

## RESEARCH ARTICLE

# Developing a Six-Item Short Form of Ryff's Psychological Well-Being Measure: Demonstration of Short Form Development Using Exhaustive Search

Mohsen Joshanloo 

Department of Psychology, Keimyung University, Daegu, South Korea

**Correspondence:** Mohsen Joshanloo ([mjoshanloo@gmail.com](mailto:mjoshanloo@gmail.com))

**Received:** 23 April 2025 | **Revised:** 25 February 2026 | **Accepted:** 20 March 2026

**Keywords:** exhaustive search | MIDUS | Ryff's measure of psychological well-being | scale shortening | short form development

## ABSTRACT

Using the Midlife in the United States data, this study developed a 6-item short form of Ryff's six-dimensional measure of psychological well-being, with a single item selected to represent each dimension. The study demonstrated the utility of an exhaustive search framework in which all permissible item combinations were systematically evaluated to identify an optimal ultra-short version. Model performance was assessed using global goodness of fit indices (e.g., RMSEA, CFI, SRMR), standardized factor loadings, and coefficient omega to ensure adequate construct representation, unidimensionality, and internal consistency. The resulting short form showed excellent model fit, strong internal consistency, and satisfactory representation of the targeted construct across the development sample and an independent replication sample. Measurement invariance analyses supported full invariance across gender and race, and partial invariance across age and education. Taken together, the findings indicate that the proposed ultra-short scale provides a valid and reliable index of global psychological well-being and is suitable for settings in which domain-level scores are not required and in which time or resource constraints preclude administration of longer measures.

## 1 | Introduction

Constraints inherent in psychological research, such as limited resources, participant fatigue, time-restricted assessments, and the demands of large-scale data collection, necessitate the strategic development of abbreviated psychological scales (Ziegler et al. 2014). The primary methodological challenge lies in achieving optimal parsimony without sacrificing the psychometric robustness of the original measure. In general, a good psychological scale, and in particular a good short-form scale, must meet at least three criteria (Strauss and Smith 2009; Widaman et al. 2011):

1. **Construct Representation:** The short form must preserve the theoretical foundation of the original scale by adequately capturing its core content and the psychological

construct it aims to measure (Whitely 1983). This is achieved by ensuring that each conceptual dimension is represented by items in the short form, and items with high factor loadings are chosen. In this study, the six-item set was required to include one representative item from each of the six original dimensions (autonomy, environmental mastery, personal growth, positive relations, purpose in life, and self-acceptance) to ensure adequate construct representation.

2. **Construct Homogeneity:** The scale must exhibit variation along a single, coherent psychological dimension, ensuring that items consistently represent a unified construct (Smith et al. 2012). This requirement can be evaluated using model fit indices to confirm unidimensionality. In this study, although the six selected

items were each drawn from a separate domain, they were expected to behave as indicators of a single latent factor of psychological well-being (i.e., a general well-being construct).

3. Internal Consistency: Despite having fewer items, the scale must maintain high internal consistency, demonstrating that its items are interrelated. Metrics such as omega reliability provide accurate assessments of internal consistency (Flora 2020).

However, these criteria are not always perfectly aligned. For example, a model with the best goodness-of-fit indices may not achieve the highest reliability, and a highly reliable scale may fail to represent the full conceptual breadth of the construct. Overemphasis on internal consistency, for instance, might result in a short form that is psychometrically robust but lacks theoretical comprehensiveness (Smith et al. 2012). These tensions highlight the importance of adopting a multi-criteria framework that balances construct representation, homogeneity, and internal consistency in short-form scale development.

Conventional methods for scale reduction, such as selecting items based on item-total correlations, factor loadings (from Principal Component Analysis or Confirmatory Factor Analysis), or face validity, provide a solid foundation for short-scale development (Browne et al. 2018; Widaman et al. 2011). However, these approaches may involve too many subjective decisions and may overlook optimal configurations, particularly if used in isolation. Recent advances in computational methodologies, facilitated by high-performance computing and parallel processing capabilities now available on standard personal computers, have transformed the landscape of long or short-form scale development (Gonzalez 2020; Dörendahl and Greiff 2020). Researchers can now systematically evaluate thousands or even millions of potential configurations in a feasible time-frame, enabling a more comprehensive and objective approach to model selection.

This computational framework offers two critical advantages. First, it allows for the systematic application of multiple evaluation criteria to assess and compare short-form scales. Second, by exhaustively or randomly testing numerous models, it minimizes subjectivity and reduces the risk of selecting suboptimal configurations to some extent. This study used these computational advancements to implement a systematic approach to scale reduction using real data, focusing on the challenging task of developing a unidimensional short form for a long multidimensional construct.

## 1.1 | The Need for a Short Form of the Psychological Well-Being Measure

Ryff's six-dimensional model of psychological well-being (Ryff 1989) is one of the most widely cited frameworks for understanding psychological and eudaimonic well-being (Vittersø 2016). Its associated measure, comprising 42 items (seven per dimension), is extensively used in research but presents practical challenges due to its length in some contexts. Although shorter versions have been developed, the briefest available form includes 18 items (Garcia et al. 2023). Given the

robustness and theoretical grounding of Ryff's model, there is a pressing need for a very short version that measures general psychological well-being as a unidimensional concept. A 6-item single-item-per-dimension measure could address this gap, offering a practical solution for studies requiring brevity but also committed to the theoretical framework of the scale. Alternative short measures for eudaimonic well-being have recently been developed to address the need for more concise assessment tools (e.g., Diener et al. 2009; Margolis et al. 2022). However, these scales are based on different theoretical foundations that do not align with Ryff's conceptual model of psychological well-being.

Notably, constructing such a short form is inherently challenging. Identifying the top-performing item from each dimension does not guarantee that the resulting six items will collectively meet all psychometric criteria. For instance, items with high individual factor loadings may not combine to produce a model with good overall fit or reliability. This underscores the need for systematic and rigorous approaches to short-form scale development. As shown below, the computational framework shown here is well-suited to solve this practical issue.

## 1.2 | The Present Study

The purpose of this study was to develop the shortest possible form of Ryff's scale that could be deployed in time- and resource-constrained contexts, in which respondent burden is a central concern and a single summary score is often needed. The one-item-per-domain rule was adopted to preserve broad content representation of Ryff's six theoretical domains while maintaining minimal length. Given this representation constraint, six items constitute the minimum possible length, because fewer than six items would necessarily omit at least one domain.

This study employs an *Exhaustive* or *Brute-Force* Search strategy to identify the top six items to maintain. All possible combinations of six items, with one item selected from each dimension, were evaluated to identify the optimal short form. Since each dimension has seven items, and we need to choose one item from each dimension, the total number of combinations is  $7^6 = 117,649$ . The selection process was guided by three equally weighted criteria:

1. Goodness of fit for a one-factor model, ensuring construct homogeneity. The Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Residual (SRMR) were used (Kline 2023).
2. Construct representation, achieved by including one item per dimension and prioritizing items with high standardized factor loadings.
3. Internal consistency, assessed via omega total coefficient.

Some steps in the procedure inevitably involved subjective judgment. A unidimensional general well-being factor was selected to produce an ultra-short index suitable for settings with limited questionnaire space or time. A six-item length was chosen because it allowed the briefest form that still maintained

coverage of the original domains. The requirement of one item per dimension was adopted as an explicit operational definition of construct representation, ensuring that the general factor was indicated by content sampled from each of Ryff's six domains rather than being disproportionately shaped by any single domain. Several criteria were selected to assess the models, namely three global fit indices, reliability, and the standardized factor loadings, and these criteria were assigned equal weight in the composite evaluation. Once the target structure (one-factor), scale length (six items), content constraint (one item per Ryff dimension), and evaluation method had been defined, all admissible item sets were evaluated exhaustively using the same estimation method and the same set of psychometric indices, and models were ranked using a deterministic composite performance score. This approach reduces subjectivity in item selection relative to traditional short-form development approaches that rely on iterative manual item pruning.

### 1.3 | Justification of the Chosen Method Over Available Alternatives

This study uses confirmatory factor analysis (CFA) as the underlying method in combination with an exhaustive search strategy. Other approaches commonly used for short-form development include principal component analysis (PCA) and item response theory (IRT), as well as research strategies that rely on random or heuristic search procedures.

PCA was not used because it serves a different purpose than common factor analysis and does not represent a latent construct model. Factor analysis is based on the common factor model and is intended to explain the correlations among measured variables by positing underlying latent factors, while also modeling unique factors that represent variance not shared with the latent factors, including specific variance and measurement error (Watkins 2021). In contrast, PCA is a data-reduction technique that creates components as linear composites of observed variables. It does not distinguish common variance from unique variance and does not treat the extracted dimensions as latent variables (Fabrigar and Wegener 2012). For these reasons, many researchers prefer factor analysis over PCA for construct measurement and scale development (Irwing and Hughes 2018; Watkins 2021). For example, Gorsuch (1990) recommended that "common factor analysis should be routinely applied as the standard analysis because it recognizes we have error in our variables, gives unbiased instead of inflated loadings, and is more elegant as a part of the standard model used in univariate and multivariate analysis" (p. 39).

CFA was selected instead of IRT because it is widely regarded as the most efficient and standard approach for evaluating dimensionality during scale development. Factor analysis has long been the primary method for assessing dimensionality and remains more commonly used than IRT in early-stage scale construction (Nandakumar 1993). For many applications, factor analysis is also recommended as a preliminary tool to evaluate unidimensionality before fitting IRT models (Reeve et al. 2007; Revicki et al. 2015). In line with this practice, CFA was used here as the most direct framework for testing whether candidate item sets show adequate fit to a single-factor structure.

With respect to the search strategy, exhaustive search was chosen because evaluating all candidate models was feasible. That is, the number of admissible item sets (117,649) was small enough to estimate every model within the available computing resources. When the hypothesis space is computationally manageable, exhaustive search offers a direct way to identify the globally best-performing candidate under the specified objective function. Random and heuristic strategies, including greedy search, genetic algorithms, and simulated annealing (Rajwar et al. 2023), are valuable when the search space is too large to enumerate or when the estimation of all models is costly.

## 2 | Methods

### 2.1 | Measure and Participants

Psychological well-being was assessed using Ryff's 42-item measure, which evaluates six dimensions: self-acceptance, positive relations with others, autonomy, environmental mastery, purpose in life, and personal growth (Ryff 1989). Each subscale comprises seven items. Items were administered on a 7-point Likert-type scale ranging from 1 (strongly agree) to 7 (strongly disagree). Prior to analysis, 21 negatively keyed items were reverse-coded. This measure is included in the Midlife in the United States (MIDUS) project, a nationally representative sample of American adults. Detailed information about this project can be obtained at <https://midus.wisc.edu>. The current study utilized data from the third wave of MIDUS, which included 3294 individuals (54.9% female). The mean age of participants was 63.64 years (SD = 11.35).

### 2.2 | Software and Statistical Analysis

Analyses were conducted using R. To explore all potential item combinations for each dimension of psychological well-being, the *expand.grid* function was used to generate all possible configurations of items. To assess the psychometric properties of each item combination, the *lavaan* package was used, with maximum likelihood estimation. One-factor confirmatory factor analysis models were specified for each combination of items, and fit indices and standardized factor loadings were extracted. Omega (total) coefficients for each model were calculated using the *psych* package. Parallel processing techniques were implemented to optimize computational efficiency. The analysis resulted in a dataset comprising 117,649 rows, each representing a unique set of six items. The first six columns of the dataset contained the six items (identified by their MIDUS variable names), while additional columns provided evaluation metrics.

### 2.3 | Data and Code

The raw MIDUS data and materials can be obtained at <https://www.midus.wisc.edu/data/index.php>. The R script and the data file containing the results for all models are available at [https://osf.io/4pgkb/?view\\_only=b73e2e8420044d5cba76763e40910fdd](https://osf.io/4pgkb/?view_only=b73e2e8420044d5cba76763e40910fdd).

### 3 | Results

To calculate a composite performance score for all tested models and facilitate comparison, the following steps were taken:

1. The six standardized loadings for each model were averaged to produce a single-factor loading metric.
2. All evaluation metrics (fit indices, omega coefficient, and average standardized loading) were z-standardized to bring them on a comparable scale.
3. Equal importance was assigned to model fit, reliability, and factor loading in calculating the overall performance score. First, an average score was computed for model fit. Then, the average fit score, reliability score, and average loading score (all z-standardized) were summed to produce a final composite score.

Because the candidate models were evaluated using multiple fit and quality criteria that are reported on different scales (e.g., incremental fit indices, residual-based indices, and reliability estimates), each criterion was z-standardized across the full set of enumerated models prior to aggregation. This standardization placed all indices on a common metric, thereby preventing any single index from dominating the composite score solely due to its numerical range and ensuring that each criterion contributed equally to the ranking. The z-standardized values were then combined to yield a single composite score for model comparison. Equal weighting was used for model fit, reliability, and factor loadings because no compelling theoretical or methodological rationale was available to privilege one criterion over the others for the intended purpose of developing a broadly usable ultra-short index. It was assumed that each criterion captured a distinct and substantively important aspect of scale quality, such that overemphasis on any single component could have yielded an instrument with strong performance in one domain at the expense of another (e.g., good fit but weak reliability, or high loadings but suboptimal overall fit).

This process resulted in a single performance score for each model, enabling models to be ranked based on their overall performance. The results for the top five and bottom five models are presented in Table 1. As shown, the top models demonstrated excellent performance metrics, while the bottom models exhibited unacceptable properties. Although the first model achieved a slightly higher overall performance score compared to the second model, the latter was selected as the best-performing model. This decision was based on the observation that one of the factor loadings in the first model was close to 0.3, whereas all factor loadings in the second model exceeded 0.4. Although a minimum loading threshold was not prespecified in the present study, the factor-analytic literature indicates that acceptable salient-loading cutoffs commonly range from 0.30 to 0.50 (Morin et al. 2020), with higher loadings generally reflecting stronger indicators. Accordingly, a minimum standardized loading closer to 0.40 was considered as preferable to 0.30 when selecting among the highest-ranked candidate models. Given the similar overall performance between the two models, the second model was judged to be superior for having a higher minimum loading. The resulting items for the best-performing model are displayed in Table 2. The omega coefficient for the selected model was 0.810 (95% CI = 0.798–0.821). Table 3 reports the factor loadings for

TABLE 1 | Top and bottom models.

Aut	Env	Growth	Rel	Purp	Self	RMSEA	CFI	SRMR	Omega	L1	L2	L3	L4	L5	L6	Average loading	Total score
Top 5 models																	
C1SE1M	C1SE1FF	C1SE1O	C1SE1J	C1SE1Q	C1SE1R	0.020	0.998	0.008	0.809	0.316	0.711	0.665	0.559	0.777	0.784	0.635	11.366
<b>C1SE1EE</b>	<b>C1SE1LL</b>	<b>C1SE1GG</b>	<b>C1SE1P</b>	<b>C1SE1Q</b>	<b>C1SE1R</b>	<b>0.025</b>	<b>0.997</b>	<b>0.009</b>	<b>0.810</b>	<b>0.438</b>	<b>0.515</b>	<b>0.612</b>	<b>0.693</b>	<b>0.789</b>	<b>0.785</b>	<b>0.639</b>	<b>11.140</b>
C1SE1EE	C1SE1LL	C1SE1GG	C1SE1J	C1SE1Q	C1SE1R	0.012	0.999	0.008	0.790	0.442	0.524	0.614	0.549	0.769	0.799	0.616	11.094
C1SE1Y	C1SE1FF	C1SE1O	C1SE1J	C1SE1W	C1SE1R	0.020	0.998	0.009	0.802	0.423	0.738	0.661	0.563	0.641	0.761	0.631	11.072
C1SE1EE	C1SE1LL	C1SE1O	C1SE1P	C1SE1Q	C1SE1R	0.029	0.996	0.011	0.816	0.430	0.514	0.660	0.699	0.794	0.778	0.646	11.048
Bottom 5 models																	
C1SE1A	C1SE1Z	C1SE1C	C1SE1D	C1SE1E	C1SE1PP	0.118	0.538	0.065	0.402	0.265	0.433	0.273	0.292	0.075	0.524	0.310	-20.048
C1SE1G	C1SE1Z	C1SE1C	C1SE1D	C1SE1E	C1SE1F	0.128	0.581	0.072	0.431	0.331	0.390	0.126	0.287	-0.069	0.722	0.298	-20.068
C1SE1G	C1SE1H	C1SE1C	C1SE1D	C1SE1E	C1SE1F	0.142	0.557	0.080	0.441	0.307	0.436	0.156	0.279	-0.028	0.737	0.315	-21.367
C1SE1G	C1SE1H	C1SE1C	C1SE1D	C1SE1E	C1SE1PP	0.136	0.517	0.076	0.432	0.183	0.605	0.399	0.185	0.220	0.376	0.328	-21.396
C1SE1G	C1SE1Z	C1SE1C	C1SE1D	C1SE1E	C1SE1PP	0.129	0.520	0.072	0.395	0.311	0.365	0.162	0.319	-0.023	0.637	0.295	-21.946

Note: The first six columns contain the MIDUS variable names. The selected model is shown in boldface. Abbreviations: Aut, Autonomy; Env, Environmental Mastery; Growth, Personal Growth; L, Standardized factor loading; Purp, Purpose in Life; Rel, Positive Relations with Others; Self, Self-Acceptance.

**TABLE 2** | The resulting 6-item version of psychological well-being measure (Model 2).

Variable name	Dimension	Item
C1SE1EE	Autonomy	I tend to worry about what other people think of me.
C1SE1LL	Mastery	I have been able to build a living environment and a lifestyle for myself that is much to my liking.
C1SE1GG	Growth	I gave up trying to make big improvements or changes in my life a long time ago.
C1SE1P	Relations	I often feel lonely because I have few close friends with whom to share my concerns.
C1SE1Q	Purpose	I don't have a good sense of what it is I'm trying to accomplish in life.
C1SE1R	Self-Acceptance	I feel like many of the people I know have gotten more out of life than I have.

Note: Response scale: 1 Strongly agree; 2 Somewhat agree; 3 A little Agree; 4 Neither agree or disagree; 5 A little disagree; 6 Somewhat disagree; 7 Strongly disagree. The item indicated with (R) needs to be reverse-coded.

**TABLE 3** | Factor loadings for the main and replication models.

Item	Dimension	Unstandardized	P	Low 95% CI	High 95% CI	Standardized	Low 95% CI	High 95% CI
Main model								
C1SE1EE	Autonomy	1.000	—	—	—	0.438	0.405	0.471
C1SE1LL	Mastery	0.856	0.000	0.766	0.947	0.515	0.484	0.545
C1SE1GG	Growth	1.245	0.000	1.125	1.366	0.612	0.585	0.639
C1SE1P	Relations	1.528	0.000	1.386	1.669	0.693	0.670	0.716
C1SE1Q	Purpose	1.600	0.000	1.457	1.743	0.789	0.770	0.808
C1SE1R	Self-Acceptance	1.668	0.000	1.521	1.815	0.785	0.766	0.805
Replication model								
RA1SF1EE	Autonomy	1.000	—	—	—	0.391	0.355	0.427
RA1SF1LL	Mastery	1.031	0.000	0.909	1.153	0.554	0.524	0.585
RA1SF1GG	Growth	1.286	0.000	1.137	1.434	0.583	0.553	0.612
RA1SF1P	Relations	1.690	0.000	1.505	1.875	0.668	0.642	0.693
RA1SF1Q	Purpose	1.806	0.000	1.614	1.998	0.785	0.764	0.805
RA1SF1R	Self-Acceptance	1.939	0.000	1.733	2.144	0.806	0.786	0.826

**TABLE 4** | Correlations with 7-item psychological well-being subscales and external variables.

Variable	r
Autonomy	0.560
Environmental Mastery	0.778
Personal Growth	0.678
Positive Relations with Others	0.705
Purpose in Life	0.740
Self-Acceptance	0.828
MIDUS Life Satisfaction	0.503
MIDUS Positive Affect	0.536
MIDUS Negative Affect	-0.523
Life Orientation Test: Optimism	0.477
Life Orientation Test: Pessimism	-0.544

Note: All correlations are significant at  $p < 0.001$ .

the model together with their confidence intervals. Table 4 shows that the total score of the short scale is moderately to strongly correlated with the original 7-item psychological well-being dimensions, providing additional support for the validity

of the new measure and indicating that all six domains are adequately represented in the short form. In addition, moderate to strong associations were observed with selected well-being indicators (life satisfaction, negative affect, positive affect, pessimism, and optimism) in the expected directions. Overall, these findings provide evidence of convergent validity for the new measure.

It is noteworthy that the model with the highest average factor loading (consisting of items C1SE1EE, C1SE1FF, C1SE1O, C1SE1P, C1SE1Q, and C1SE1R) showed substantially poorer fit than the selected model (RMSEA = 0.065, CFI = 0.982, SRMR = 0.024), although its coefficient omega was higher (0.842) than that of the selected model. Its overall performance ranking was 180th.

### 3.1 | Post Hoc Analysis: Measurement Invariance of the Short Scale

Measurement invariance (Schmitt and Kuljanin 2008) was evaluated in a multiple-group confirmatory factor analysis framework by sequentially testing configural, metric, and scalar invariance across gender (male vs. female), age (18–54, 55–64,

and  $\geq 65$ ), education (high school diploma or less vs. any higher degree), and race (White vs. non-White). Descriptives are presented in Table 5. Configural invariance served as the baseline model and was specified with the same factor structure in all groups but without equality constraints on factor loadings or item intercepts. Adequate fit of this model indicates that the construct is represented by the same pattern of relationships between items and the latent factor across groups. Metric invariance was then tested by constraining factor loadings to be equal across groups. Support for metric invariance indicates that the items relate to the latent construct to a comparable extent in each group. Scalar invariance was subsequently tested by additionally constraining item intercepts to equality across groups while retaining the loading constraints. Support for scalar invariance indicates that individuals with the same level of the latent construct are expected to have the same observed item scores regardless of group membership, thereby permitting meaningful comparisons of latent means across groups (Millsap 2012).

Invariance decisions were based on changes in global fit indices between nested models ( $\Delta\text{CFI} \geq -0.010$  and  $\Delta\text{RMSEA} \geq 0.015$  indicating non-invariance; Chen 2007). When full invariance was not supported, partial invariance was pursued. All models

**TABLE 5** | Descriptives related to measurement invariance groups.

Variable	Group	N	%
Gender	Male	1484	45.1
	Female	1810	54.9
Age	39–54	816	24.8
	55–64	982	29.8
	$\geq 65$	1496	45.4
Education	High school diploma or less	955	29.0
	Beyond high school diploma	2328	70.7
Race	White	2923	88.7
	Other	344	10.4

**TABLE 6** | Measurement invariance tests.

Variable	Model	Chi-Square	df	RMSEA	$\Delta\text{RMSEA}$	CFI	$\Delta\text{CFI}$	Eliminated constraints
Gender	Configural	33.844	18	0.025	—	0.997	—	—
	Full metric	36.022	23	0.020	-0.005	0.997	0.000	—
	Full scalar	70.679	28	0.032	0.012	0.992	-0.005	—
Education	Configural	34.019	18	0.025	—	0.997	—	—
	Full metric	83.168	23	0.042	0.017*	0.988	-0.009	—
	Partial metric	39.378	22	0.023	-0.002	0.997	0.000	Loading of C1SE1LL
	Partial scalar	45.991	25	0.024	-0.001	0.996	-0.001	Intercepts of C1SE1LL C1SE1GG
Age	Configural	59.112	27	0.035	—	0.994	—	—
	Full metric	115.902	37	0.047	0.012	0.984	-0.010*	—
	Partial metric	69.198	35	0.032	-0.003	0.993	-0.001	Loading of C1SE1LL
	Partial scalar	93.838	41	0.036	0.004	0.990	-0.003	Intercepts of C1SE1LL C1SE1GG
Race	Configural	34.540	18	0.025	—	0.997	—	—
	Full metric	40.756	23	0.023	-0.002	0.996	-0.001	—
	Full scalar	45.817	28	0.021	-0.002	0.996	0.000	—

Note: Thresholds used to indicate lack of invariance were:  $\Delta\text{CFI} \geq -0.010$ ,  $\Delta\text{RMSEA} \geq 0.015$  for metric and scalar invariance. Asterisks indicate  $\Delta$  values that exceeded the prespecified thresholds.

were estimated in Mplus. Results are presented in Table 6. Full metric and scalar invariance were supported across gender and race. Across age and education, partial invariance was established, with items C1SE1LL and C1SE1GG identified as non-invariant for both grouping variables.

Based on these results, future applications of the scale should consider the possibility of measurement noninvariance across demographic groups, particularly age and education. Researchers are encouraged to evaluate measurement invariance before conducting substantive group comparisons and, when noninvariance is present, to use structural equation modeling approaches that allow noninvariant parameters to vary across groups rather than assuming full comparability of observed scores.

### 3.2 | Replication in an Independent Dataset

To evaluate the generalizability of the final 6-item solution, its performance was examined in an independent MIDUS sample (MIDUS Refresher; 2011–2014) used as an out-of-sample test set. Specifically, the model was identified in the development sample and subsequently evaluated in this independent, previously unseen sample. This approach functionally approximates a holdout train-test validation procedure (Lantz 2023). Results indicated that the 6-item one-factor model retained very good fit and excellent reliability in the independent sample ( $N = 3577$ ; mean age = 50.5,  $SD = 14.38$ ; female = 51.9%), with  $\text{RMSEA} = 0.051$ ,  $\text{CFI} = 0.986$ ,  $\text{SRMR} = 0.019$ ,  $\omega = 0.804$  (95%  $\text{CI} = 0.792\text{--}0.816$ ), and an average standardized loading of 0.631. Table 3 reports the factor loadings for the model together with their confidence intervals.

## 4 | Discussion

This study sought to develop a 6-item short form of Ryff's psychological well-being measure that possesses strong psychometric properties and adequate construct representation.

This goal was achieved using an exhaustive search strategy. The resulting short form demonstrates exceptional psychometric properties and effectively represents all six dimensions of Ryff's model while maintaining unidimensionality, as confirmed in an independent sample. Item redundancy and excessive content overlap were minimized by design because the final scale retained only one item from each of Ryff's six domains, thereby reducing the likelihood that multiple items would assess the same narrow content. Consequently, any remaining inter-item associations are more appropriately interpreted as reflecting the intended general well-being factor and the inherently inter-related nature of the domains, rather than as evidence of duplicated item content.

The exhaustive search framework demonstrated its efficacy in systematically evaluating all possible item combinations while balancing multiple psychometric criteria. This method addresses the limitations of traditional approaches to scale reduction, which may rely on subjective decisions or overlook optimal configurations because they usually test only a few alternative models or use a single criterion for item selection. By incorporating model fit indices, standardized factor loadings, and omega reliability, and testing numerous alternative models, the framework ensures a comprehensive and objective assessment of construct representation, unidimensionality, and internal consistency. This efficiency opens new possibilities for researchers to explore extensive item combinations and optimize scale performance.

Ryff's original model conceptualizes psychological well-being as six related dimensions that can be examined separately when domain-level inferences are required (Ryff 1989). In the present study, unidimensionality was intentionally targeted because the aim was to develop an ultra-brief instrument suitable for contexts in which respondent burden is a primary constraint and a single summary index of general psychological well-being is needed. This short form can facilitate the inclusion of psychological well-being assessments in large-scale surveys, longitudinal studies, or clinical settings where time and resources are limited. To preserve alignment with Ryff's multidimensional framework while estimating a one-factor model, one item from each of the six domains was required, such that the resulting total score can be interpreted as a broad, general well-being indicator rather than as a substitute for the full multidimensional assessment. When the six dimensions are of substantive interest, the original long form or other multidimensional short forms should be used instead of the ultra-short unidimensional version.

#### 4.1 | Possible Adjustments and Extension

The analysis for this study was completed in approximately 1 h on a standard personal computer, indicating that the exhaustive method can be performed within a manageable timeframe for most applied uses in psychology on modern computing systems. However, as the number of factors and items increases, the number of models to test may become computationally demanding. In such cases, a random sample of item combinations could serve as a computationally efficient alternative to an exhaustive search. Additionally, optimization algorithms such as genetic algorithms (Mitchell 1998) or ant colony optimization (Leite et al. 2008) offer innovative approaches to navigate the vast search space of item combinations. These algorithms

efficiently reduce computational complexity by prioritizing promising combinations based on iterative evaluations, thereby focusing on high-performing models without the need to exhaustively evaluate every possible configuration. Moreover, the framework can flexibly incorporate other performance metrics for each model. For example, if correlations with external variables or between latent variables in a multidimensional model are of interest, they can be included as performance metrics. The method can accommodate any calculable metric for model evaluation. Finally, in this study, alternative criteria were given equal weight in determining model performance; however, in some research contexts, researchers may choose to use different weights per criterion.

#### 4.2 | Limitations

Some limitations should be considered when interpreting the findings of this study. First, the analyses were based on two independent adult samples from the MIDUS project. Although these are large and nationally representative samples, the psychometric properties of the identified short form may differ across other populations, cultural contexts, or age groups. Replication in independent and demographically diverse samples is necessary to establish the generalizability of the results. It should be acknowledged that the present findings were obtained exclusively in a United States sample, which is embedded in a predominantly individualistic sociocultural context. Consequently, the validity, reliability, and measurement properties of the ultra-short PWB index should not be assumed to generalize to populations in culturally dissimilar settings, where the meaning, salience, and normative expression of well-being may differ (Joshani et al. 2021). Cross-cultural validation studies are therefore warranted, including applications in collectivistic and non-Western contexts, to evaluate the replicability of the factor structure, internal consistency, and convergent validity, as well as the extent of measurement invariance across cultural groups and languages.

Additionally, the use of a unidimensional measure necessarily simplifies the complex, multidimensional nature of psychological well-being, as originally conceptualized in Ryff's model. This trade-off is common in the development of short forms. Nonetheless, the demand for short instruments remains high, particularly in research contexts with time or space constraints, such as experience sampling designs and large-scale national or international surveys (Rammstedt and Beierlein 2014). In such settings, a short scale that yields a total score of psychological well-being is often more practical. Short forms reduce respondent burden and administrative costs, and they can enhance research efficiency by improving response rates and reducing item nonresponse, thereby increasing statistical power (Franke et al. 2013).

Although the selected six-item solution showed essentially identical performance in an independent MIDUS sample, it is still important to note that some sample dependency can occur when a very large number of candidate models is evaluated. Under these conditions, the exact rank ordering of the best-performing solutions can be partly shaped by the specific sample used. A different sample could have produced a slightly different top set of items. The successful replication supports

the trustworthiness of the proposed short form, but it does not entirely rule out the possibility of overfitting to the training sample. This limitation is unavoidable and should be acknowledged with appropriate caution. The short form introduced here fulfills its intended purpose, even with the understanding that a different short form could have emerged had another sample been used.

Some additional practical limitations should be considered when the ultra-short PWB index is applied in clinical or other applied settings. Ryff's psychological well-being measure was not designed as a clinical assessment instrument, and this limitation necessarily extends to the present short form. The short form is intended to provide a brief global indicator of psychological well-being rather than a comprehensive evaluation for clinical decision-making. Accordingly, it should not be used as a substitute for diagnostic interviewing, disorder-specific symptom measures, or full-length multidimensional well-being batteries. Moreover, because the instrument includes only one item per Ryff domain, domain-level interpretation is not supported and should be avoided. Applied use should also be evaluated in light of the measurement invariance findings reported in the present study. Full invariance was not supported across age and education groups, with non-invariance concentrated in two items (C1SE1LL and C1SE1GG). This pattern indicates that observed mean differences may partially reflect differential item functioning rather than true differences in the underlying construct, particularly for comparisons across age strata or educational levels. Therefore, interpretations that depend on score comparability across these groups (e.g., benchmarking individuals against subgroup norms) should be made cautiously unless non-invariance is addressed statistically (e.g., via partial invariance modeling within a structural equation modeling framework).

## 5 | Conclusion

Methodologically, this study demonstrates and highlights the value of computational approaches in advancing psychological measurement. By using accessible computing resources and applying rigorous statistical techniques, researchers can enhance the validity, reliability, and objectivity of scale development processes. Substantively, the study contributed to psychological research by developing and validating a brief, psychometrically robust unidimensional measure of psychological well-being, based on large, nationally representative samples. The resulting short form provides a practical tool for use in time-constrained or large-scale research contexts, while maintaining adequate reliability and construct validity. Although it simplifies the complexity of Ryff's original multidimensional model, it meets the growing demand for efficient assessment tools in psychological research. Future studies should assess the generalizability of this measure across culturally and demographically diverse populations and examine its performance in applied settings.

## Funding

The author received no specific funding for this work.

## Ethics Statement

MIDUS data collection is reviewed and approved by the Education and Social/Behavioral Sciences and the Health Sciences IRBs at the University of Wisconsin-Madison (<https://pmc.ncbi.nlm.nih.gov/articles/PMC4280664/#:~:text=Ethics,the%20University%20of%20Wisconsin%2DMadison>).

## Consent

All participants provided informed consent before participating in the study.

## Conflicts of Interest

The author declares no conflicts of interest.

## Data Availability Statement

The raw MIDUS data and materials can be obtained at <https://www.midus.wisc.edu/data/index.php>. The R script and the data file containing the results for all models are available at [https://osf.io/4pgkb/?view\\_only=b73e2e8420044d5cba76763e40910fdd](https://osf.io/4pgkb/?view_only=b73e2e8420044d5cba76763e40910fdd).

## References

- Browne, M., M. Rockloff, and V. Rawat. 2018. "An SEM Algorithm for Scale Reduction Incorporating Evaluation of Multiple Psychometric Criteria." *Sociological Methods & Research* 47, no. 4: 812–836. <https://doi.org/10.1177/0049124116661580>.
- Chen, F. F. 2007. "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Structural Equation Modeling: A Multidisciplinary Journal* 14, no. 3: 464–504. <https://doi.org/10.1080/1070510701301834>.
- Diener, E., D. Wirtz, W. Tov, et al. 2009. "New Well-Being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings." *Social Indicators Research* 97, no. 2: 143–156. <https://doi.org/10.1007/s11205-009-9493-y>.
- Dörendahl, J., and S. Greiff. 2020. "Are the Machines Taking Over? Benefits and Challenges of Using Algorithms in (Short) Scale Construction." *European Journal of Psychological Assessment* 36, no. 2: 217–219. <https://doi.org/10.1027/1015-5759/a000597>.
- Fabrigar, L. R., and D. T. Wegener. 2012. *Exploratory Factor Analysis*. Oxford University Press.
- Flora, D. B. 2020. "Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates." *Advances in Methods and Practices in Psychological Science* 3, no. 4: 484–501. <https://doi.org/10.1177/2515245920951747>.
- Franke, G. R., A. Rapp, and J. "Mick" Andzulis. 2013. "Using Shortened Scales in Sales Research: Risks, Benefits, and Strategies." *Journal of Personal Selling & Sales Management* 33, no. 3: 319–328. <https://doi.org/10.2753/pss0885-3134330306>.
- Garcia, D., M. Kazemitabar, and M. H. Asgarabad. 2023. "The 18-item Swedish Version of Ryff's Psychological Wellbeing Scale: Psychometric Properties Based on Classical Test Theory and Item Response Theory." *Frontiers in Psychology* 14: 1208300. <https://doi.org/10.3389/fpsyg.2023.1208300>.
- Gonzalez, O. 2020. "Psychometric and Machine Learning Approaches to Reduce the Length of Scales." *Multivariate Behavioral Research* 56, no. 6: 903–919. <https://doi.org/10.1080/00273171.2020.1781585>.
- Gorsuch, R. L. 1990. "Common Factor Analysis Versus Component Analysis: Some Well and Little Known Facts." *Multivariate Behavioral Research* 25, no. 1: 33–39. [https://doi.org/10.1207/s15327906mbr2501\\_3](https://doi.org/10.1207/s15327906mbr2501_3).
- Irwing, P., and D. J. Hughes. 2018. "Test Development." In *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on*

- Survey, Scale and Test Development, edited by P. Irwing, T. Booth and D. J. Hughes, 3–47. Wiley Blackwell.
- Joshanloo, M., E. Van de Vliert, and P. E. Jose. 2021. “Four Fundamental Distinctions in Conceptions of Wellbeing Across Cultures.” In *The Palgrave Handbook of Positive Education*, edited by M. L. Kern and M. L. Wehmeyer, 675–703. Springer International Publishing.
- Kline, R. B. 2023. *Principles and Practice of Structural Equation Modeling*. Guilford Publications.
- Lantz, B. 2023. *Machine Learning With R: Learn Techniques for Building and Improving Machine Learning Models, From Data Preparation to Model Tuning, Evaluation, and Working With Big Data*. Packt Publishing Ltd.
- Leite, W. L., I. C. Huang, and G. A. Marcoulides. 2008. “Item Selection for the Development of Short Forms of Scales Using an Ant Colony Optimization Algorithm.” *Multivariate Behavioral Research* 43, no. 3: 411–431. <https://doi.org/10.1080/00273170802285743>.
- Margolis, S., E. Schwitzgebel, D. Ozer, R. L. Martinez, and S. Lyubomirsky. 2022. “Creating Eudaimonic and Non-Eudaimonic Goods in the Pursuit of the Good Life: The Riverside Eudaimonia Scale and the Rich & Sexy Well-Being Scale.” *International Journal of Wellbeing* 12, no. 1: 1–20. <https://doi.org/10.5502/ijw.v12i1.1575>.
- Millsap, R. E. 2012. *Statistical Approaches to Measurement Invariance*. Routledge.
- Mitchell, M. 1998. *An Introduction to Genetic Algorithms*. MIT Press.
- Morin, A. J. S., N. D. Myers, and S. Lee. 2020. “Modern Factor Analytic Techniques: Bifactor Models, Exploratory Structural Equation Modeling (ESEM), and Bifactor-ESEM.” In *Handbook of Sport Psychology: Exercise, Methodologies, & Special Topics*, edited by G. Tenenbaum, R. C. Eklund and N. Boiangin, 1044–1073. John Wiley & Sons, Inc.
- Nandakumar, R. 1993. “Assessing Essential Unidimensionality of Real Data.” *Applied Psychological Measurement* 17, no. 1: 29–38. <https://doi.org/10.1177/014662169301700108>.
- Rajwar, K., K. Deep, and S. Das. 2023. “An Exhaustive Review of the Metaheuristic Algorithms for Search and Optimization: Taxonomy, Applications, and Open Challenges.” *Artificial Intelligence Review* 56, no. 11: 13187–13257. <https://doi.org/10.1007/s10462-023-10470-y>.
- Rammstedt, B., and C. Beierlein. 2014. “Can’t We Make It Any Shorter?” *Journal of Individual Differences* 35, no. 4: 212–220. <https://doi.org/10.1027/1614-0001/a000141>.
- Reeve, B. B., R. D. Hays, J. B. Bjorner, et al. PROMIS Cooperative Group. 2007. “Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS).” *Medical Care* 45, no. 5: S22–S31. <https://doi.org/10.1097/01.mlr.0000250483.85507.04>.
- Revicki, D. A., W.-H. Chen, and C. Tucker. 2015. “Developing Item Banks for Patient-Reported Health Outcomes.” In *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*, edited by S. P. Reise and D. A. Revicki, 334–363. Routledge/Taylor & Francis Group.
- 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, no. 6: 1069–1081. <https://doi.org/10.1037/0022-3514.57.6.1069>.
- Schmitt, N., and G. Kuljanin. 2008. “Measurement Invariance: Review of Practice and Implications.” *Human Resource Management Review* 18, no. 4: 210–222. <https://doi.org/10.1016/j.hrmr.2008.03.003>.
- Smith, G. T., J. L. Combs, and C. M. Pearson. 2012. “Brief Instruments and Short Forms.” In *APA Handbook of Research Methods in Psychology: Foundations, Planning, Measures, and Psychometrics*, edited by H. Cooper, P. M. Camic and D. L. Long, et al., 393–409. American Psychological Association.
- Strauss, M. E., and G. T. Smith. 2009. “Construct Validity: Advances in Theory and Methodology.” *Annual Review of Clinical Psychology* 5: 1–25. <https://doi.org/10.1146/annurev.clinpsy.032408.153639>.
- Vittersø, J. 2016. “The Most Important Idea in the World: An Introduction.” In *International Handbooks of Quality-of-Life*, 1–24. Springer International Publishing.
- Watkins, M. W. 2021. *A Step-by-Step Guide to Exploratory Factor Analysis With SPSS*. Routledge.
- Whitely, S. E. 1983. “Construct Validity: Construct Representation Versus Nomothetic Span.” *Psychological Bulletin* 93, no. 1: 179–197. <https://doi.org/10.1037/0033-2909.93.1.179>.
- Widaman, K. F., T. D. Little, K. J. Preacher, and G. M. Sawalani. 2011. “On Creating and Using Short Forms of Scales in Secondary Research.” In *Secondary Data Analysis: An Introduction for Psychologists*, edited by K. H. Trzesniewski, M. B. Donnellan and R. E. Lucas, 39–61. American Psychological Association.
- Ziegler, M., C. J. Kemper, and P. Kruey. 2014. “Short Scales—Five Misunderstandings and Ways to Overcome Them.” *Journal of Individual Differences* 35, no. 4: 185–189. <https://doi.org/10.1027/1614-0001/a000148>.