UC SANTA BARBARA

Gevirtz Graduate School of Education
International Center for School Based Youth Development
January 2019

An Analytic Approach for Deciding Between 4- and 6-point Likert-Type Response Options

Melissa Gordon Wolf
Karen Nylund-Gibson
Erin Dowdy
Michael Furlong

PROJECT CoVITALITY
University of California Santa Barbara

The research reported in this paper was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant # R305A160157 to the University of California, Santa Barbara. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

ABSTRACT

This paper presents a framework for choosing between 4-and 6-point response options for use with online surveys. Using data that have both 4- and 6-point Likert-type items, we compare correlations, fit of factor analytic models, and several different reliability estimates as a way of identifying if there is empirical support for choosing a response option with more categories. Results indicate that the instrument had slightly better psychometric properties with a 4-point response option, although the estimates for both response options were acceptable. From a statistical perspective, there was no rationale to switch to a 6-point response option when a 4-point response option was already in place.

INTRODUCTION

Selecting the number of response options to include on a rating scale in a psychological measure is an under-scrutinized challenge of instrument development for researchers. One popular rating scale format, known as a Likert scale (1932), is often used to assess the magnitude of an attitude or belief about a construct. A plethora of research over the last 80 years has examined how many response options are optimal when using Likert scales. Typically, this research has focused on selecting the number of response options that optimize the psychometric properties of the instrument, while simultaneously both reducing the cognitive burden on the respondent and preserving the richness of information that can be gleaned from the data. Understanding the cognitive burden of a chosen measure involves a range of things including what is being measured and how many responses options are distinguishable by the respondent given the target population (e.g., age, knowledge level). These characteristics are often studied using interviews of respondents. In this paper, we focus on the statistical attributes related to determining the number of response options.

When concerned with optimizing the psychometric properties of an instrument, the focus is often on the reliability and validity evidence in support of the instrument's score inferences (Lietz, 2010). Reliability, often estimated by measures of internal consistency, assesses the degree to which participants respond similarly to items designed to measure the same underlying construct. Though there are many ways to estimate reliability coefficients, the most commonly used methods include calculating Cronbach's alpha, item-item correlations, item-total correlations, Coefficient Omega, Coefficient H, and factor loading strengths (Borgers, Hox, & Sikkel, 2004; Leung, 2011). To find evidence of validity that support of the uses of the instrument, researchers often focus on establishing a consistent factor structure for the set of items (i.e., construct validity) and correlating the instrument with similar and dissimilar constructs (i.e., convergent and divergent validity, respectfully). Descriptive statistics, such as skew and kurtosis, are generally evaluated as well (Dawes, 2002).

To date, there is no commonly accepted standard for determining the number of response options (Krosnick & Presser, 2010). When evaluating the ideal number of response options, recommendations from the literature are varied. For example, Rodgers, Andrews, and Herzog (1992) investigated the effect of using items with 2 to 10 response categories and concluded that expected values of validity coefficients increased by approximately .04 with each additional response option. In contrast, Bendig (1954) and Mattell and Jacoby (1971) studied the effect of using response categories ranging from 2 to 9 and found negligible impact on reliability estimates when the number of response categories was increased. Bloom, Fischer, and Orme (2003) suggest that 9-point scale may be the ideal number of response options as long as the respondent is able to make distinctions among the presented response options. However, Borgers et al. (2004) suggested the use of a 4-point scale as an optimum

after conducting studies varying the number of response options and including a neutral point on the resulting reliability. Meanwhile, Preston and Coleman (2000) found that predictive validity and item-item correlations improved with a larger number of response options.

Other studies have found no differences in terms of reliability and validity evidence when altering the number of response options. Chang (1994) used a model approach to evaluate 4- and 6-point scales and concluded that the scale points had no effect on criterion-related validity. J. Lee and Paek (2014) and Lozano et al. (2008) also found virtually no difference in the psychometric properties for an instrument when using a 4- and 6-point scale. Further, Dawes (2002) found that item skew and kurtosis were the same between a 5-point and 11-point scale. Additionally, research by J. Lee and Paek (2014) found that the psychometric properties of 2- and 3-point response options were less optimal than those with four or more, although they report no differences in the typical measures of validity and reliability for response options greater than four.

Although it is important to establish strong psychometric properties of an instrument, statistics should not guide theory in terms of instrument development. Since many statistical approaches rely on correlations among variables, there needs to be variation among the items. If more response options are used with the intent of finding more variation, it may result in larger reliability coefficients, though this is a statistical artifact and not a reflection of "good" items (J. Lee & Paek, 2014; Lozano, García-Cueto, & Muñiz, 2008; Muñiz, García-Cueto, & Lozano, 2005). Additionally, items that have a smaller number of response options are sensitive to small sample sizes and violations of normality when they are used in factor analysis, which necessitate the use of alternative estimation methods for modeling categorical or ordinal data (Rhemtulla et al., 2012). The creation of a psychological measure and the selection of response options should be grounded primarily in theory rather than optimal psychometric properties or "convenience and tradition" (Lee & Paek, 2014, p. 664).

Study Purpose

In this current study, we set out to explore if there was any empirical evidence in support of a 4- or 6-point response option for the items on one measure designed to measure the psychological strengths in students, the Social Emotional Health Survey-Secondary (SEHS-S; Furlong, You, Renshaw, Smith, & O'Malley, 2014). Since its inception, the SEHS-S has been administered using a 4-point response option. However, considering alternative response formats are an important, yet oft-overlooked aspect of ongoing scale development and refinement; thus, an alternative 6-point response format was considered.

The goal is to provide enough response options to capture the underlying variation in the population, but not too many as to create too much distinction that artificially creates variation. In addition, we were curious if adding more response options would result in better model fit/predictive validity, more variability in responses, and finer discriminations between categories and persons. In addition to evaluating if a 4- or 6-point response option format would enhance the psychometric properties of the SEHS-S, we also aimed to provide a more generalized methodological contribution. Our hope is that by presenting the rational and analyses used to determine the optimal number of response options for the SEHS-S, we can provide an example of how the procedures can be used with other psychological measures.

METHOD

Participants

The data used in this study come from students in two public secondary schools in the western United States in Grades 9-12. All students in both schools were invited to participate in the survey if they received parental permission. Slightly more than half (52.4%) of the sample identified as female. Participants identified predominantly as Latinx (62.1%) and White (22.1%). We used two independent datasets: (a) a sample of n = 1,866 where the SEHS-S was measured using a 4-point response option, and (b) a second, independent sample of n = 1,889 where the SEHS-S was measured on a 6-point response scale. Survey data were collected online using Qualtrics.

The Social Emotional Health Survey-Secondary (SEHS-S)

The Social Emotional Health Survey-Secondary (SEHS-S) is an instrument designed to assess student's psychological strengths (Furlong etal., 2014). This scale is widely used with strong empirical results supporting its psychometric properties (Furlong et al., 2014; S. Y. Lee, You, & Furlong, 2016; You et al., 2013; You, Furlong, Felix, & O'Malley, 2015). The SEHS-S has a hypothesized higher-order structure, depicted in Figure 1 (see Appendix), such that 36 survey items map onto 12 hypothesized sub-factors (each with three items), which map onto four hypothesized overall factors (each with three sub-factors), and one hypothesized overall measure of covitality.

To evaluate the differences between 4- and 6-point response options, the SEHS-S items were administered to two different populations with both 4-point and 6-point response options. The response options for the 4-point scale were: not at all true of me, a little true of me, pretty much true of me, and very much true of me. The response options for the 6-point scale were: mot at all like me, not like me, not much like me, somewhat like me, like me, and very much like me. Lower values correspond to lower levels of self-reported strengths.

To evaluate the convergent and discriminant validity of the SEHS-S, we utilized a one-item measure of life satisfaction (convergent) and the aggregate of a 10-item measure of social emotional distress (SEDS; discriminant). The SEDS is a complimentary instrument to the SEHS-S designed to evaluate a person's strengths and weaknesses simultaneously; in other words, to evaluate the "whole student" (Dowdy, Furlong, Nylund-Gibson, Moore, & Moffa, 2018).

Analytic Approach

We used several approaches to see if there was psychometric evidence to guide our selection of a 4- or 6-point response option. Our choice of methods for making comparisons was based on similar studies which compared response options (Leung, 2011) and criterion found in the literature (Lietz, 2010; Rodgers et al., 1992). Specifically, we used the following set of analyses to compare the response options: descriptive statistics (skew and kurtosis), reliability indices, model fit using confirmatory factor analysis (CFA), factor loadings, inter-item correlations, and predictive validity.

We utilized three different measures of reliability: Cronbach's alpha, omega, and coefficient H (Mcneish, 2017). Higher values indicate that there is more shared covariance between the items than unique variance, giving us confidence that the items are reflective of the same underlying construct. Ideally, we expect to find an alpha of at least .70 (on a scale of 0 to 1) to indicate a sufficiently high reliability coefficient (Streiner, 2003). We also evaluated the average inter-item correlations for each

factor; values between .20 and .50 are considered to be satisfactory (Clark & Watson, 1995). It is worth noting that inter-item correlations have an upper bound to account for item redundancy; these guidelines are contradictory to the recommendations for alpha, whereby the higher values, closer to 1, are considered best without bound. To evaluate the normality of the response option distributions, we estimated the skew and kurtosis of each item with the hope of finding values between +/- 2 (Trochim & Donnelly, 2006).

Given the ordinal nature of our response scales, some studies suggest that more response options might elucidate more normally distributed responses (Bloom et al., 2003; Leung, 2011). As it is not advisable to use maximum likelihood estimation for categorical response data with less than five categories (Rhemtulla et al., 2012), we utilized two estimation methods: maximum likelihood with robust standard errors (MLR), which is typically utilized for continuous data, and robust unweighted categorical least squares (UWLS[1]) which is typically utilized for categorical data (J. Lee & Paek, 2014; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Model fit was evaluated using the root mean square error of approximation (RMSEA) and the comparative fit index (CFI) (Brown, 2015). According to Brown (2015), good model fit for the RMSEA suggests a value less than .08, while good model fit for the CFI suggests a minimum value of .90. All models were fit in Mplus, version 8 (Muthen & Muthen, 1998-2017). Any missing data was imputed using multiple imputation.

Table 1. Model fit indices of the confirmatory factor analyses for the 4-point and 6-point models estimated using maximum likelihood estimation with robust standard errors, and unweighted least squares estimation.

Estimation	Response Option	χ^2	*df*	CFI	RMSEA (90% CI)	SRMR
MLR	4-point	1654.4	544	.958	.034 [.032, .035]	.046
MLR	6-point	1860.09	578	.936	.034 [.033, .036]	.048
UWLS	4-point	2861.49	544	.938	.049 [.047, .050]	n/a
UWLS	6-point	4213.93	578	.897	.058 [.056, .059]	n/a

Note. n/a = SRMR is not available when using UWLS estimation.

RESULTS

The skew and kurtosis for both the 4- and 6-point scale were within acceptable ranges (+/- 2). In addition, the model fit for all the CFA models (except the 6-point response option estimated using UWLS) was good: the models had CFI's greater than .90, RMSEA's less than .05, and SRMR's less than .05 (see Table 1). Upon evaluating the item and factor loadings across both estimation methods and both response options, we determined that all factor loadings were similarly high. However, the loadings were higher for the 4-point response option for more than 90% of the items across both estimation techniques (see Tables 2 and 3 in Appendix).

Factor reliabilities in Table 4, calculated using Omega total were similar although

[1] Robust weighted least squares is typically considered for categorical response options (Lipsitz et al. 2017) but recent research shows that robust unweighted least squares performs better (Rhemtulla et al., 2012).

higher for the 4-point response option. Given the popularity of Cronbach's alpha in the literature for the evaluation of reliability, and thus popular criteria for selecting the number of response options, it is included in Table 4 as well (the results are similar although downward biased due to the lack of tau-equivalence). The inter-item correlations in Table 4 were larger than the recommended upper bound value of .50 for both the 4- and 6-point response options (except for the Self Control factor). The 6-point response option had lower correlations, however, which were closer to the upper bound recommendation of .50 found in the literature (Clark & Watson, 1995). This is an expected finding since the alpha values for the 4-point response option were higher, and it is impossible to have very high alphas without high inter-item correlations when there are only three items per factor.

Table 4. First-level factor reliabilities and inter-item correlations for 4- and 6-point response scales using MLR.

	Omega		Coefficient H		Cronbach's Alpha		Average Inter-Item Correlation	
	4-point	6-point	4-point	6-point	4-point	6-point	4-point	6-point
Self-Efficacy	.755	.682	.758	.690	.754	.674	.510	.411
Self-Awareness	.614	.660	.614	.671	.612	.664	.443	.399
Persistence	.710	.638	.714	.644	.708	.633	.447	.366
School Support	.871	.790	.872	.796	.870	.783	.691	.550
Family Coherence	.897	.880	.905	.880	.895	.879	.740	.708
Peer Support	.885	.839	.902	.847	.879	.834	.712	.631
Emotional Regulation	.755	.620	.761	.639	.744	.587	.501	.338
Empathy	.846	.756	.872	.776	.837	.746	.636	.499
Self-Control	.668	.467	.673	.489	.670	.469	.404	.229
Optimism	.855	.802	.870	.806	.849	.799	.652	.571
Zest	.863	.846	.863	.847	.861	.847	.676	.639
Gratitude	.918	.852	.924	.858	.917	.850	.786	.651

Note. Bolded values indicate the higher reliability when comparing 4- and 6-point response options for each coefficient.

Upon evaluating the predictive validity via a structural equation modeling framework (estimated using UWLS; see Figure 2 in Appendix), we determined that the structural relations were stronger for the 4-point response option, as was the model fit (the CFI for the 6-point response option was under .90).

DISCUSSION

Using evaluative tools commonly found in the literature, we set out to see if there was any empirical evidence to suggest that the 6-point response option for the SEHS-S provided better model fit or discriminated better between response options than the previously used 4-point response option. Based on our results, we did not find any evidence suggesting that the use of a 6-point response option

would produce better model fit. Model fit, loadings, and reliabilities were similar across both response options, with better fit and higher loadings for the 4-point scale, which was not that surprising given that some research suggests there are not significant differences between a 4- and 6-point response option in terms of optimal psychometric properties (J. Lee & Paek, 2014). Considering the predictive validity of the SEHS-S, model fit statistics were subpar for the 6-point response option. Given the expected large sample sizes, the normally distributed response patterns, the limited number of response options, and considering that the 4-point response scale on the SEHS-S is widely adopted, we found no reason to switch to a 6-point response option. Practical considerations, and the lack of empirical evidence in support of the 6-point response option, suggest that it is advisable to continue to use the 4-point response option.

Beyond the pragmatic implications of this study's findings for the optimal presentation of the SEHS-S items, this study also provided an example of an empirical approach for other researchers engaged in instrument development when they are evaluating the optimum number of response options for use with adolescents when assessing psychological mindsets. The literature indicated that we could expect at best a modest improvement between a 4- and 6-point response option. However, we found that the psychometric properties were generally slightly better with a 4-point response option. With other considerations being similar, fewer response options place a lower cognitive demand on students when completing surveys. Researchers should therefore use theory, along with the statistical methods detailed herein, to guide the development of their measure and their selection of response options.

REFERENCES

Bloom, M., Fischer, J., & Orme, J. G. (2003). *Evaluating practice: Guidelines for the accountable professional* (4th ed.). Boston, MA: Allyn & Bacon.

Borgers, N., Hox, J., & Sikkel, D. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality and Quantity, 38*(1), 17–33. https://doi.org/10.1023/B:QUQU.0000013236.29205.a6

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford.

Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*(3), 205–215. http://doi.org/10.1177/014662169401800302

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Bloom, M., Fischer, J., & Orme, J. G. (2003). *Evaluating practice: Guidelines for the accountable professional* (4th ed.). Boston, MA: Allyn & Bacon.

Borgers, N., Hox, J., & Sikkel, D. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality and Quantity, 38*(1), 17–33. https://doi.org/10.1023/B:QUQU.0000013236.29205.a6

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: NY: Guilford.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Dowdy, E., Furlong, M. J., Nylund-Gibson, K., Moore, S., & Moffa, K. (2018). Initial validation of the Social Emotional Distress Survey – Secondary to support complete mental health screening. *Assessment for Effective Intervention, Intervention, 43,* 241–248. https://doi.org/10.1177/1534508417749871

Furlong, M. J., You, S., Renshaw, T. L., Smith, D. C., & O'Malley, M. D. (2014). Preliminary development and validation of the Social and Emotional Health Survey for secondary students. *Social Indicators Research, 117*, 1011–1032. http://link.springer.com/article/10.1007/s11205-013-0373-0

Krosnick, J. a., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263–313). West Yorkshire, England: Emerald Group. https://doi.org/10.1111/j.1432-1033.1976.tb10115.x

Lee, J., & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment, 32*(7), 663–673. https://doi.org/10.1177/0734282914522200

Lee, S. yeon, You, S., & Furlong, M. J. (2016). Validation of the Social Emotional Health Survey–Secondary for Korean students. *Child Indicators Research, 9*(1), 73–92. https://doi.org/10.1007/s12187-014-9294-y

Lei Chang. (1994). A Psychometric Evaluation of 4-Point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*(3), 205–215. https://doi.org/10.1177/014662169401800302

Leung, S.-O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-Point Likert Scales. *Journal of Social Service Research, 37*(4), 412–421. https://doi.org/10.1080/01488376.2011.580697

Lietz, P. (2010). Research into questionnaire design: A summary of the literature. International *Journal of Market Research, 52*(2), 249. https://doi.org/10.2501/S147078530920120X

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

Mcneish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*. https://doi.org/10.1037/met0000144

Rhemtulla, M., Brosseau-Liard, P., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(354–373), 354–373.

Rodgers, W. L., Andrews, F. M., & Herzog, A. R. (1992). Quality of survey measures: A structural equation modeling approach. *Journal of Official Statistics, 8*(3), 251–275.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001

Trochim, W. M., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Cincinnati, OH: Atomic Dog.

You, S., Furlong, M. J., Felix, E., & O'Malley, M. (2015). Validation of the social and emotional health survey for five sociocultural groups: Multigroup invariance and latent mean analyses. *Psychology in the Schools, 52*(4), 349–362. https://doi.org/10.1002/pits.21828

You, S., Furlong, M. J., Dowdy, E., Renshaw, T. L., Smith, D. C., & O'Malley, M. D. (2013). Further Validation of the Social and Emotional Health Survey for High School Students. *Applied Research in Quality of Life, 9*(4), 997–1015. https://doi.org/10.1007/s11482-013-9282-2

APPENDIX

Table 2. Factor loadings for the higher order factor model estimated using MLR estimation.

Table 3. Factor loadings for the higher order factor model estimated using unweighted least squares estimation.

Figure 1. The higher-order factor structure of the SEHS-S.

Figure 2. Predictive validity of the 4-point and 6-point models estimated using ULS estimation.

Table 2. ***Factor loadings for the higher-order factor model estimated using MLR estimation.***

Item	Belief in Self (BS) 4-point	Belief in Self (BS) 6-point	Belief in Others (BO) 4-point	Belief in Others (BO) 6-point		Emotional Competence (EC) 4-point	Emotional Competence (EC) 6-point	Engaged Living (EL) 4-point	Engaged Living (EL) 6-point
BS: Self-Efficacy 1	**.687**	.606			EC: Emotional Regulation 1	**.750**	.667		
BS: Self-Efficacy 2	**.702**	.615			EC: Emotional Regulation 2	**.732**	.636		
BS: Self-Efficacy 2	**.747**	.713			EC: Emotional Regulation 3	**.652**	.472		
BS: Self-Awareness 1	.677	**.708**			EC: Empathy 1	**.674**	.595		
BS: Self-Awareness 2	*	.590			EC: Empathy 2	**.870**	.744		
BS: Self-Awareness 3	**.654**	.578			EC: Empathy 3	**.860**	.791		
BS: Persistence 1	**.711**	.659			EC: Self Control 1	**.606**	.441		
BS: Persistence 2	**.674**	.614			EC: Self Control 2	**.605**	.391		
BS: Persistence 3	**.625**	.550			EC: Self Control 3	**.689**	.590		
BO: School Support 1			**.835**	.774	EL: Optimism 1			**.787**	.757
BO: School Support 2			**.811**	.682	EL: Optimism 2			**.886**	.795
BO: School Support 3			**.849**	.779	EL: Optimism 3			**.767**	.721
BO: Family Coherence 1			**.842**	.830	EL: Zest 1			**.834**	.819
BO: Family Coherence 2			**.910**	.852	EL: Zest 2			**.810**	.795
BO: Family Coherence 3			.833	**.844**	EL: Zest 3			**.825**	.800
BO: Peer Support 1			.763	**.757**	EL: Gratitude 1			**.882**	.825
BO: Peer Support 2			**.865**	.778	EL: Gratitude 2			**.924**	.846
BO: Peer Support 3			**.911**	.853	EL: Gratitude 3			**.857**	.761
BS: Self-Efficacy	**.919**	.900			EC: Emotional Regulation	.876	**.930**		
BS: Self-Awareness	.968	**.984**			EC: Empathy	**.604**	.565		
BS: Persistence	**.786**	.692			EC: Self Control	.929	**.971**		
BO: School Support			**.724**	.629	EL: Optimism			.886	**.922**
BO: Family Coherence			**.669**	.635	EL: Zest			.713	**.740**
BO: Peer Support			**.509**	.464	EL: Gratitude			.672	**.758**

Higher order factor loadings	4-point	6-point
Belief in Self	.914	**.987**
Belief in Others	**.970**	.960
Emotional Competence	**.759**	.640
Engaged Living	.862	**.952**

Note. Bolded values reflect the higher factor loading between the two response options. *This item was mistakenly not administered during the iteration of the survey that utilized a 4-point response option. Six of the 36 items on the 4-point response option were administered with a 5-point response option. These items represent the factors known as Zest and Gratitude. The response options for the 5-point scale were Not at all, Very little, Somewhat, Quite a lot, and Extremely.

Table 3. **Factor loadings for the higher-order factor model estimated using unweighted least squares estimation.**

Item	Belief in Self 4-point	Belief in Self 6-point	Belief in Others 4-point	Belief in Others 6-point		Emotional Competence 4-point	Emotional Competence 6-point	Engaged Living 4-point	Engaged Living 6-point
BS: Self-Efficacy 1	**.749**	.665			EC: Emotional Regulation 1	**.856**	.738		
BS: Self-Efficacy 2	**.779**	.669			EC: Emotional Regulation 2	**.749**	.666		
BS: Self-Efficacy 2	**.824**	.724			EC: Emotional Regulation 3	**.727**	.494		
BS: Self-Awareness 1	.770	**.776**			EC: Empathy 1	**.796**	.704		
BS: Self-Awareness 2	*	.622			EC: Empathy 2	**.885**	.755		
BS: Self-Awareness 3	**.699**	.602			EC: Empathy 3	**.907**	.805		
BS: Persistence 1	**.742**	.672			EC: Self Control 1	**.656**	.515		
BS: Persistence 2	**.708**	.628			EC: Self Control 2	**.597**	.341		
BS: Persistence 3	**.691**	.577			EC: Self Control 3	**.805**	.657		
BO: School Support 1			**.873**	.793	EL: Optimism 1			**.831**	.804
BO: School Support 2			**.887**	.741	EL: Optimism 2			**.890**	.805
BO: School Support 3			**.903**	.834	EL: Optimism 3			**.854**	.739
BO: Family Coherence 1			**.963**	.884	EL: Zest 1			.771	**.778**
BO: Family Coherence 2			**.931**	.895	EL: Zest 2			**.856**	.833
BO: Family Coherence 3			**.846**	.825	EL: Zest 3			**.929**	.878
BO: Peer Support 1			**.961**	.881	EL: Gratitude 1			**.903**	.847
BO: Peer Support 2			**.826**	.763	EL: Gratitude 2			**.927**	.844
BO: Peer Support 3			**.919**	.864	EL: Gratitude 3			**.925**	.831
BS: Self-Efficacy	**.891**	.895			EC: Emotional Regulation	.885	**.929**		
BS: Self-Awareness	**.969**	.956			EC: Empathy	**.668**	.611		
BS: Persistence	**.831**	.717			EC: Self Control	.904	**.923**		
BO: School Support			**.749**	.661	EL: Optimism			**.915**	.900
BO: Family Coherence			**.700**	.625	EL: Zest			.685	**.726**
BO: Peer Support			**.579**	.500	EL: Gratitude			.748	**.810**

Higher order factor loadings	4-point	6-point
BS	.912	**.986**
BO	**.984**	.974
EC	**.779**	.673
EL	.858	**.939**

Note. Bolded values reflect the higher factor loading between the two response options. *This item was mistakenly not administered during the iteration of the survey that utilized a 4-point response option. Six of the 36 items on the 4-point response option were administered with a 5-point response option. These items represent the factors known as Zest and Gratitude. The response options for the 5-point scale were Not at all, Very little, Somewhat, Quite a lot, and Extremely.

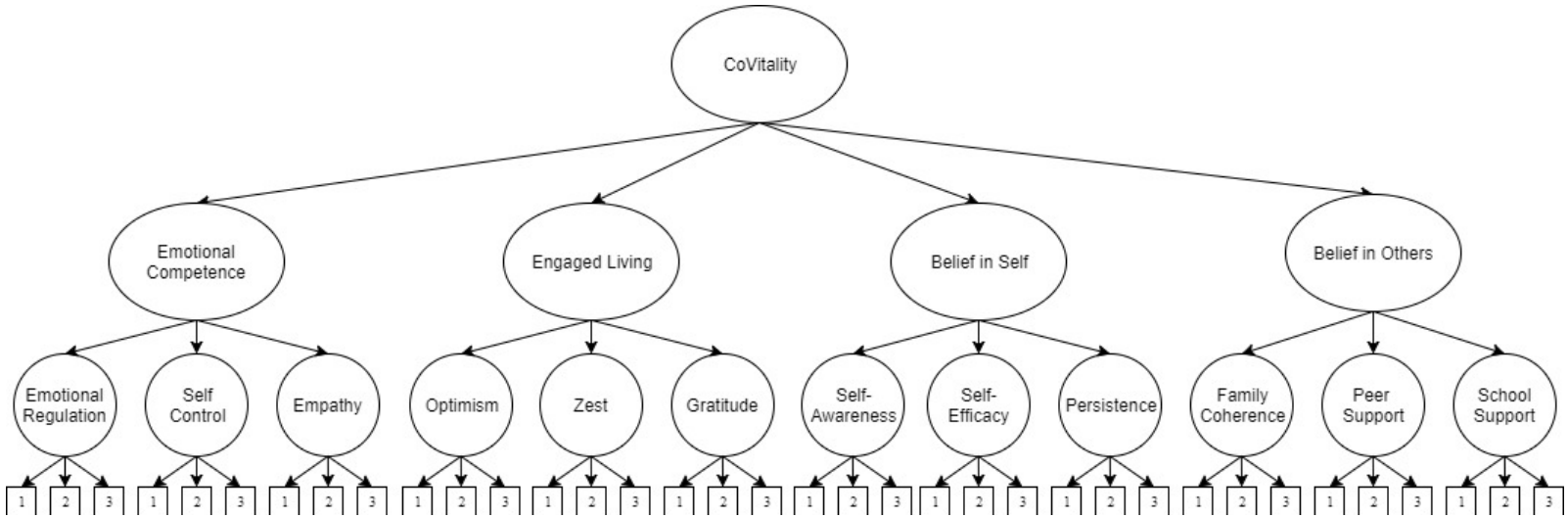

Figure 2. **The higher-order factor structure of the SEHS-S.**

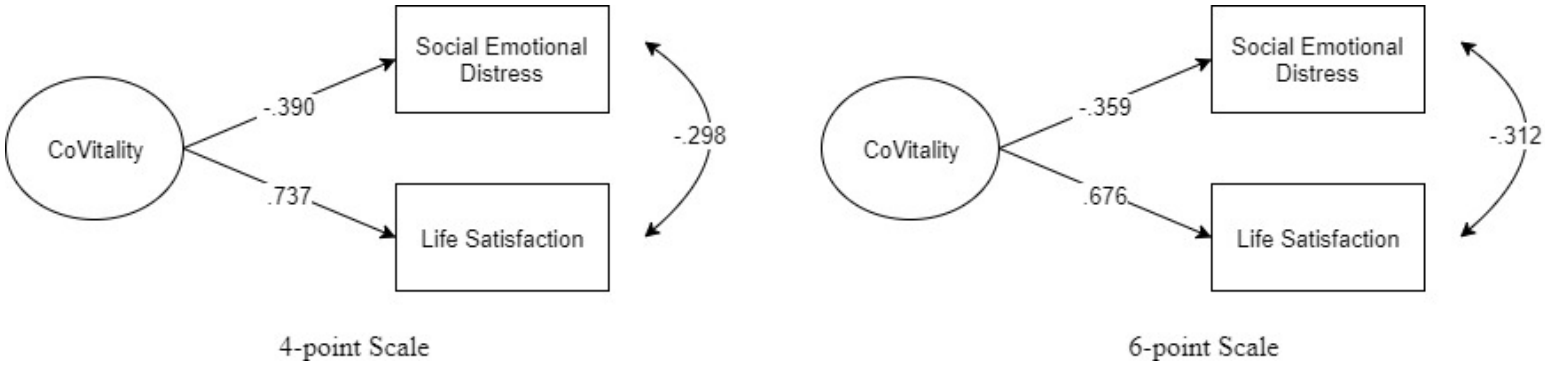

Figure 2. **Predictive validity of the 4-point and 6-point models estimated using ULS estimation.**