined the univariate normality of the SURF, PANAS subscale, and SWLS scores. A joint chi-squared probability test for normality demonstrated that the SURF means ($p > .01$), the PANAS Negative Affect subscale mean ($p > .01$), PANAS Positive Affect subscale ($p > .01$) mean, and SWLS mean ($p > .01$) were each not normally shaped data. However, because of the nature of these measures, we concluded that abnormally shaped data was to be expected. For example, it is likely that most participants demonstrated low levels of negative affect. Thus, we determined that no data transformations would be necessary.

We also conducted a chi-square difference test to determine whether our sample was distinguishable from national statistics. We determined that the sample was nationally

representative based on race ($\chi^2(7, N = 334) = 7.99, p = .33$), gender ($\chi^2(2, N = 334) = .1, p = .95$), and region ($\chi^2(3, N = 334) = 5.84, p = .12$).

Table 2

Demographic Summary and Normative Data from Each Sample

Demographic	Sample Characteristics
Sex*	47% male 52.7% female .3% other
Age	M = 14.8, SD = 1.6
Race*	White = 63.17% Hispanic/Latino = 14.07% Black/African American = 13.77% Native American/Alaska Native = .6% Native Hawaiian/Pacific Islander = .3% Asian = 4.49% Other/mixed = 3.29% Prefer not to answer = .3%
Region*	Midwest = 20.66% Northeast = 19.76% South = 40.72% West = 18.86%
Mean SURF Score	M=4.91, SD=.96

Note: *Nationally representative based on the U.S. 2020 Census

Internal Consistency

Results demonstrated that the SURF’s internal consistency was high ($\alpha = .92$; see Table 5). The average inter-item correlation for the SURF was .36. These results support our expectation that the measure would have good internal consistency. Of note, the SWLS ($\alpha = .86$), PANAS positive affect subscale ($\alpha = .94$), and the PANAS negative affect subscale ($\alpha = .92$) also showed high internal consistency.

Factor Structure

Although Cronbach's alpha provides evidence for internal consistency, it alone does not provide adequate information about the dimensionality or factor structure for the SURF. To obtain information about the factor structure of the test items, we conducted a confirmatory factor analysis (CFA). Previous research examining the factor structure for the original SURF questionnaire found that all test items loaded strongly onto one general factor, which was identified as subjective well-being. Because we expect that the SURF items are broad enough to allow for differences in interpretation between the general population and adolescents yet specific enough to retain good interpretability, we expected a one-factor model would demonstrate good fit with the observed data. We compared our primary model's fit with an alternate bi-factor model based on the current literature.

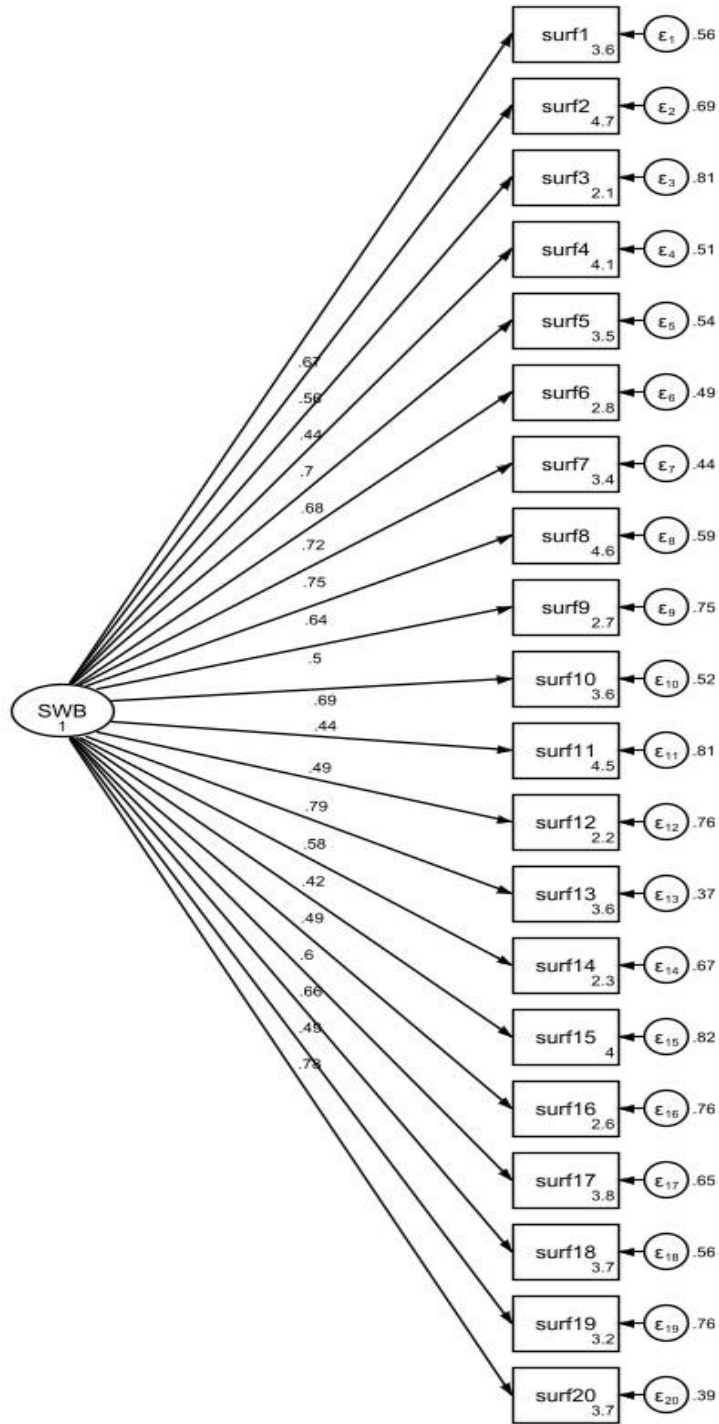
Model 1

Model 1 consisted of a one-factor model, with each item loading onto a latent variable representing subjective well-being (see Figure 1). This model was identified according to the three-indicator rule (Davis, 1993). According to this rule, a single-factor model is identified if it has three or more indicators and if no error terms are correlated. Our primary one-factor model of subjective well-being demonstrated adequate fit to the data ($\chi^2(170, N = 334) = 528.51, p < .001$; model fit statistics can be seen in Table 4). The model's root-mean-square error of approximation (RMSEA) = .08 which suggests moderate fit when considering the parsimony of the model (RMSEA values of $< .08$ indicate "acceptable" fit, while values $< .05$ indicate "good" fit; Schumacher & Lomax, 2016). The standardized root mean squared residual (SRMR), which reports the average difference between the observed and implied covariances for the surf items was .06, which suggests moderate fit. The confirmatory fit index (CFI), which compares how well the identified model compares to a null model, was .87. This suggests the identified model

fits 87% better than a null model (a CFI above .90 is said to have adequate fit; Hu & Bentler, 1999). The Bayesian Information Criterion (BIC), which can be compared to the BIC of other models to examine relative fit, was 22106.87. Taken together, these fit statistics suggest the model demonstrated adequate fit to the data. Per our a priori specifications, we examined model fit indices to examine possible changes to our model to be conducted post-hoc, although we determined no changes were theoretically founded.

Figure 1

Diagram of Model 1 Structure and Factor Loadings



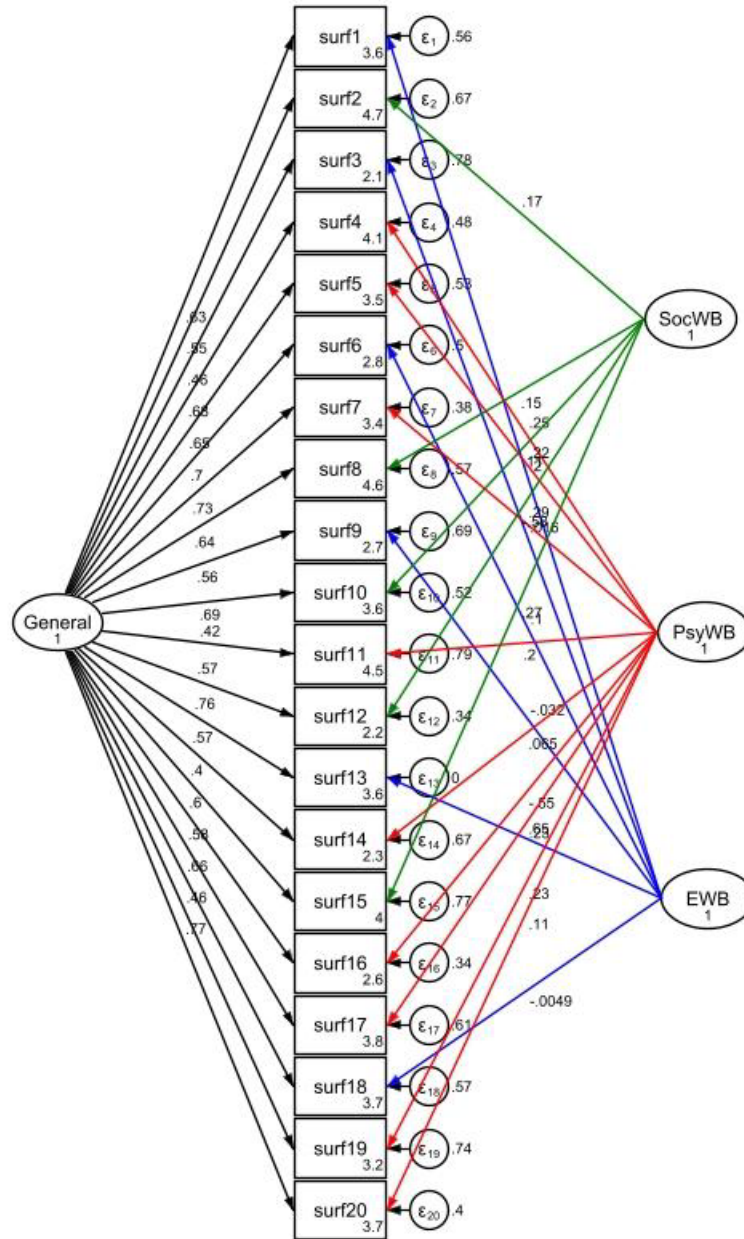
Note: SWB = subjective well-being

Model 2

The second model we examined was a bi-factor model of subjective well-being. In this model, all items loaded on to a broad factor of well-being, but also loaded onto one of three factors representing social, psychological, and emotional well-being (see Figure 2). These factors were determined based on previous literature regarding the factor structure of subjective well-being (Clarke et al., 2011; Reinhardt et al., 2020; Shannon et al., 2020). It is important to note that upon running this model we identified that item 13 had a negative residual variance. To obtain model convergence we set that item's error variance term to zero and proceeded to run the model. This model was theoretically identified by using the t-rule (the number of observed values in the covariance matrix exceeded the number of estimated parameters). The bi-factor model demonstrated adequate model fit, with $\chi^2(151, N = 334) = 433.23, p > .00$; RMSEA = .08; SRMR = .06; CFI = .90; BIC = 22122.00. Overall, these data suggest this model demonstrated a similar fit to Model 1.

Figure 2

Diagram of Model 2 Structure and Factor Loadings



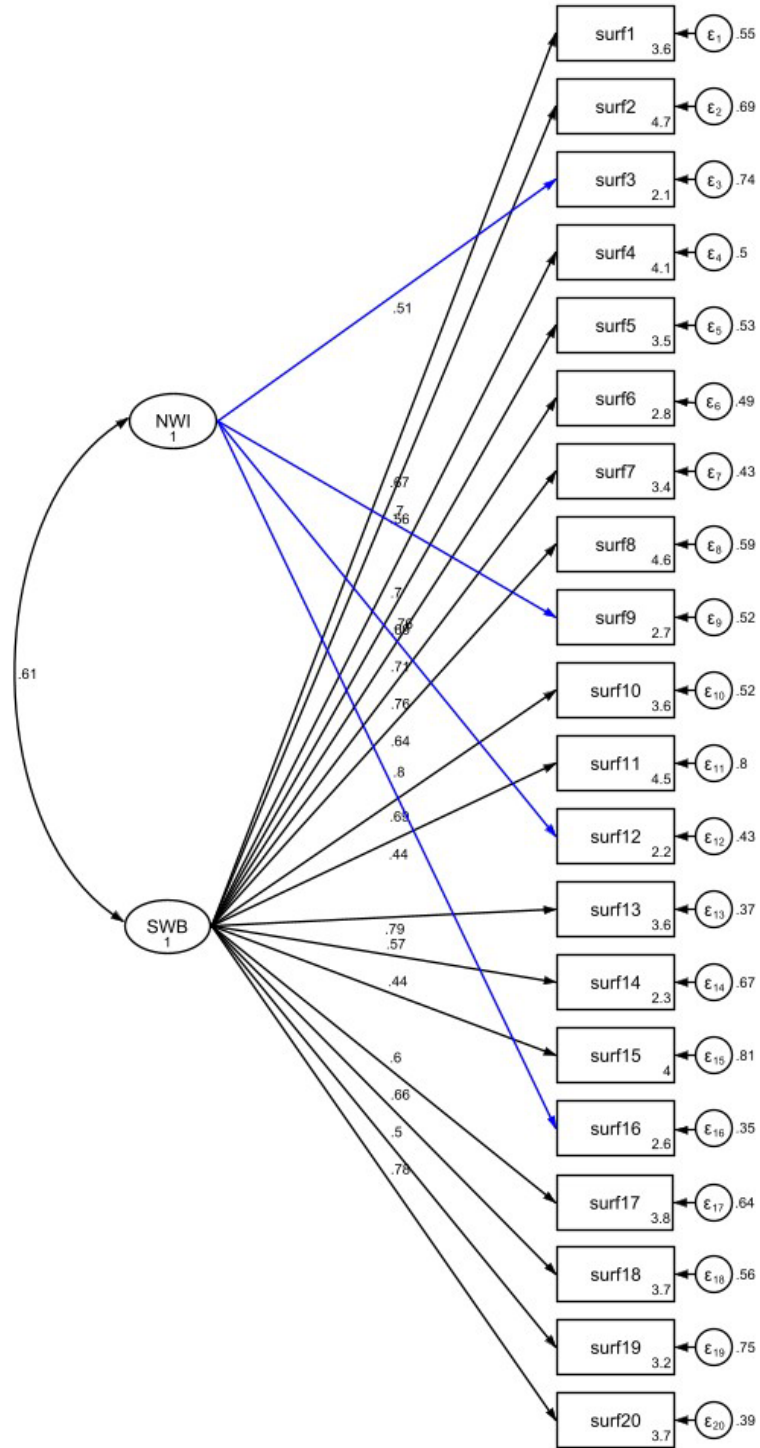
Note: Note: SocWB = social well-being; PsyWB = psychological well-being; EWB = emotional well-being.

Model 3

The third model we examined was a two-factor model of subjective well-being, with one factor representing subjective well-being and a second factor representing negatively worded items (see Figure 3). We ran this model to explore whether negatively worded items may have created an artifact in the data. Because the SURF consisted of an unequal number of positively and negatively worded items, we suspected that this may cause some variance in items to be due to response bias as opposed to true score. This model was also identified according to the three-indicator rule (Davis, 1993). According to this rule, a two-factor model is identified if each latent variable has three or more indicators, no error terms are correlated, and if each indicator loads onto only one factor. Model 4 demonstrated good fit to the data, with $\chi^2(169, N = 334) = 326.88, p > .00; RMSEA = .05; SRMR = .05; CFI = .94; BIC = 21911.05$. Taken together, this model demonstrated good fit to the data and accounted for approximately 98% of the variance in SURF total scores.

Figure 3

Diagram of Model 3 Structure and Factor Loadings



Note: SWB = subjective well-being; NWI = negatively worded items

Model 4

Lastly, Model 4 was also a one-factor model of subjective well-being similar to Model 1, although we correlated the error variance terms between the negatively worded items (see Figure 4). This model was theoretically identified by using the t-rule. Model 4 demonstrated good fit to the data, with $\chi^2(164, N = 334) = 307.09, p < .001$; RMSEA = .05; SRMR = .04; CFI = .95; BIC = 21920.317. Taken together, this model demonstrated good fit to the data and accounted for approximately 92% of the variance in SURF total scores. BIC comparisons suggest that this model demonstrated best fit relative to the other models.

Table 3

SURF Items, Factor Descriptors, and Loadings for the Modified One-factor Model.

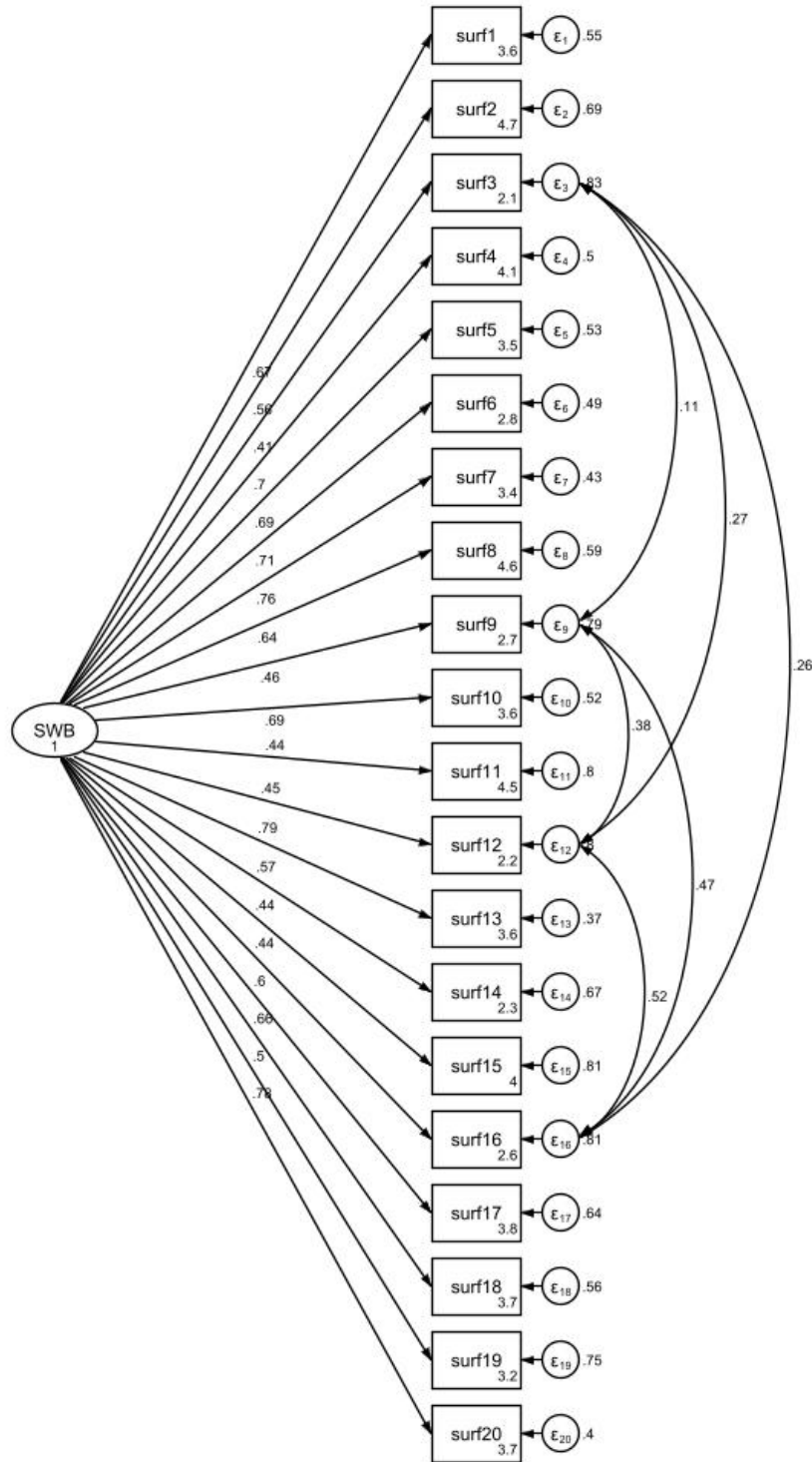
Item	Factor Loading	R ²
1. My life is full of joy.	.67	.45
2. Other people genuinely appreciate me and care about me.	.56	.31
3. I often feel like I should be happier than I am. *	.41	.17
4. The things I do in life are valuable and worthwhile.	.70	.50
5. I am very satisfied with the way I am living my life.	.69	.47
6. I usually wake up excited for the day ahead.	.71	.51
7. I feel a strong sense of purpose and meaning in my life.	.76	.57
8. My relationships are supportive and rewarding.	.64	.41
9. Most days, I experience more negative emotions than positive ones. *	.46	.21

10. I feel a genuine sense of connection to other people.	.69	.48
11. I regularly spend time doing things I enjoy	.44	.20
12. I often feel like no one understands me. *	.45	.20
13. I feel happy and peaceful most of the time.	.79	.63
14. There is very little (if anything) I would change about my life.	.57	.33
15. I do many things that contribute to others' well-being.	.44	.19
16. I often feel like it's a struggle to get through the day. *	.44	.19
17. I can achieve all the goals I set for myself.	.59	.36
18. I feel intense gratitude to be alive.	.66	.44
19. I often have experiences where I feel connected to something greater than myself.	.49	.24
20. I would say I'm making very good progress in life.	.78	.60

Note: *Reverse Scored Items

Figure 4

Diagram of Model 4 Structure and Factor Loadings



Note: SWB = subjective well-being

In summary, the primary one-factor model and the bi-factor model both demonstrated adequate fit. The two-factor model (including the latent variables of subjective well-being and negatively worded items) and the modified one-factor model each demonstrated good fit. We concluded that the modified one-factor model (Model 4) demonstrated best fit with the data. We discuss these results further below.

Table 4

Fit Statistics for the Models Tested

Model		χ^2 (N = 334)	df	RMSEA	SRMR	CFI	BIC
1	One-factor	528.51	169	.08	.06	.87	22106.87
2	Bifactor Orthogonal	433.23	150	.08	.06	.90	22122.00
3	Two-factor	326.88	168	.05	.05	.94	21911.05
4	One-factor (with error term covariances)	307.09	164	.05	.04	.95	21920.32

Convergent and Discriminant Validity

To examine the validity of the SURF, we correlated the total scores of the PANAS subscales and the SWLS with the SURF to estimate convergent and discriminant validity. The SURF total scores demonstrated a significant positive correlation with the SWLS ($r = .70$, 95% CI [0.64, 0.75], $p < 0.001$) and the PANAS positive affect subscale ($r = 0.61$, 95% CI [0.54, 0.67], $p < 0.001$). SURF total scores also demonstrated a significant weak negative correlation with the PANAS negative affect subscale, a measure of impaired subjective well-being ($r = -0.20$, 95% CI [-0.30, -0.09], $p < 0.001$). All convergent validity correlations can be found in Table 5.

Table 5*Cronbach's Alpha and Correlations between SURF and Other Measures*

	Cronbach's Alpha	Correlations with SURF (CI)
SURF	.92	1.0
SWLS	.86	.70* (95% CI [0.64, 0.75])
PANAS	.85	.32* (95% CI [0.22, 0.41])
PANAS (Positive Affect Subscale)	.94	.61* (95% CI [0.54, 0.57])
PANAS (Negative Affect Subscale)	.92	-.20* (95% CI [-0.30, -0.09])

Note: * $p < .001$

Discussion

SURF Psychometric Properties and Structure

The purpose of this study was to examine the reliability, factor structure, and validity of the SURF, a measure of subjective well-being, in an adolescent population. Results from this study provided evidence for good internal consistency and convergent/discriminant validity.

Regarding factor structure, our primary one-factor model performed similarly to the alternate bi-factor model. The two-factor and modified one-factor models both showed good fit and suggest that accounting for bias resulting from negatively worded items allowed resulted in the models' good fit. We ultimately suggest Model 4 demonstrates best fit while also balancing parsimony.

Our comparison of the one-factor model (Model 1) to the bi-factor model (Model 2) showed that they fit the observed data very similarly. Additionally, the specific latent variables included in the bi-factor model accounted for a very small portion of variance in the scores after extracting the variance accounted for by the general subjective well-being factor. Interestingly, this finding appears to support prior research examining other bi-factor well-being models where

the secondary factors are weak relative to the primary factor. This suggests that the second-level factors in Model 2 provided very little utility beyond what the general factor was able to account.

We also examined two additional CFA models in an exploratory manner to better understand the impact of item bias on model fit. Regarding the two-factor model, we expected that bias in the participant's response patterns may result in a negatively worded item factor. After examining this model's performance, we expect that it is likely that this would be the case. However, we also recognize that a second possibility as to why this model displayed good fit resides in the content of the negatively worded item factor. It is possible that the negatively worded item factor may represent a substantive construct such as depression or low mood. Items in this factor appear to reflect the absence of positive emotion, supportive relationships, and meaning. However, these negatively worded items were created as counterparts to positively worded items assessing the same content. Thus, we would expect that the positively and negatively worded items would load onto the same factor in absence of response bias. Thus, we conclude that the most likely explanation for the good fit of this model is the presence of response bias, although it is possible that the content of these items may have also impacted model fit.

Model 4, was composed of a single factor representing subjective well-being but which allowed the negatively worded items to covary, also demonstrated good fit. Regarding the model fit, after accounting for the error introduced through item response bias, the negatively worded items still loaded significantly onto the broad subjective well-being factor. This suggests that the SURF items represent a unitary construct. In comparison to the two-factor model, we determined that this model demonstrates similar fit, but it also maximizes parsimony. We determined that this model demonstrated best fit with the data.

Overall, the results from our examination of the SURF's structure suggests that a one-factor model demonstrated best fit with the data, and that the SURF items reflect a unitary construct representing subjective well-being. While the two-factor model and modified one-factor model demonstrated similar fit, we expect that these fit well because they accounted for the presence of item response bias. We retained the modified one-factor model because it appears to be the most parsimonious option while still demonstrating good fit with the data (see Table 4).

SURF Performance in Adults Versus Adolescents

The results from this study align with past research conducted in adult and adolescent populations. First, the SURF demonstrated similar internal consistency among both populations ($\alpha > .90$). Regarding the structure of the SURF, a prior exploratory factor analysis (EFA) suggested that the SURF structure consisted of one broad factor which the authors described as subjective well-being (Linford, Bekker, et al., 2022). This study also demonstrated that a one-factor model best characterizes the SURF, although we also suggest that future measurement studies be aware of how negatively worded items may influence the test results. Overall, the data on the SURF when used on both adults and adolescents suggests that the SURF measures a unitary construct representing subjective well-being.

Lastly, the SURF demonstrated good convergent validity when compared to the PANAS-pos and the SWLS, and discriminant validity when compared to the PANAS-neg for both adolescents and adults. While we hoped to include additional measures of discriminant validity (e.g., YOQ-30.2; Wells et al., 1996), our available resources limited our ability to include other measures. Future research may focus on expanding the discriminant validity of the SURF.

Other Contributions to Adolescent Well-Being Measurement

In addition to examining the psychometric properties of the SURF, this study adds to the current research in significant ways. First, the SURF represents a measure of adolescent subjective well-being that is accessible, quick to administer, and free to use with acknowledgement. Many current measures of adolescent subjective well-being are lacking in one or more of these areas, making them inappropriate for widespread use. The SURF provides a useful alternative to these measures.

Second, the SURF provides a reliable measurement tool whose psychometric properties have been examined using a nationally representative sample within the United States. While there are a few measures that currently exist, only one of them used a representative US sample in their validation studies. Because of the difficulty in collecting data on adolescent participants, many researchers use convenience sampling methods or existing infrastructure (e.g., school systems) to provide them with participants. Although data may be more easily gathered through these methods, oftentimes this limits the generalizability of the findings. Because our data represents responses from a wide range of adolescents across the United States, we may be more confident of SURF's utility when used in the broader US population.

Third, the SURF items contain content not yet included in existing adolescent subjective well-being measurement. This includes items on mindfulness, transcendence, and gratitude, which have been shown to be important to well-being. Additionally, because it is difficult to capture the breadth of this construct with a single measure, it may be necessary to use multiple measures to best understand adolescent subjective well-being. The SURF is a tool which can overlap with other measures to produce a clearer understanding of adolescent subjective well-being. Overall, the SURF is an effective and research-supported measure that can help researchers and practitioners better understand adolescent well-being and how to improve it.

Conclusion

Overall, this study contributes to the current literature by providing a reliable and valid measure of subjective well-being. This study also had several strengths worth noting. First, this study employed a nationally representative sample which adds to the generalizability and utility of this measure. Similarly, a second strength of our study was the size of our sample. Clarke and Watson suggested that researchers collect at least ten responses per scale item, with an ideal ratio of 15:1 or 20:1 (Clark & Watson, 2016; Morgado et al., 2017). We analyzed the responses of 334 participants, which resulted in a ratio of approximately 17:1. Including a sufficient number of participants helps ensure the data obtained is reliable. We consider this a strength because, according to a review of scale development, researchers concluded that approximately half of the studies included in their review did not meet the Clark & Watson's suggested 10:1 ratio (Morgado et al., 2017).

A third strength of our study relates to our adherence to recommendations suggesting transparency in research methods and planned analyses (Nosek et al., 2018; Silberzahn et al., 2018). Unfortunately, we did not pre-register the analyses for this study. However, prior to conducting any analyses, we specified a priori which statistical procedures we planned on running given our research questions. Additionally, we specified which analyses were conducted in an exploratory manner after running our main analyses. Having a data-analysis plan reduces bias which may result from questionable research practices, which researchers have shown to be extremely common among social scientists (John et al., 2012; Simmons et al., 2011). Thus, our commitment to adhering to our data-analysis plan and reporting the results in a transparent manner helps increase replicability and gives evidence for the robustness of our findings.

Limitations

Although our study has notable strengths, it is also important to recognize limitations that may have impacted our results. Many of these limitations are related to our decision to use Qualtrics Online Sample to assist with distributing our study and collecting the data. First, although working with online panels allowed us to collect a nationally representative sample, some research regarding Amazon's Mechanical Turk (MTurk), a similar online data collection site, suggests that response quality gathered from online data collection agencies may often be low. In addition to data screening methods introduced by Qualtrics, we also introduced safeguards (e.g., validity metrics) to ensure high response quality (Kees et al., 2017).

Second, some research suggests that online data collection sites may saturate the responses with participants who are not representative of the intended sample, despite responses to the demographic questions (Kees et al., 2017). Although we did not detect any abnormalities that might suggest problems with our sample, this still represents a risk and is a limitation to our study.

Third, because the online panel oversaw distributing the study measures, the second administration of the study was distributed to participants at an incorrect time, not in accordance with our outlined methods (see Methods section). This ultimately invalidated the re-test responses and restricted us from analyzing important results such as test-retest data.

Future Directions

This study highlights several future directions to improve the psychometric properties of the SURF. First, we recommend that additional data samples be collected to replicate these results. Although this study provides a good first step, the robustness of these findings would be increased if other adolescent samples found similar results. Future studies may also consider collecting samples which are less influenced by bias, and which are of higher quality.

In addition to replication, there are several things future studies may consider to extend our understanding of the SURF's psychometric properties. We suggest that future studies prioritize collecting retest data. This will provide additional evidence for the reliability and stability of SURF scores over time. We suggest that future studies also include more measures of convergent and discriminant validity, possibly including objective measures, such as those that examine physical environment, access to resources, and physiology. Although there is still some evidence for this, including other similar and dissimilar measures would provide greater evidence that the SURF measures what it purports to measure.

Lastly, we recommend that researchers include the SURF in intervention studies aimed to improve adolescent subjective well-being. One goal we had when developing the SURF was that the measure would provide a way for researchers to track changes in adolescent subjective well-being over time, and in response to intervention. By including the SURF as an outcome measure researchers may calculate the reliable change index (RCI) of the SURF and determine whether this measure is an appropriate tool for that use. Addressing these issues will help further establish the SURF as a valid and reliable measure of adolescent subjective well-being.

References

- Alfonso, V.C., Allison, D.B., Rader, D.E. et al. (1996). The extended satisfaction with life scale: Development and psychometric properties. *Soc Indic Res*, 38, 275–301.
- Allgaier, A. K., Pietsch, K., Frühe, B., Prast, E., Sigl-Glöckner, J., & Schulte-Körne, G. (2012). Depression in pediatric care: Is the WHO-Five Well-Being Index a valid screening instrument for children and adolescents? *General Hospital Psychiatry*, 34(3), 234-241.
- Bartholomaeus, J. D., Iasiello, M. P., Jarden, A., Burke, K. J., & van Agteren, J. (2020). Evaluating the psychometric properties of the PERMA Profiler. *Journal of Well-Being Assessment*, 4(2), 163–180. <https://doi.org/10.1007/s41543-020-00031-3>
- Buckley, M., Storino, M., & Saarni, C. (2003). Promoting emotional competence in children and adolescents: Implications for school psychologists. *School Psychology Quarterly*, 18(2), 177.
- Butler, J., & Kern, M. L. (2016). The PERMA-Profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing*, 6(3).
- Clark, L. A., & Watson, D. (2016). *Constructing validity: Basic issues in objective scale development* (p. 203). American Psychological Association. <https://doi.org/10.1037/14805-012>
- Clarke, A., Friede, T., Putz, R., Ashdown, J., Martin, S., Blake, A., Adi, Y., Parkinson, J., Flynn, P., & Platt, S. (2011). Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Validated for teenage school students in England and Scotland. A mixed methods assessment. *BMC Public Health*, 11(1), 1–9.

- Copeland, E. P., Nelson, R. B., & Traugher, M. C. (2010). Wellness dimensions relate to happiness in children and adolescents. *Advances in School Mental Health Promotion*, 3(4), 25–37. <https://doi.org/10.1080/1754730X.2010.9715689>
- Crum, A. J., Salovey, P., & Achor, S. (2013). Rethinking stress: The role of mindsets in determining the stress response. *Journal of Personality and Social Psychology*, 104(4), 716–733. <https://doi.org/10.1037/a0031201>
- Damon, W. (2004). What is positive youth development? *The ANNALS of the American Academy of Political and Social Science*, 591(1), 13–24. <https://doi.org/10.1177/0002716203260092>
- Davis, W. R. (1993). The FC1 rule of identification for confirmatory factor analysis: A general sufficient condition. *Sociological Methods & Research*, 21(4), 403–437. <https://doi.org/10.1177/0049124193021004001>
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95(3), 542–575. APA PsycArticles. <https://doi.org/10.1037/0033-2909.95.3.542>
- Diener, E., & Biswas-Diener, R. (2011). *Happiness: Unlocking the mysteries of psychological wealth*. John Wiley & Sons.
- Diener, E., Colvin, C. R., Pavot, W. G., & Allman, A. (1991). The psychic costs of intense positive affect. *Journal of Personality and Social Psychology*, 61(3), 492.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49(1), 71–75.
- Diener, E., & Ryan, K. (2009). Subjective well-being: A general overview. *South African Journal of Psychology*, 39(4), 391–406. <https://doi.org/10.1177/008124630903900402>

- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(3), 440–464. https://doi.org/10.1207/s15328007sem1303_6
- Dodge, R., Daly, A. P., Huyton, J., & Sanders, L. D. (2012). The challenge of defining wellbeing. *International Journal of Wellbeing*, *2*(3), Article 3. <https://www.internationaljournalofwellbeing.org/index.php/ijow/article/view/89>
- Fernandes, H. M., Vasconcelos-Raposo, J., & Teixeira, C. M. (2010). Preliminary analysis of the psychometric properties of Ryff's scales of psychological well-being in Portuguese adolescents. *The Spanish Journal of Psychology*, *13*(2), 1032-1043.
- 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*(3), 1011–1032.
- Howell, A. J. (2009). Flourishing: Achievement-related correlates of students' well-being. *The Journal of Positive Psychology*, *4*(1), 1–13. <https://doi.org/10.1080/17439760802043459>
- 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.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532.
- Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's mechanical turk. *Journal of Advertising*, *46*(1), 141–155. <https://doi.org/10.1080/00913367.2016.1269304>

- Kern, M. L., Benson, L., Steinberg, E. A., & Steinberg, L. (2016). The EPOCH measure of adolescent well-being. *Psychological Assessment, 28*(5), 586–597.
<https://doi.org/10.1037/pas0000201>
- Keyes, C. L.(1998). Social well-being. *Social Psychology Quarterly, 61*, 121–140.
<http://dx.doi.org/10.2307/2787065>
- Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology, 73*(3), 539.
- Keyes, C. L. (2006). The subjective well-being of America’s youth: Toward a comprehensive assessment. *Adolescent & Family Health, 4*(1), 3–11.
- Larsen, R. J., Diener, E. D., & Emmons, R. A. (1985). An evaluation of subjective well-being measures. *Social Indicators Research, 17*(1), 1-17.
- Lerner, J. V., Phelps, E., Forman, Y., & Bowers, E. P. (2009). Positive youth development. In *Handbook of adolescent psychology: Individual bases of adolescent development, Vol. 1*, (3rd ed., pp. 524–558). John Wiley & Sons Inc.
<https://doi.org/10.1002/9780470479193.adlpsy001016>
- Linford, L., Bekker, J., Ameen, J., & Warren, J. (2022). Implementation of a positive psychology curriculum in a high school setting: A mixed methods pilot study. *Journal of Positive School Psychology, 6*(1), 25-37.
- Linford, L., Salazar, G., Jackman, K., Warren, J. (2022). *The Survey on Flourishing (SURF): Development of a comprehensive measure of well-being*. Manuscript in preparation.
- Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologist, 53*(2), 205.

- Maygar, J., & Keyes, C. (2019). Defining, measuring, and applying subjective well-being. In M. Gallagher & S. Lopez (Eds.), *Positive psychological assessment: A handbook of models and measures* (2nd ed., pp. 380-415). <http://dx.doi.org.proxy1.ncu.edu/10.1037/0000138-025>
- Merritt, S. M. (2012). The two-factor solution to Allen and Meyer's (1990) affective commitment scale: Effects of negatively worded items. *Journal of Business and Psychology*, 27(4), 421–436. <https://doi.org/10.1007/s10869-011-9252-3>
- Moore, K. A., & Keyes, C. L. M. (2003). A brief history of well-being in children and adults. In *Well-being: Positive development across the life course* (pp. 1–11). Lawrence Erlbaum Associates Publishers.
- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, 30. <https://doi.org/10.1186/s41155-016-0057-1>
- Neto, F. (1993). The satisfaction with life scale: Psychometrics properties in an adolescent sample. *Journal of Youth and Adolescence*, 22(2), 125–134. <https://doi.org/10.1007/BF01536648>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.
- Ong, A. D. (2010). Pathways linking positive emotion and health in later life. *Current Directions in Psychological Science*, 19(6), 358–362. <https://doi.org/10.1177/0963721410388805>
- Pavot, W., & Diener, E. (1993). The affective and cognitive context of self-reported measures of subjective well-being. *Social Indicators Research*, 28(1), 1-20.

- Pavot, W., Diener, E. D., Colvin, C. R., & Sandvik, E. (1991). Further validation of the Satisfaction with Life Scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment*, *57*(1), 149-161.
- Pérez-Garín, D., Molero, F., & Bos, A. E. R. (2015). Internalized mental illness stigma and subjective well-being: The mediating role of psychological well-being. *Psychiatry Research*, *228*(3), 325–331. <https://doi.org/10.1016/j.psychres.2015.06.029>
- Reinhardt, M., Horváth, Z., Morgan, A., & Kökönyei, G. (2020). Well-being profiles in adolescence: Psychometric properties and latent profile analysis of the mental health continuum model – a methodological study. *Health and Quality of Life Outcomes*, *18*(1), 95. <https://doi.org/10.1186/s12955-020-01332-0>
- Rose, T., Joe, S., Williams, A., Harris, R., Betz, G., & Stewart-Brown, S. (2017). Measuring mental wellbeing among adolescents: A systematic review of instruments. *Journal of Child and Family Studies*, *26*(9), 2349–2362. <https://doi.org/10.1007/s10826-017-0754-0>
- Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, *52*(1), 141–166. <https://doi.org/10.1146/annurev.psych.52.1.141>
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*, 1069–1081. <http://dx.doi.org/10.1037/0022-3514.57.6.1069>
- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719–727. <https://doi-org.erl.lib.byu.edu/10.1037/0022-3514.69.4.719>

- Seligman, M. (2018). PERMA and the building blocks of well-being. *The Journal of Positive Psychology, 13*(4), 333–335.
- Seligman, M. (2011). Flourish: A visionary new understanding of happiness and well-being. *Policy, 27*(3), 60–61.
- Shannon, S., Breslin, G., Prentice, G., & Leavey, G. (2020). Testing the factor structure of the Warwick-Edinburgh Mental Well-Being Scale in adolescents: A bi-factor modelling methodology. *Psychiatry Research, 293*, 113393.
<https://doi.org/10.1016/j.psychres.2020.113393>
- Shek, D. T., Dou, D., Zhu, X., & Chai, W. (2019). Positive youth development: Current perspectives. *Adolescent Health, Medicine and Therapeutics, 10*, 131–141.
<https://doi.org/10.2147/AHMT.S179946>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., ... Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science, 1*(3), 337–356. <https://doi.org/10.1177/2515245917747646>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., ... & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes, 5*(1), 1-13.

- Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-cultural Psychology, 38*(2), 227-242.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063.
- Weich, S., Brugha, T., King, M., McManus, S., Bebbington, P., Jenkins, R., Cooper, C., McBride, O., & Stewart-Brown, S. (2011). Mental well-being and mental illness: Findings from the Adult Psychiatric Morbidity Survey for England 2007. *The British Journal of Psychiatry, 199*(1), 23–28. <https://doi.org/10.1192/bjp.bp.111.091496>
- Wells, M. G., Burlingame, G. M., Lambert, M. J., Hoag, M. J., & Hope, C. A. (1996). Conceptualization and measurement of patient change during psychotherapy: Development of the Outcome Questionnaire and Youth Outcome Questionnaire. *Psychotherapy: Theory, Research, Practice, Training, 33*(2), 275.