Disentangling Gratitude: A Theoretical and Psychometric Examination of the Gratitude Resentment and Appreciation Test- Revised Short (GRAT-RS)

Current APA-Style Citation:

Hammer, J. H., & Brenner, R. E. (2017). Disentangling gratitude: A theoretical and psychometric examination of the Gratitude Resentment and Appreciation Test- Revised Short (GRAT-RS). *Journal of Personality Assessment, 14,* 1-10. doi: 10.1080/00223891.2017.1344986

Joseph H. Hammer, Ph.D.
243 Dickey Hall
University of Kentucky
Lexington, KY 40506
(847) 809-8785
jhhammer@gmail.com

Rachel E. Brenner, M.S.
W112 Lagomarcino Hall
Department of Psychology
Iowa State University
Ames, IA 50011
(407) 506-5225
rbrenner@iastate.edu

Abstract

The current study extended our theoretical and applied understanding of gratitude through a psychometric examination of the most popular multidimensional measure of gratitude, the Gratitude, Resentment, and Appreciation Test – Revised Short form (GRAT-RS). Namely, the dimensionality of the GRAT-RS, the model-based reliability of the GRAT-RS total score and three subscale scores, and the incremental evidence of validity for its latent factors were assessed. Dimensionality measures (e.g., Explained Common Variance) and CFA results with 426 community adults indicated that the GRAT-RS conformed to a multidimensional (bifactor) structure. Model-based reliability measures (e.g., Omega Hierarchical) provided support for the future use of the Lack of a Sense of Deprivation raw subscale score, but not for the raw GRAT-RS total score, Simple Appreciation subscale score, or Appreciation of Others subscale score. Structural equation modeling results indicated that only the general gratitude factor and the Lack of a Sense of Deprivation specific factor accounted for significant variance in life satisfaction, positive affect, and distress. These findings support the *three pillars of gratitude* conceptualization of gratitude over competing conceptualizations, the position that the specific forms of gratitude are theoretically distinct, and the argument that appreciation is distinct from the superordinate construct of gratitude.

Keywords: Gratitude, factor analysis, reliability, validity, bifactor analysis, scale development, psychometrics

Disentangling Gratitude: A Theoretical and Psychometric Examination of the Gratitude Resentment and Appreciation Test- Revised Short (GRAT-RS)

“We can only be said to be alive in those moments when our hearts are conscious of our treasures.”

-Thornton Wilder (2007, p.149)

Gratitude is a general disposition towards acknowledging and appreciating the positive in the world (Wood, Froh, & Geraghty, 2010). Distinguished from future-focused emotions, such as optimism and hope, feelings of gratitude are focused on the present moment and without expectation for the future. The extant literature underscores the benefits of gratitude, which has been consistently linked to greater well-being, life satisfaction, self-esteem, positive affect, and healthier relationship functioning (Bernstein & Simmons, 1974; see Wood et al., 2010, for a review). Gratitude may also hold clinical relevance, having demonstrated associations with reduced depressive symptomology (Disabato et al., 2016; Fredrickson, Tugade, Waugh, & Larkin, 2003; Wood, Maltby, Gillet, Linley, & Joseph, 2008; Van Dusen, 2014), suicidal ideation and past suicide attempts (Li, Zhang, Li, Li, & Ye, 2012), and post-traumatic symptoms (Van Dusen, Tiamiyu, Kashdan, & Elhi, 2015). Moreover, longitudinal studies suggest a causal relationship wherein gratitude predicts depressive symptoms over time (e.g., Disbato, Kashdan, Short, Jarden, 2016; Van Dusen, 2014; Wood, Maltby, Gillet, et al., 2008).

There is ongoing debate about the theoretical structure of gratitude. Some scholars have argued that gratitude is best conceptualized as a unitary construct (McCullough, et al. 2004; Wood, Maltby, Stewart, et al., 2008). In factor analytic terms, gratitude would be operationalized with a unidimensional model. Others suggest that gratitude manifests itself not as a singular construct, but as independent-yet-related forms of gratitude (Diessner & Lewis, 2007; Thomas &

Watkin, 2003). This corresponds to an oblique (i.e., correlated, common factors) model. In turn, the *three pillars of gratitude* conceptualization suggests that gratitude simultaneously exists in three specific forms but also as a superordinate construct (Watkins, 2009, as cited in Watkins, 2014; Watkins, Woodward, Stone, & Kolts, 2003). The superordinate gratitude construct can be conceptualized as a general factor that exists independently of its specific forms (i.e., operationalized by a bifactor model) or as a higher-order factor that is defined by the common element that runs through all three specific forms (i.e., operationalized by second-order model). A compelling psychometric rationale for preferring one of these four theoretical structures over the others remains unarticulated. In addition, embedded within this debate is the question of whether or not appreciation (sometimes considered a specific form of gratitude) is distinct from the superordinate construct of gratitude (Fagley, 2012; Manela, 2012; Wood, Maltby, Stewart, et al., 2008). Beyond the appreciation debate, there is also continued disagreement on whether specific forms of gratitude are theoretically distinct (Wood, Maltby, Stewart, et al., 2008) and/or provide incremental clinical utility beyond the general gratitude factor (Fagley, 2012).

Intertwined with these questions regarding gratitude theory and clinical utility are questions about the reliability and validity of gratitude assessments. One of the most widely used multidimensional measures of gratitude is the Gratitude Resentment and Appreciation Test-Revised Short (GRAT-RS; Thomas & Watkins, 2003). The GRAT-RS was designed to operationalize the three pillars of gratitude (Thomas & Watkin, 2003; Watkins et al., 2003, Watkins, 2009) via three GRAT-RS subscales. Lack of a Sense of Deprivation (LOSD) purports to measure the absence of the belief that life has treated oneself unfairly. Simple Appreciation (SA) purports to measure the appreciation of simple day-to-day pleasures. Appreciation of Others (AO) purports to measure the recognition of the importance of appreciating and expressing appreciation of the contribution that others have in their lives. The developers contend that the GRAT-RS subscale scores reliably and validly measure their respective forms of gratitude, while the GRAT-RS total score reliably and validly measures the superordinate gratitude construct (Thomas & Watkins, 2003). However, moderate to strong correlations observed between the three GRAT-RS subscales, consistent with the theoretical debate regarding the extent to which facets of gratitude are distinct, indicate that the dimensionality of the GRAT-RS, the model-based reliability of its total and subscale scores, and incremental evidence of validity for its latent factors require additional investigation through a bifactor analysis lens.

In summary, there are outstanding questions about the theoretical structure of gratitude, the distinctness of appreciation from gratitude, the incremental clinical utility of the specific forms of gratitude, and the dimensionality, reliability, and validity of the GRAT-RS scores. Therefore, using promising but under-utilized methods of bifactor analysis and ancillary bifactor measures, the current study sought to address these theoretical and applied questions.

Dimensionality of the GRAT-RS

Thomas and Watkins (2003) derived the 16-item GRAT-RS from the original three subscales of the 44-item Gratitude Resentment and Appreciation Test (GRAT; Watkins et al., 2003). Froh and colleagues (2011) subjected the GRAT-RS to an exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) using a sample of 681 college-enrolled adults. After dropping one item, Froh and colleagues concluded the results supported the anticipated three-factor oblique structure. Subsequently, Froh and colleagues conducted a CFA using the 15-item version of the GRAT-RS in a sample of a youth ranging from age 10 to 19 (N = 1,405), concluding the GRAT-RS demonstrated the same dimensionality as in the adult sample. Unidimensional, second-order, or bifactor solutions were not explored in either sample.

To date, no published studies have subjected the GRAT-RS to an oblique CFA in an adult sample independent from the EFA sample. As noted by Hammer and Toland (2016), “It is important to use an independent sample to confirm an instrument’s internal structure (Worthington & Whittaker, 2006) because it is possible for an initial dataset to be idiosyncratic, leading to the identification of an initial factor structure that does not generalize well to subsequent samples (i.e., a lack of structural generalizability)” (p. 2). The present study conducts an independent sample CFA examining the dimensionality of the GRAT-RS among adults.

Furthermore, the GRAT-RS has not been subjected to a bifactor CFA nor a second-order (i.e., higher order) CFA. The GRAT-RS subscale scores have demonstrated moderate to high intercorrelations (i.e., r = .25 to .83 [Magno & Orillosa, 2012]). This shared variance among the subscales suggests that the GRAT-RS may be defined by a strong general factor that is distinct from the GRAT-RS subscale factors. A second-order factor model, however, would specify three first-order factors (LOSD, SA, and AO) and one second-order gratitude factor that is defined by the shared variance of LOSD, SA, and AO. In other words, the second-order gratitude factor only measures what the three pillars of gratitude share in common, rather than merely being (a) an amalgam of the entire content domains of the three forms of gratitude or (b) an independent general gratitude factor distinct from the three specific forms.

In contrast, a bifactor model specifies one general gratitude factor reflecting the common variance of all items as well as three separate, narrower specific factors (LOSD, SA, and AO) reflecting the variance of their assigned items. That is, each item simultaneously loads on the general factor and its assigned specific factor. The general and specific factors are first-order factors set orthogonal to each other. Thus, the three specific factors account for variance in their respective items after partialling out the variance accounted for by the general factor.

A bifactor CFA and ancillary bifactor measures (see Hammer & Toland, 2016, November, for a video walkthrough) allow researchers to answer unresolved questions about the dimensionality and model-based reliability of instrument scores (Reise, Bonifay, & Haviland, 2013). However, the possibility of a second-order or bifactor solution for the GRAT-RS has not been investigated. The present study examined the dimensionality of the GRAT-RS by testing four alternative models (e.g., unidimensional, three-factor oblique, second-order, bifactor) and examining ancillary bifactor measures pertaining to dimensionality. Determining which of these models best fits the GRAT-RS data will help provide a psychometric rationale for preferring one theoretical structure over the rest. If a bifactor solution provides the best fit, this would also provide an initial affirmative answer to the question of whether appreciation (operationalized in two forms via the SA and AO specific factors) is distinct from the general factor of gratitude.

In addition to refining our theoretical understanding of gratitude, these results can clarify the dimensionality of a prominent gratitude assessment, the GRAT-RS. Dimensionality evidence (e.g., Explained Common Variance) has a direct bearing on how users can permissibly model and score the GRAT-RS. For example, even when a unidimensional measurement model provides a poor fit for a given instrument, under certain circumstances it remains permissible for test users to model the instrument using a unidimensional solution (Rodriguez, Reise, & Haviland, 2016a). In practice, some researchers or clinicians may wish to treat the GRAT-RS as unidimensional for the purposes of scoring, but this may result in model misspecification that can lead to factor loadings that are too high or low. In turn, this can lead to inaccurate conclusions regarding group differences on gratitude or how strongly gratitude correlates with other constructs (Rodriguez, et al., 2016a). Thus, the degree of Average Relative Measurement Parameter Bias (ARMPB) associated with forcing the GRAT-RS into a unidimensional solution was also calculated in the present study. The presence of significant ARMPB would suggest the possibility of inaccurate parameter estimates and, in turn, indicate that researchers and clinicians should avoid operationalizing gratitude via a unidimensional solution for the GRAT-RS.

Reliability of the Raw GRAT-RS Scores

Cronbach's alpha (α) estimates for the GRAT-RS total (i.e., composite) score have ranged from .86 to .92 among adults (Spangler, 2010; Thomas & Watkins, 2003). Estimates for the three subscales scores have ranged from .80 (Lack of a Sense of Deprivation), .76 to .87 (Simple Appreciation), and .75 to .76 (Appreciation of Others) among adults (Diessner & Lewis, 2007; Diessner, Solom, Frost, Parsons, & Davidson, 2008). However, if the GRAT-RS best conforms to a bifactor structure, this would indicate that the raw scores' internal consistency estimates are dictated by both general and specific sources of common variance (Rodriguez et al., 2016a). In other words, the GRAT-RS total score's alpha coefficient may be artificially inflated by reliable variance from the specific factors, and the GRAT-RS subscale scores' alpha coefficients may be artificially inflated by reliable variance from the general factor. Therefore, model-based reliability measures (e.g., Omega Hierarchical) are needed to determine whether the raw GRAT-RS total score and subscale scores are reliable measures of their intended constructs. This complies with the *Standards for Educational and Psychological Testing,* which state that it is only appropriate to use scores—particularly subscale scores—that demonstrate evidence of "distinctiveness and reliability" (*The Standards*; Standard 1.14; American Education Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 27). The present study used model-based reliability measures to determine whether it may be appropriate to use the raw total and/or subscale scores of the GRAT-RS in future research and clinical practice. The finding that certain scores are unreliable indicators of their intended gratitude construct would have significant implications for the clinical utility of gratitude assessment with the GRAT-RS.

Evidence of Validity for the GRAT-RS Factors

Regarding convergent evidence of validity, the total GRAT-RS score shows moderate to strong associations with other established measures of gratitude (Froh et al., 2011; Solom, Watkins, McCurrah, & Scheibe, 2016; Toussaint & Friedman, 2009; Van Dusen, 2014), including the original GRAT (i.e., $r = .95$ [Toussaint & Friedman, 2009], $r = .97$ [Watkins, 2014]). The GRAT-RS total score has also demonstrated concurrent evidence of validity through correlations with theoretically-linked criterion variables. For example, broaden and build theory asserts that gratitude should be linked to action tendencies that promote social engagement (Fredrickson, 2004), and indeed the GRAT-RS total score has been positively associated with interpersonal forgiveness and inversely associated with interpersonal revenge and avoidance (Toussaint & Friedman, 2009). The GRAT-RS has also been linked with greater life satisfaction, psychological well-being, happiness, positive affect, spiritual transcendence, and self-esteem (Diessner & Lewis, 2007; Solom et al., 2017; Van Dusen, 2014; Toussaint & Friedman, 2009; Watkins, Cruz, Holben, & Kolts, 2008), and inversely linked with negative affect, thought suppression, envy, cynicism, material values, pathological narcissism, and ruminating about sad memories (Diessner & Lewis, 2007; Magno & Orillosa, 2012; Solom et al., 2017; Watkins et al., 2008). Important to clinical contexts, the GRAT-RS has demonstrated predictive validity through a longitudinal relationship with depression. Specifically, in one study, the Time 1 GRAT-RS total score significantly predicted depression six weeks later (Time 2), controlling for Time 1 levels of self-reported depression, life satisfaction, positive and negative affect, and Time

2 levels of self-reported academic stress (Van Dusen, 2014). This relationship remained significant when life satisfaction at Time 2 was added to the model.

Researchers have stated that the GRAT-RS subscale scores demonstrated convergent evidence of validity via correlations with theoretically-linked criterion variables (e.g., life satisfaction, depression, positive affect, negative affect, materialism, spiritual transcendence, achievement emotions; Diessner & Lewis, 2007; Froh et al., 2011; Magno & Orillosa, 2012; Thomas & Watkins, 2003; Toussaint & Friedman, 2009). However, incremental convergent evidence of validity for the GRAT-RS factors has not been reported. The present investigation used structural equation modeling (SEM) to examine incremental convergent evidence of validity for the general gratitude factor and the three specific factors (LOSD, SA, and AO). This can provide an initial answer to the question of whether specific forms of gratitude—particularly those measuring appreciation (SA and AO)—have incremental clinical utility beyond the general gratitude factor (Fagley, 2012). The finding that certain factors fail to account for unique variance in key outcomes would hold implications for the theoretical and clinical utility of these gratitude factors. For example, if AO was uniquely linked to well-being beyond general gratitude, clinicians utilizing gratitude may consider supplemental interventions aimed toward increasing expression of appreciation to others. Factors failing to demonstrate added value may require enhanced scrutiny in future research, and possibly theoretical revision.

The Present Study

In summary, this study had three primary goals. The first goal of the present study was to examine the theoretical structure of gratitude through the comparison of four models. A unidimensional model will correspond the conceptualization of gratitude as a unitary construct (e.g., Wood, Maltby, Stewart, et al., 2008). An oblique model comprised of three latent correlated factors will correspond to the view of gratitude as comprised of three independent-yet-related forms of gratitude (e.g., Diessner & Lewis, 2007). In turn, the three pillars of gratitude conceptualization, which suggests that gratitude simultaneously exists in three specific forms but also as a superordinate construct, will be operationalized by a higher-order model and bifactor model. A higher-order model considers gratitude as a common element (i.e., second-order factor) that runs through all three first-order factors. A bifactor approach will model the superordinate gratitude construct as a general factor that exists independently of its specific forms (i.e., operationalized by a bifactor model). Given the moderate to strong correlations between the three GRAT-RS subscales reported in the literature (Magno & Orillosa, 2012), and the unique substantive value argued for specific gratitude facets (Fagley, 2012), we anticipated that a bifactor structure would best account for the item covariation for the GRAT-RS. Assuming the GRAT-RS conforms to a bifactor structure, ancillary bifactor measures pertaining to dimensionality would be examined to determine whether the GRAT-RS should be viewed as having a primarily unidimensional or primarily multidimensional structure and, in turn, inform our theoretical understanding of gratitude as a unidimensional or multifaceted construct.

The second goal, assuming the GRAT-RS conforms to a bifactor structure, was to use model-based reliability estimates to determine whether it may be appropriate to use the raw GRAT-RS total score and/or any of the three raw subscale scores. Past research determined that some multidimensional instruments should only be scored as univocal measures (e.g., Brewster, Hammer, Sawyer, Eklund, & Palamar, 2016; Brouwer, Meijer, Zevalkink, 2012; Hammer & Toland, 2016), whereas others should only have certain subscales scored while the total score and remaining subscales scores should be abandoned (e.g., Li, Toland, & Usher, 2016). Given extant research on the GRAT-RS' mixed-strength interfactor correlations, it was anticipated that model-based reliability estimates would support the use of some raw scores but not others.

Third, assuming the GRAT-RS conforms to a bifactor structure, SEM would then be used to examine incremental convergent evidence of validity. The gratitude literature consistently suggests that gratitude is linked to greater life satisfaction and positive affect (see Wood et al., 2010, for a review) as well as lower distress (Disabato et al., 2016; Fredrickson, Tugade, Waugh, & Larkin, 2003; Wood, Maltby, Gillet et al., 2008; Van Dusen, 2014). Therefore, we utilized measures of these constructs to examine incremental convergent evidence of validity for the GRAT-RS general and specific factors. Given the inclusion of two appreciation-based specific factors (i.e., SA and AO), findings would contribute to the burgeoning line of research assessing the incremental utility of appreciation over and above gratitude (Fagley, 2012).

Method

Participants and Procedure

Participants were 426 adults (300 women, 124 men) who completed the online survey study through Mechanical Turk (MTurk). Results from research examining the viability of samples recruited through MTurk suggests that the data is as reliable as data obtained through traditional methods (Buhrmester, Kwang, & Gosling, 2011; Goodman, Cryder, & Cheema, 2012; Paolacci, Chandler, Ipeirotis, 2010; Shapiro, Chandler, & Mueller, 2013). Participants were informed that participation was voluntary and received $0.10 for their participation. Interested participants were directed to an online survey that began with an informed consent page, followed by the instrument battery and demographic items, and ended with a debriefing page. Approval was obtained for this project from Iowa State University's Office of Responsible Research.

Participants ranged in age from 18 to 74 years old (*M* = 36.7, *SD* = 12.7, *Mdn* = 33). Approximately 79.3% of the sample identified as White, 6.6% as African American/Black, 5.4% as Latino/a, 3.3% multiracial, 3.3% Asian American or Pacific Islander, and 1.4% American Indian/Native American, 0.7% other race/ethnicity; 1% preferred not to answer. Approximately 88.3% of participants identified as heterosexual, 6.6% bisexual, 2.5% lesbian, 0.7% questioning, 0.7% asexual, 0.2% gay, and 0.8% chose to self-identify with a text response (e.g., pansexual). Approximately 50% reported being married or in a committed relationship or civil union, 31% single, 17% separated or divorced, 2% widowed, and 1% preferred not to answer. Approximately 10.6% reported having some high school education or a high school diploma, 7.0% earned a Technical and Further Education (TAFE) degree, 62.4% had college experience or earned a four-year college degree, and 19.9% earned a graduate or professional degree.

Measures

Gratitude. The GRAT-RS is designed to measure gratitude. Items are rated on a 1 (*strongly disagree*) to 9 (*strongly agree*) Likert-type scale. Scores were derived by averaging individual item scores. Higher scores reflect higher levels of reported gratitude. Extant evidence for the psychometric properties of the GRAT-RS was discussed in the introduction. In the present study, the internal consistency estimates (α) were as follows: GRAT-RS = .82, 95% CI [.80, .85], LOSD = .86, 95% CI [.84, .88], SA = .85, 95% CI [.83, .87], AO = .84, 95% CI [.82, .86]. The means (standard deviations) for the scores were as follows: GRAT-RS = 6.68 (1.20), LOSD = 5.89 (1.77), SA = 7.21 (1.37), AO = 7.06 (1.47).

Distress. The *K6* (Kessler et al., 2002) is a 6-item scale measure of non-specific psychological distress developed for use in the U.S. National Health Interview Survey. Participants read the sentence stem, “During the past 30 days, about how often did you feel…”

and rate items such as “nervous” and “hopeless” on a 5-point Likert scale from 0 (*All the time)* to 4 (*None of the time).* The six item scores are summed to calculate a total score, with higher scores indicating greater tendency toward mental illness. The K6 has been linked to blinded clinical diagnoses in general U.S. population samples (Kessler et al., 2003), and self-report ratings of severe mental illness in samples from 14 countries (Kessler et al., 2010). Internal consistency estimates have been reported for undergraduate (α = .84; Lannin, Vogel, Brenner, Abraham, & Tucker, 2016) and community adult samples (.89 < α’s < .92; Kessler et al., 2002). In the current study, the K6 yielded an internal consistency estimate of .89, 95% CI = [87, .90] and a mean (standard deviation) of 2.31 (.90).

Positive affect. The positive affect subscale from the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) was used to measure positive affect over the past week. The 10-item positive affect subscale (PANAS PA) consists of items, such as “excited” and “strong” (Watson et al., 1988, p. 1070). Participants indicate the extent to they experienced each emotion over the past week on a 5-point Likert-type scale ranging from 1 (*Very slightly or not at all*) to 5 (*Extremely*). The items are averaged to create a total score, with higher scores indicating more positive affect. Previous studies produced concurrent evidence of validity for the PANAS PA through associations with other measures of positive and negative mood (Watson et al., 1998), as well as happiness in undergraduate and adult community samples (Wei et al., 2011). The PANAS PA yielded an internal consistency estimate of reliability of .89 in an adult community sample (Wei et al., 2011) and .89 in the current study, 95% CI = [.88, .91] and a mean (standard deviation) of 3.30 (.79).

Life satisfaction. The Satisfaction with Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) is a five-item measure of life satisfaction. Participants indicate the extent to which the items reflect how they view their lives using a 7-point Likert scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*). An example item is "In most ways, my life is close to my ideal." The items are averaged to create a total score, with higher scores indicating greater life satisfaction. The SWLS has demonstrated positive correlations with interviewer ratings of one-hour interviews of life satisfaction (Deiner et al., 1985) and self-reports of happiness (Wei et al., 2011). The SWLS has demonstrated internal consistency in community adults ($\alpha = .90$; Wei, Liao, & Shaffer, 2011) and yielded test-retest reliabilities of .84 for two-week and 1-month intervals across undergraduate students (Pavot, Diener, Colvin, & Sandvik, 1991). In the current study, the SWLS demonstrated an internal consistency of .92, 95% CI = [.87, .90] and a mean (standard deviation) of 4.15 (1.57).

Quality check. The survey contained two instructed response items (e.g., respond with 'strongly disagree' for this item'). The survey included a yes/no self-reported single-item indicator known as the *SRSI Use Me* ("In your honest opinion, should we use your data?," Meade & Craig, 2012, p. 16). The *SRSI Use Me* has been highly correlated with number of bogus items missed and provides a clear cutoff point (Meade & Craig, 2012).

Results

Data Cleaning

The initial dataset contained 487 individuals. Cases with any missing data ($n = 0$), incorrect responses to one or both instructed response items ($n = 28$), with a duplicate person response ($n = 26$) indicated by participants' anonymous MTurk Worker ID, and with responses of "no" to the SRSI Use Me item ($n = 7$) were deleted, resulting in a final sample of $N = 426$. The final sample was used for all analyses and reliability estimates. No variables exceeded cutoffs of 3 and 10 for high univariate skewness and kurtosis values, respectively (Weston & Gore, 2006).

Dimensionality of the GRAT-RS

The dimensionality of the GRAT-RS was first tested via a series of CFAs with M*plus* version 6.11 (Muthén & Muthén, 1998-2012). Specifically, four competing measurement models—unidimensional, three-factor oblique, second-order, bifactor—were examined. We used M*plus*' MLR option for maximum likelihood estimation, which calculates the Satorra and Bentler (1988) corrected/scaled chi-square test statistic (S-B χ^2) and associated fit indices that use it, to protect against deviations from multivariate normality. Model fit was evaluated using the S-B χ^2 statistic, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Standard Root Mean Square Residual (SRMR). The following fit criteria were used: RMSEA $\leqslant .06$, CFI $\geqslant .95$, TLI $\geqslant .95$, SRMR $< .08$ (Weston & Gore, 2006). The metric for each factor was scaled by setting the variance of each factor to one. All analyses were done at the 5% significant level. Only the bifactor solution met all fit criteria (see Table 1). Thus, the CFA results suggest that the GRAT-RS best conforms to a bifactor structure rather than a unidimensional, three-factor oblique, or second-order structure.

We next calculated the Explained Common Variance (ECV; Reise, Moore, & Haviland, 2010), an index of unidimensionality attributable to the general factor and each of the three specific factors. ECV quantifies how unidimensional versus multidimensional a set of items is. Rodriguez et al. (2016a) indicate that an ECV > .70 and a Percent of Uncontaminated Variance (PUC) > .70 suggest that the presence of some multidimensionality is not severe enough to disqualify the interpretation of the instrument as primarily unidimensional. PUC is the percentage of "unique correlations in a correlation matrix that are influenced by a single factor…The higher the PUC, the more the matrix is saturated with information relevant to estimating the parameters of a single factor and the less likely the parameter estimates in a unidimensional model will be biased" (Rodriguez, Reise, & Haviland, 2016b; p. 146).

Table 2 summarizes the factor loadings for the unidimensional and bifactor solutions for the GRAT-RS. Four findings inform the dimensionality of the GRAT-RS. First, the ECV (.45) and PUC (.70) failed to reach the cutoff for essential unidimensionality. The ECV of .45 means that only 45% of the common variance in the set of items is due to the general dimension. Second, half of the items (particularly the SA items) demonstrated stronger general factor loadings than specific factor loadings, while the other half of the items demonstrated the opposite pattern (particularly the LOSD items). Third, half of the items were better measures of their respective specific factors than the general factor, indicated by Individual Explained Common Variance (IECV) coefficients below .50 for eight of the 16 items (see Table 2). Fourth, the average difference between items' loading in the unidimensional solution and on the general factor in the bifactor solution was .04, with differences ranging from a low of .02 to a high of .09 (see Table 2). Thus, the Average Relative Measurement Parameter Bias (ARMPB; Hammer, 2016; Rodriguez et al., 2016a) associated with forcing the GRAT-RS into a unidimensional solution was 14%, which is close to violating the 10-15% upper limit discussed by Muthen, Kaplan, and Hollis (1987). The ARMPB is the mean of "the difference between each item's loading in the unidimensional solution and its general factor loading in the bifactor (i.e., the truer model), divided by the general factor loading in the bifactor" (Rodriguez et al., 2016b; p. 145). In summary, these four findings collectively suggest that the dimensionality of the construct of gratitude, as operationalized by the GRAT-RS, is best represented as a multidimensional (bifactor) solution and that the use of a unidimensional solution may be misleading. This also indicates that specific forms of gratitude, including appreciation (operationalized by the SA and

AO specific factors), are dimensionally distinct from the superordinate construct of gratitude (operationalized by the general gratitude factor).

Model-Based Reliability of the GRAT-RS Total Score

Coefficient Omega (ω) is similar to Cronbach's alpha, but a more suitable indicator of reliability for instruments with a bifactor structure. It is "a factor analytic model-based estimate of the reliability (true score variance over observed score variance) of unit-weighted test scores" (Rodriguez et al., 2016b; p. 145). *Coefficient Omega* for the total score measures the proportion of explained total score variance that can be attributed to all common factors (i.e., the general factor and all three specific factors). *Coefficient Omega Hierarchical* (ωH) measures the proportion of explained total score variance that can be attributed to the general factor after accounting for the three specific factors. It quantifies the degree to which the raw total score is a univocal indicator of the general factor (Rodriguez et al., 2016b; p. 145). The *Percentage of Reliable Variance* (PRV) for the total score (i.e., ωH divided by ω for the total score) offers a more contextualized view of the total score's reliability. Namely, by only including reliable variance and not error into account, the PRV assesses overall explained total score variance into account in a way that ωH does not. An ωH > .75 and total score PRV of > 75% would both indicate that the raw GRAT-RS total score is an appropriate measure of the general gratitude factor (Li et al., 2016; Reise et al., 2013, p. 137).

Results indicated that the general gratitude factor only accounted for 65% of the explained variance (ωH = .65) and 70% of the reliable variance (total score *PRV* = 70%) in the raw GRAT-RS total score. Therefore, the evidence did not strongly support the use of the raw GRAT-RS total score as a measure of the latent general construct of gratitude.

Model-Based Reliability of the GRAT-RS Subscale Scores

Coefficient Omega for subscales measures the proportion of explained subscale score variance that can be attributed to all common factors (i.e., the general factor and the specific factor corresponding with that subscale). *Coefficient Omega Hierarchical Subscale* (ωHS) measures the proportion of explained subscale score variance uniquely accounted for by the corresponding specific factor, after partialling out the variance accounted for by the general factor. A low ωHS precludes meaningful interpretation of the raw subscale score as a clear indicator of a specific factor (Rodriguez et al., 2016b; p. 146). The PRV for the subscale score (i.e., ωHS for that subscale score divided by ω for that subscale score) likewise offers a more contextualized view of the subscale score's reliability. An ωHS > .75 and subscale score PRV of > 75% would indicate that the raw subscale score in question is an appropriate measure of its corresponding specific factor (Li et al., 2016; Reise et al., 2013, p. 137). Table 3 summarizes the ω, ωH, ωHS, PRV, and ECV coefficients for the bifactor solution of the GRAT-RS.

The LOSD specific factor (ωHS = .74, PRV = 85%) approximated or exceeded the cutoffs, whereas the SA (ωHS = .09, PRV = 10%) and AO (ωHS = .32, PRV = 38%) specific factors did not. Thus, evidence supports the use of the raw LOSD subscale score as a measure of the LOSD specific factor, but not does support the use of the raw SA or AO subscale scores.

Evidence of Validity for the GRAT-RS Latent Factors

Although model-based reliability measures provide guidance regarding the utility of raw total and subscale scores, these methods do not address the validity of the general and specific latent factors themselves. Importantly, the specific factors' demonstration of theoretical dimensional distinctness does not necessary mean that they will demonstrate incremental clinical utility beyond the general factor (see Evidence of Validity for the GRAT-RS Latent Factors section). Latent factors can only be modeled in the context of SEM, where their relationship with theoretically-linked criterion variables can be studied with greater precision. The use of bifactor CFA to separate general factor variance from specific factor variance allows researchers to examine incremental convergent evidence of validity for each of the four GRAT-RS latent factors. Factors whose raw scores demonstrated a small PRV (see above) are the least likely to demonstrate incremental evidence of validity. However, it is best to conduct a direct test, particularly when examining latent factors with borderline-sufficient model-based reliability.

In line with extant research, we specified three separate SEM structural models in which the (a) GRAT-RS items were set to load in accordance with the bifactor model, (b) the criterion variable items (i.e., life satisfaction, positive affect, or distress) were set to load on the latent criterion variable factor, and (c) the GRAT-RS general and specific factors were simultaneously regressed onto the latent criterion variable factor. Results indicated that the general factor (life satisfaction $\beta = .32$, positive affect $\beta = .44$, distress $\beta = -.39$) and LOSD specific factor (life satisfaction $\beta = .46$, positive affect $\beta = .12$, distress $\beta = -.41$) accounted for significant variance in all three criterion variables, whereas the SA and AO specific factors never did (p's $> .14$). Therefore, the general gratitude factor and the LOSD specific factors were the only factors to demonstrate incremental convergent evidence of validity. The SA and AO specific factors did not demonstrate incremental clinical utility beyond the general gratitude factor.

Discussion

The current study re-examined the factor structure of the GRAT-RS and extended our theoretical understanding of the gratitude construct. An advantage of the GRAT-RS over existing gratitude measures is its provision of a more comprehensive assessment of gratitude through its inclusion of three purportedly distinct yet related facets of gratitude. However, there is debate regarding the composition of gratitude; is gratitude a unitary construct (Wood, Maltby, Stewart, et al., 2008), or are there aspects of gratitude, such as appreciation, that fall under the umbrella of gratitude but are uniquely associated with psychological health and well-being over and above general trait gratitude (Adler & Fagley, 2005; Fagley, 2012)? Although the GRAT-RS is one of the more commonly used multidimensional measures of gratitude, sparse examination of its factor structure presented a serendipitous gap in the literature where theory, practice, and psychometrics intersect. That is, we extended our theoretical and applied understanding of gratitude through an investigation of the dimensionality of the GRAT-RS, the model-based reliability of total and subscale scores, and incremental evidence of validity for its latent factors.

Dimensionality

The GRAT-RS best conformed to a bifactor structure. This suggests that the covariation among the 16 GRAT-RS items may be best accounted for by (a) a single general gratitude factor that reflects the common variance across all items and (b) three specific factors that capture additional (i.e., unique) common variance among clusters of items (Reise, 2012, p. 688). Ancillary bifactor measures reinforced the finding that the GRAT-RS' dimensionality is best represented with a multidimensional (bifactor) solution. In contrast, results did not provide support for the use of a unidimensional, three-factor oblique, or second-order measurement model to conceptualize the GRAT-RS. Given the 14% ARMPB found in the present study, test users should carefully consider how much measurement parameter bias they are willing to accept before choosing to force the GRAT-RS into a unidimensional model.

Regarding theoretical implications, these findings support the *three pillars of gratitude* conceptualization (Watkin, 2009, as cited in Watkins, 2014), wherein the superordinate gratitude construct functions as a general factor that exists independent of its pillars (operationalized by the orthogonal specific factors). Findings also support the theoretical distinction of the specific forms of gratitude (Fagley, 2012), and the distinction of appreciation from the superordinate construct of gratitude (Fagley, 2012; Manela, 2012; Wood, Maltby, Stewart, et al., 2008).

Reliability

Contrary to popular opinion, "the finding that an instrument conforms to a bifactor structure does not, by itself, provide a compelling rationale for the calculation and interpretation of the total or subscale scores of that instrument" (Hammer & Toland, 2016, p. 11). Therefore, model-based reliability measures were calculated to determine whether it may be appropriate to use the raw GRAT-RS total and subscale scores in research and clinical settings. Given that only 70% of the reliable total score variance was driven by the general gratitude factor, there is a lack of strong evidence in favor of using the raw GRAT-RS total score as a measure of the general gratitude factor. Rather, researchers wishing to measure general trait gratitude using the GRAT-RS would likely best capture this construct through a latent bifactor structure.

Given that only 10% and 38% of the reliable subscale score variance was driven by the SA and AO specific factors, respectively, there is a lack of strong evidence in favor of using the raw SA and AO subscale scores as measures of the SA and AO specific factors. In other words, these scores primarily re-measure the general gratitude factor, rather than the specific factors they were designed to measure. In contrast, because 85% of the reliable subscale score variance was driven by the LOSD specific factor, there is sufficient evidence that raw LOSD subscale scores can be interpreted as representing meaningful information about the LOSD specific factor.

Thus, model-based reliability estimates provided strong support for the use of one of the four raw GRAT-RS scores. This outcome is not uncommon when examining well-respected instruments from a bifactor framework (e.g., Rodriguez et al., 2016b; Li et al., 2016). When a factor cannot reliably be measured by its intended raw score, it should instead be measured as a latent SEM factor. Researchers and clinicians using the GRAT-RS are encouraged to carefully consider the risks of using insufficiently-reliable raw scores. This can result in misleading interpretations, which threatens the clinical utility of gratitude assessment via the GRAT-RS.

Validity

Building upon the model-based reliability measures, which provided guidance regarding the utility of raw total and subscale scores, the use of bifactor CFA to separate general factor variance from specific factor variance allowed for the examination of incremental convergent evidence of validity for each of the four GRAT-RS latent factors. Results indicated that the general gratitude factor and the LOSD specific factor each accounted for significant variance in all three criterion variables (i.e., life satisfaction, positive affect, or distress), whereas the SA and AO specific factors did not. This incremental convergent evidence of validity in favor of the general gratitude factor and LOSD specific factor support the continued measurement and study of these two constructs. In contrast, additional research on the incremental convergent validity of the SA and AO specific factors in the context of other criterion variables is needed.

The finding that only certain latent factors operationalized by an instrument accounted for variance in key criteria is likewise common in the bifactor literature (e.g., Brenner, Heath, Vogel, & Credé, 2017; Brewster et al., 2016; Schroder, Dawood, Yalch, Donnellan, Moser, 2016). However, the finding that the appreciation factors (i.e., SA and AO) did not account for incremental variance in these key variables is notable, as this was a fundamental clinical utility question of interest raised by Fagley (2012). Namely, Fagley found that certain appreciation-based subscales predicted life satisfaction above a general measure gratitude in a hierarchical regression and, in turn, concluded that this supports the clinical utility of appreciation above and beyond gratitude. In their study, however, Fagley used regression of correlated manifest variables rather than partialling out a general gratitude factor. If appreciation is conceptually subordinate to gratitude, as supported by this study, shared variance between the appreciation and gratitude measure(s) may present as a confound and overinflate the contribution of a specific factor (Tracey, 2012). To build upon our understanding of the relationship between appreciation and gratitude, future researchers can examine unidimensional, oblique, higher-order, and bifactor models using responses to a variety of purported gratitude and appreciation measures.

This research also presents implications for clinicians utilizing gratitude. The finding that the general gratitude factor was uniquely associated with lower distress as well as greater life satisfaction and positive affect further supports the use of general gratitude exercises. Exercises utilizing appreciation of simple pleasures or interpersonal relationships may be useful as part of this general gratitude contribution, rather than offering incremental value above other general gratitude interventions. Given the incremental convergent evidence of validity of LOSD over and above the general gratitude factor in predicting these outcomes, it may behoove clinicians to use interventions aimed at addressing how fairly one believes they have been treated in life. This might include mindfulness exercises aimed toward decreasing rumination about being treated unfairly. Focusing on LOSD may offer incremental clinical utility beyond other gratitude and appreciation interventions. Such clinical suppositions must be tested directly in future research.

Cautions, Limitations, and Additional Future Directions

As of this writing, scholarship on bifactor modeling is rapidly evolving, with guidelines for analysis and interpretation in flux. Thus, we offer these interpretations as key information for the ongoing conversation regarding the theoretical structure of gratitude and the psychometric properties of the GRAT-RS, rather than definitive conclusions. It is also important to reiterate the utility of bifactor analysis and ancillary bifactor measures, as demonstrated in the present paper. These promising yet under-utilized methods offer researchers the ability to understand the dimensionality of instruments and the constructs they operationalize with a much greater degree of precision than merely comparing alternative CFA models (Rodriguez et al., 2016b).

In addition, the generalizability of these findings to other versions, adaptations, and translations of the GRAT-RS cannot be assumed. For example, it is not certain to what degree the present results would replicate themselves with the original GRAT. Future researchers could examine the original GRAT using analyses like those used in the present study. Likewise, given the community adult makeup of this sample, it cannot be assumed that these findings will automatically generalize to other populations. Most the sample was also female, white, and reported moderate to strong gratitude. Thus, the generalizability of these findings should be tested directly in future research, rather than assumed. For example, perhaps the GRAT-RS demonstrates a different factor structure or more reliable gratitude, SA, and AO factor scores among university students, religious practitioners, or populations scoring low in gratitude.

The present results prompt future research beyond the clinical and theoretical directions mentioned earlier in this discussion. First, the SA and AO specific factors would benefit from additional investigation of the evidence—incremental and otherwise—for or against their validity. Second, if the present findings are replicated with diverse samples, a redevelopment of gratitude theory and the GRAT-RS may be indicated. Such a redevelopment could benefit from the use of exploratory bifactor analysis to help guide the creation of an instrument that will reliably measure the constructs of most interest. For example, if the specific factors (LOSD, SA, and AO) are of most interest, then developing and retaining items that primarily measure these constructs rather than the general gratitude construct would be desirable. Researchers could examine the internal structure of the full-length GRAT, which may yield a different factor structure or more reliable factor scores than found with the GRAT-RS in the current study. Finally, the GRAT-RS, like many popular social science instruments, would benefit from investigation of how sensitive the GRAT-RS scores are to changes in the gratitude construct they purport to measure (Borsboom, Mellengbergh, & van Heerden, 2004). It is essential to demonstrate that gratitude has a causal effect on responding on the GRAT-RS.

References

Adler, M. G., & Fagley, N. S. (2005). Appreciation: Individual differences in finding value and meaning as a unique predictor of subjective well□being. *Journal of personality*, *73*, 79-114.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Bernstein, D. M., & Simmons, R. G. (1974). The adolescent kidney donor: The right to give. *American Journal of Psychiatry, 131*, 1338-1343.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071. doi: 10.1037/0033-295X.111.4.1061

Brenner, R. E., Heath, P. J., Vogel, D. L., & Credé, M. (2017). Two is more valid than one: Examining the factor structure of the Self-Compassion Scale (SCS). *Journal of Counseling Psychology.* Advance online publication. doi:10.1037/cou0000211

Brewster, M. E., Hammer, J., Sawyer, J. S., Eklund, A., & Palamar, J. (2016). Perceived experiences of atheist discrimination: Instrument development and evaluation. *Journal of Counseling Psychology.* Advance online publication. doi: 10.1037/cou0000156.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6(*1*)*, 3-5.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49(*1*)*, 71-75.

Diessner, R., & Lewis, G. (2007). Further validation of the gratitude, resentment, and appreciation test (GRAT). *The Journal of Social Psychology, 147*, 445-447.

Diessner, R., Solom, R. D., Frost, N. K., Parsons, L., & Davidson, J. (2008). Engagement with beauty: Appreciating natural, artistic, and moral beauty. *The Journal of psychology*, *142*, 303-332.

Disabato, D. J., Kashdan, T. B., Short, J. L., & Jarden, A. (2016). What predicts positive life events that influence the course of depression? A longitudinal examination of gratitude and meaning in life. *Cognitive Therapy and Research*, doi:10.1007/s10608-016-9785-x

Fagley, N. S. (2012). Appreciation uniquely predicts life satisfaction above demographics, the Big 5 personality factors, and gratitude. *Personality and Individual Differences*, *53*, 59-63.

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philosophical transactions-royal society of london series b biological sciences*, 1367-1378.

Fredrickson, B. L., Tugade, M. M., Waugh, C. E., & Larkin, G. R. (2003). What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *Journal of Personality and Social Psychology, 84,* 365-376.

Froh, J. J., Bono, G., & Emmons, R. (2010). Being grateful is beyond good manners: Gratitude and motivation to contribute to society among early adolescents. *Motivation and Emotion, 34,* 144-157.

Froh, J. J., Fan, J., Emmons, R. A., Bono, G., Huebner, E. S., & Watkins, P. (2011). Measuring gratitude in youth: Assessing the psychometric properties of adult gratitude scales in children and adolescents. *Psychological Assessment, 23,* 311-324.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26,* 213-224.

Hammer, J. H. (2016, October). Average Relative Parameter Bias (ARPB) Calculator: A Microsoft Excel-based tool to calculate the ARPB statistic. Available from http://DrJosephHammer.com/.

Hammer, J. H., & Toland, M. D. (2016, November). *Bifactor analysis in Mplus*. [Video file]. Retrieved from http://sites.education.uky.edu/apslab/upcoming-events/

Hammer, J. H., & Toland, M. D. (2016). Internal Structure and Reliability of the Internalized Stigma of Mental Illness Scale (ISMI-29) and Brief Versions (ISMI-10, ISMI-9) among Americans with Depression. *Stigma and Health*. Advance online publication. doi: 10.1037/sah0000049

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., ... & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32,* 959-976.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., ... & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of general psychiatry*, 60, 184-189. doi:10.1002/mpr.310.

Kessler, R. C., Green, J. G., Gruber, M. J., Sampson, N. A., Bromet, E., Cuitan, M., ... & Lara, C. (2010). Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *International journal of methods in psychiatric research, 19(0 1),* 4-22.

Lannin, D. G., Vogel, D. L., Brenner, R. E., Abraham, T. W., & Heath, P. J. (2016). Does self-stigma reduce the probability of seeking mental health information? *Journal of Counseling Psychology, 63,* 351-358. doi:10.1037/cou0000108

Lanham, M. E. (2011). The relationship between gratitude and burnout in mental health professionals (Unpublished doctoral dissertation). University of Dayton, Dayton, OH.

Li, C. R., Toland, M. D., Usher, E. L. (2016). *Dimensionality, scoring, and interpretation of the Short Grit Scale*. Manuscript in preparation.

Li, D., Zhang, W., Li, X., Li, N., & Ye, B. (2012). Gratitude and suicidal ideation and suicide attempts among Chinese adolescents: Direct, mediated, and moderated effects. *Journal of Adolescence, 35(*1*),* 55-66.

Manela, T. (2013). Gratitude and appreciation. *American Philosophical Quarterly., 53,* 281-294).

Magno, C., & Orillosa, J. (2012). Gratitude and achievement emotions. *Philippines Journal of Counseling Psychology, 14(*1*),* 29-43.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17,* 437-455.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431-462.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Authors.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, *5*, 411-419.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*, 667-696.

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

Reise, S. P., Moore, T. N., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*, 544–559.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5–26.

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

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

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for statistics in covariance structure analysis (UCLA Statistics Series 2). *Los Angeles: University of California at Los Angeles, Department of Psychology*.

Schroder, H. S., Dawood, S., Yalch, M. M., Donnellan, M. B., & Moser, J. S. (2016). Evaluating the domain specificity of mental health–related mind-sets. *Social Psychological and Personality Science*, *7*, 508-520.

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science, 1,* 213-220.

Solom, R., Watkins, P. C., McCurrach, D., & Scheibe, D. (2017). Thieves of thankfulness: Traits that inhibit gratitude. *The Journal of Positive Psychology*, *12*(2), 120-129.

Spangler, K. (2010). *Does gratitude enhance self-esteem?* (Unpublished master's thesis). Eastern Washington University, Cheney, Washington.

Thomas, M. & Watkins, P. (2003, May). *Measuring the grateful trait: Development of revised GRAT.* Paper presented at the annual convention of the Western Psychological Association, Vancouver, British Columbia, Canada.

Toussaint, L., & Friedman, P. (2009). Forgiveness, gratitude, and well-being: The mediating role of affect and beliefs. *Journal of Happiness Studies, 10,* 635-654. doi: 10.1007/s10902-008-9111-8

Van Dusen, J. P. (2014). *Relationships amongst gratitude, well-being and depression* (Unpublished doctoral dissertation). University of Toledo, Toledo, OH

Van Dusen, J. P., Tiamiyu, M. F., Kashdan, T. B., & Elhai, J. D. (2015). Gratitude, depression and PTSD: Assessment of structural relationships. *Psychiatry Research, 230(*3*),* 867-870.

Watkins, P. C., (2009, April). *Gratitude: Amplifier of the good*. Invited paper presented to the 89th annual convention of the Western Psychological Association, Portland, OR.

Watkins, P. C. (2014). What is Gratitude and How Can It Be Measured? In P.C. Watkins (Ed), *Gratitude and the good life* (pp.13-40). doi:10.1007/978-94-007-7253-3_2

Watkins, P. C., Cruz, L., Holben, H., & Kolts, R. L. (2008). Taking care of business? Grateful processing of unpleasant memories. *The Journal of Positive Psychology, 3*(2), 87-99.

Watkins, P. C., Woodward, K., Stone, T., & Kolts, R. L. (2003). Gratitude and happiness: Development of a measure of gratitude and relationships with subjective well-being. *Social Behavior and Personality, 31*, 431–452.

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*, 1063-1070.

Weston, R., & Gore, P. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist, 34*, 719-751.

Wilder, T. (2014). *The Cabala and The Woman of Andros* [HarperCollins e-book version]. Retrieved from https://www.harpercollins.com

Wood, A. M., Froh, J. J., & Geraghty, A. W. A. (2010). Gratitude and well-being: A review and theoretical integration. *Clinical Psychology Review, 30*, 890-905

Wood, A. M., Maltby, J., Gillett, R., Linley, P. A., & Joseph, S. (2008). The role of gratitude in the development of social support, stress, and depression: Two longitudinal studies. *Journal of Research in Personality, 42*, 854-871.

Wood, A. M., Maltby, J., Stewart, N., & Joseph, S. (2008). Conceptualizing gratitude and appreciation as a unitary personality trait. *Personality and individual differences, 44*, 621-632.

Worthington, R. L., & Whitaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*, 806-838.

Table 1

Goodness of Fit Statistics for All Tested Measurement Models

Model	S-B $\chi2$	df	RMSEA [90% CI]	CFI	TLI	SRMR
GRAT-RS Unidimensional	958.61	104	.139 (.131, .147)	.632	.576	.139
GRAT-RS Three-Factor Oblique	301.77	101	.068 (.060, .077)	.914	.897	.076
GRAT-RS Second Order	301.77	101	.068 (.060, .077)	.914	.897	.076
GRAT-RS Bifactor	181.99	88	.050 (.040, .060)	.960	.945	.040

Note: The chi square for all models was statistically significant at the $p < .001$ level. GRAT-RS = Gratitude Resentment and Appreciation Test- Revised Short. Statistics are based on MLR estimation. S-B $\chi2$ = Satorra and Bentler (1988) corrected/scaled chi-square test statistic, RMSEA = Root Mean Square Error of Approximation, CI = Confidence Interval, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standard Root Mean Square Residual.

Table 2

Confirmatory Factor Analysis Standardized Loadings, IECVs, and ARMPBs

	Uni	Bifactor					RMPB
		Gen	F1	F2	F3	IECV	
Specific Factor 1: Lack of a Sense of Deprivation							
2. Life has been good to me.	**.52**	**.46**	**.39**			**.58**	.13
3. There never seems to be enough to go around and I never seem to get my share.	**.30**	**.24**	**.68**			**.11**	.25
6. I really don't think that I've gotten all the good things that I deserve in life.	**.34**	**.27**	**.67**			**.14**	.23
10. More bad things have happened to me in my life than I deserve.	**.27**	**.19**	**.74**			**.06**	.40
11. Because of what I've gone through in my life, I really feel like the world owes me something.	**.38**	**.33**	**.58**			**.24**	.15
15. For some reason I don’t seem to get the advantages that others get.	**.28**	**.20**	**.83**			**.06**	.36
Specific Factor 2: Simple Pleasures							
4. Oftentimes I have been overwhelmed at the beauty of nature.	**.47**	**.45**		**.51**		**.44**	.05
7. Every Fall I really enjoy watching the leaves change colors.	**.46**	**.44**		**.49**		**.44**	.05
9. I think that it's important to "Stop and smell the roses."	**.77**	**.79**		**.19**		**.95**	.02
12. I think that it's important to pause often to "count my blessings."	**.80**	**.83**		-.05		**1.00**	.03
13. I think it's important to enjoy the simple things in life.	**.79**	**.81**		.08		**.99**	.03
16. I think it's important to appreciate each day that you are alive.	**.81**	**.86**		-.03		**1.00**	.05
Specific Factor 3: Social Appreciation							

1. I couldn't have gotten where I am today without the help of many people.	**.54**	**.45**		**.54**	**.41**	.21
5. Although I think it's important to feel good about your accomplishments, I think that it's also important to remember how others have contributed to my accomplishments.	**.67**	**.58**		**.53**	**.55**	.15
8. Although I'm basically in control of my life, I can't help but think about all those who have supported me and helped me along the way.	**.68**	**.62**		**.46**	**.64**	.11
14. I feel deeply appreciative for the things others have done for me in my life.	**.80**	**.76**		**.33**	**.84**	.05

Note: GRAT-RS = Gratitude Resentment and Appreciation Test- Revised Short. Uni = Unidimensional Model, Gen = General Factor, F1 = Lack of a Sense of Deprivation Specific Factor, F2 = Simple Pleasures Specific Factor, F3 = Social Appreciation Specific Factor, IECV = Individual Explained Common Variance, RMPB = Relative Measurement Parameter Bias. Loadings are standardized and based on MLR estimation. All bolded standardized loadings significant at $p < .05$.

Table 3

Ancillary Bifactor Measures

	Gen	F1	F2	F3
Omega (ω)	**.92**	**.87**	**.88**	**.85**
Omega Hierarchical (ωH)	**.65**	**.23**	.02	**.03**
Omega Hierarchical Subscale (ωHS)		**.74**	**.09**	**.32**
Percentage of Reliable Variance (PRV)	**.70**	**.85**	**.10**	**.38**
Explained Common Variance (ECV)	**.45**	**.41**	**.06**	**.08**

Note: GRAT-RS = Gratitude Resentment and Appreciation Test- Revised Short. Gen = General Factor, F1 = Lack of a Sense of Deprivation Specific Factor, F2 = Simple Pleasures Specific Factor, F3 = Social Appreciation Specific Factor. All bolded coefficients significant at $p < .05$. Coefficient Omega for the total score (Gen column) measures the proportion of explained total score variance that can be attributed to all common factors (i.e., the general factor and all three specific factors). Coefficient Omega Hierarchical (ωH) measures the proportion of explained total score variance that can be attributed to the general factor after accounting for the three specific factors. The Percentage of Reliable Variance (PRV) for the total score (Gen column) is ωH divided by ω for the total score. Coefficient Omega for the subscale score measures the proportion of explained subscale score variance that can be attributed to all common factors (i.e., the general factor and the specific factor corresponding with that subscale). Coefficient Omega Hierarchical Subscale (ωHS) measures the proportion of explained subscale score variance uniquely accounted for by the corresponding specific factor, after partialling out the variance accounted for by the general factor. The PRV for the subscale score is the ωHS for that subscale score divided by ω for that subscale score. The Explained Common Variance (ECV) is an index of unidimensionality of the common variance attributable to the general factor and each of the three specific factors.

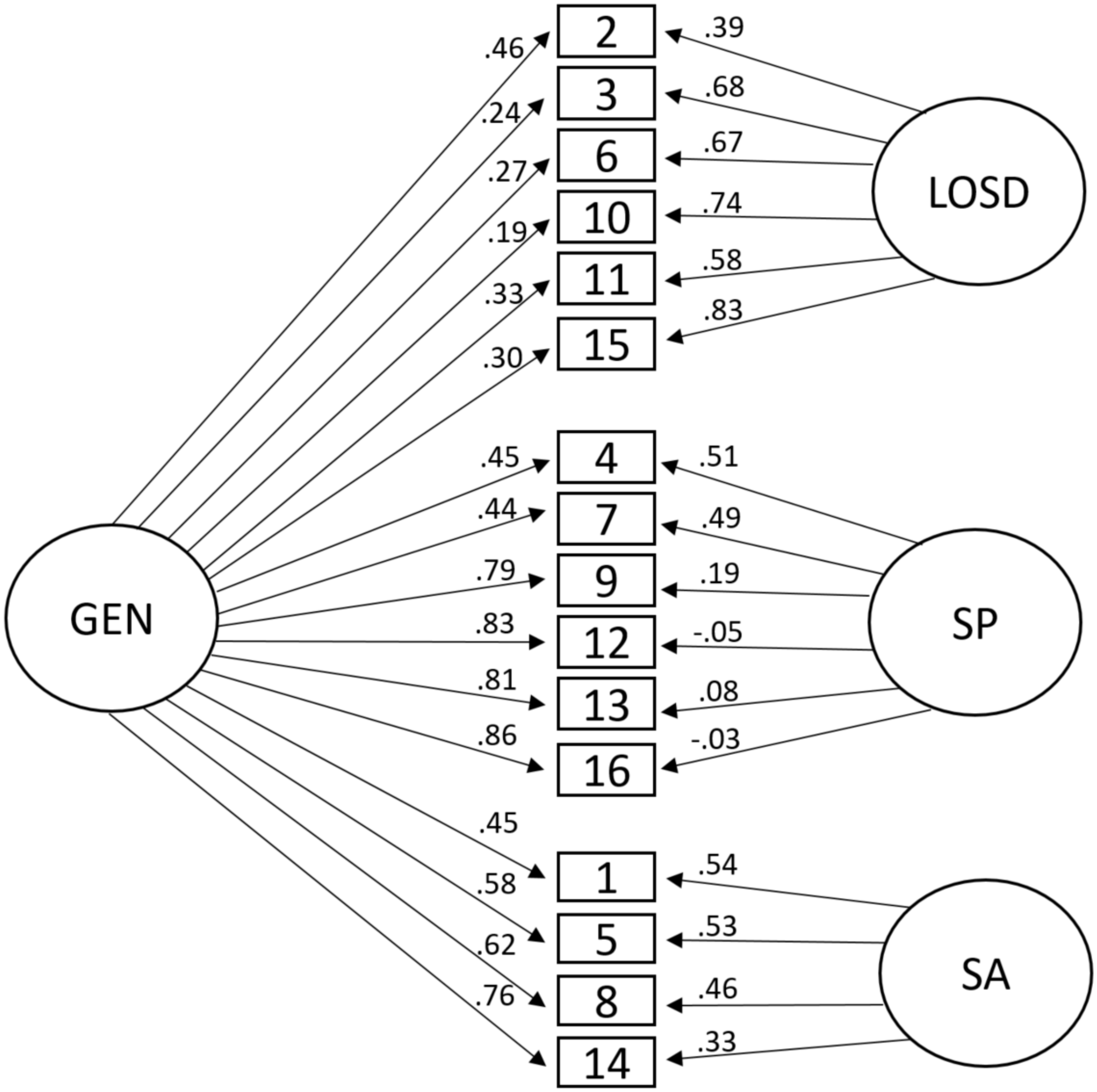

Figure 1. Bifactor measurement model displaying standardized factor loadings. Covariances, error terms, and disturbance terms are not pictured for readability. GEN = General Factor, LOSD = Lack of a Sense of Deprivation Specific Factor, SP = Simple Pleasures Specific Factor, SA = Social Appreciation Specific Factor.