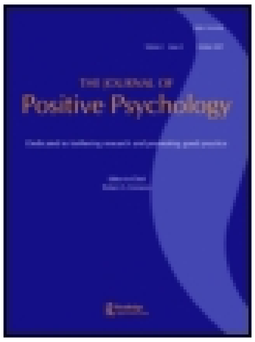
Are there virtuous types? Finite mixture modeling of the VIA Inventory of Strengths

Article *in* The Journal of Positive Psychology · August 2018

DOI: 10.1080/17439760.2018.1510021



The Journal of Positive Psychology
Dedicated to furthering research and promoting good practice

ISSN: 1743-9760 (Print) 1743-9779 (Online) Journal homepage: http://www.tandfonline.com/loi/rpos20

Are there virtuous types? Finite mixture modeling of the VIA Inventory of Strengths

Dawn M. Berger & Robert E. McGrath

THE JOURNAL OF POSITIVE PSYCHOLOGY
https://doi.org/10.1080/17439760.2018.1510021

Are there virtuous types? Finite mixture modeling of the VIA Inventory of Strengths

Dawn M. Berger[a*] and Robert E. McGrath[a]

[a]School of Psychology, Fairleigh Dickinson University, Teaneck, NJ, USA

ABSTRACT
Philosophical and religious traditions often refer to 'the virtuous person.' This terminology usually carries with it the assumption that a class of individuals exists who have achieved a virtuous state. This study attempted to test that implication. The VIA Inventory of Strengths (VIA-IS) is intended as a comprehensive assessment of character strengths, which are conceptualized as markers of virtuous character. One prior study using taxometric methods found no evidence for the existence of such a category of individuals using VIA-IS scores. Subsequent literature has suggested the superiority of finite mixture modeling for identifying categorical structure. Latent profile analyses of 1–10 classes were conducted in a stratified sample of 10,000 adults. The results provided little evidence for class structure, and support thinking of virtue as something we must continuously pursue rather than a state that we achieve.

ARTICLE HISTORY
Received 15 October 2017
Accepted 25 July 2018

KEYWORDS
Character strengths; virtue; latent profile analysis (LPA); finite mixture modeling (FMM)

In ethics and in various religions, the concept of virtue is often used to identify optimal character. It is a common practice in these traditions to refer to 'the virtuous.' This terminology suggests the existence of moral exemplars who have achieved a virtuous state that distinguishes them from the general population. Religious traditions in particular have described groups of people who have achieved a level of adaptation to the world that places them in an exalted category. In some instances, these groups are thought to be extremely rare: examples include the Christian concept of the saint or the bodhisattva of Mahayana Buddhism. Other terms are applied to broader classes of individuals whose status is still thought to set them apart from the general population, such as the swami, the guru, the Sufi, or the shaman. This assumption of the exceptional person has infiltrated into more secular thinking as well. Despite skepticism about how the concept of virtue was employed, Nietzsche's (1891/1978) overman referred to individuals who embodied those features he considered most admirable. In psychology, Rank's (1989/1932) concept of the artist and Maslow's (1998/1968) self-actualized individual reflect similar assumptions about the existence of general paragons.

The existence of a linguistic category is insufficient to assure that a distinct state exists, however. Reference to 'the virtuous' may simply be a linguistic convention for referring to individuals who are generally more virtuous than the norm. Concepts such as the saint may have been invented to personify religious ideals, or out of the wish for role models in life, as Campbell (1949) proposed for the heroic archetype. Finally, assuming the existence of people who are 'virtuous' creates the opportunity for looking at certain individuals as exemplars of virtue, and these people can then be held up as guides to how to lead a life of virtue (e.g., Hatzimalonas, 2018). Literature exists suggesting that exposure to individuals considered moral exemplars encourages others to engage in more moral behavior (e.g., Han, Kim, Jeong, & Cohen, 2017).

The question of whether there are individuals who demonstrate a qualitatively distinct state of virtue is therefore a question with interesting implications for philosophy, for many religions, and for psychologists interested in the topic of moral and adaptive exceptionalism. If a distinct state of virtue does not exist, it raises questions about the potential for identifying individuals who can serve as guides for right behavior across a variety of situations and contexts. It also brings into question whether a stable state of virtuous character is achievable.

In recent years, a number of methodological tools have emerged that can provide insight into this question of the fundamental status of virtue as a categorical state. First, Peterson and Seligman (2004) attempted the development of a comprehensive and cross-culturally valid

CONTACT Robert E. McGrath mcgrath@fdu.edu
*Present address: Dawn Berger is now at Community Health Center, Inc., Norwalk, CT.

model of the dimensions central to exemplary character. Based on a three-year research effort involving more than 50 experts in various attributes of individuals considered to be positive (e.g., humor and forgiveness), these authors identified 24 key positive personal attributes that tend to be stable within individuals, and referred to these attributes as character strengths. They conceptualized these attributes as markers of six virtues that appear ubiquitously across literary moral traditions (Dahlsgaard, Peterson, & Seligman, 2005), and that they defined as 'the core characteristics valued by moral philosophers and religious thinkers' (Peterson & Seligman, 2004, p. 13). The character strengths were in turn defined as 'the psychological ingredients – processes or mechanisms – that define virtues' (p. 13). Now referred to as the VIA Classification of Strengths and Virtues, this model was developed with cross-cultural generality in mind. Character strengths are personal traits that are morally valued, contribute to individual fulfillment, and are simultaneously beneficial to the person, to their colleagues, and to their community. The developers of the model therefore hypothesized the 24 strengths would be recognizable and relevant across human societies. Though this hypothesis has not been fully tested, the existing evidence indicates substantial relevance across cultures (McGrath, 2014, 2016), even in cultures without a strong literary tradition (Biswas-Diener, 2006). The VIA Classification is provided in Table 1.

Table 1. The VIA classification of strengths and virtues.

Virtues	Character Strengths
Wisdom & Knowledge	Creativity [originality, ingenuity]
	Curiosity [interest, novelty-seeking, openness to experience]
	Judgment & Open-Mindedness [critical thinking]
	Love of Learning
	Perspective [wisdom]
Courage	Bravery [valor]
	Perseverance [persistence, industriousness]
	Honesty [authenticity, integrity]
	Zest [vitality, enthusiasm, vigor, energy]
Humanity	Capacity to Love and Be Loved
	Kindness [generosity, nurturance, care, compassion, altruistic love, 'niceness']
	Social Intelligence [emotional intelligence, personal intelligence]
Justice	Teamwork [citizenship, social responsibility, loyalty]
	Fairness
	Leadership
Temperance	Forgiveness & Mercy
	Modesty & Humility
	Prudence
	Self-Regulation [self-control]
Transcendence	Appreciation of Beauty and Excellence [awe, wonder, elevation]
	Gratitude
	Hope [optimism, future-mindedness, future orientation]
	Humor [playfulness]
	Religiousness & Spirituality [faith, purpose]

Adapted from 'Character Strengths and Virtues: A Classification and Handbook,' by C. Peterson and M. E. P. Seligman, (2004), American Psychological Association, pp. 29–30. Copyright 2004 by Values in Action Institute.

Second, Peterson and Seligman (2004) developed a measure to detect the 24 character strengths in adults, called the VIA Inventory of Strengths. The instrument has recently been revised (McGrath, 2017), but an earlier 120-item version (VIA-120), comprised of five items for each strength, has been available free of charge online since 2013 at the website of the VIA Institute on Character (http://www.viacharacter.org). The VIA-120 has now been completed over 2.5 million times by individuals around the world. Recent research with the VIA Inventory and related measures has identified a stable three-factor structure underlying the 24 strengths that provides a more empirically reliable model of virtue than that incorporated in the VIA Classification (McGrath, 2015; McGrath, Greenberg, & Hall-Simmonds, 2018). The three virtues in this model have been called Caring, Inquisitiveness, and Self-Control.

Third, several statistical strategies have been developed with special relevance to the question of whether a set of dimensional observed variables reflect an underlying dimensional, dichotomous, or polytomous (here used to refer more than two distinct classes) structure. These models include taxometric methods (Ruscio, Haslam, & Ruscio, 2006; Waller & Meehl, 1998), cluster analysis (DiStefano, 2012), and finite mixture models (FMMs; McLachlan & Peel, 2000). Several authors have previously compared these strategies (see Beauchaine, 2003; Beauchaine & Beauchaine, 2002; Lenzenweger, McLachlan, & Rubin, 2007). To summarize some of the key differences, taxometric methods were developed primarily to evaluate the competing hypotheses of dimensional versus dichotomous underlying structure. In contrast, FMMs and cluster analysis were developed to divide cases or variables under the assumption that a certain number of latent classes exist. Though some work has been conducted attempting to extend taxometric methods to instances involving more than two classes (McGrath & Walters, 2012), FMM and cluster analysis are inherently intended for application to polytomous as well as dichotomous structures.

When the number of underlying classes is not known or is not hypothesized a priori, the use of cluster analysis and FMMs requires the comparison of models varying in the number of classes. In practice, this problem is generally addressed through an iterative process in which each analysis assumes one more class than the previous analysis. Various strategies have emerged for comparing the relative goodness of fit for these models. While various authors have suggested standards for identifying the best-fitting model in cluster analysis (e.g., Milligan & Cooper, 1985; Tonidandel &

Overall, 2004), no standard has emerged. In contrast, the expectation-maximization algorithm widely used for the estimation of FMMs allows for the computation of a likelihood function that provides the basis for various goodness of fit indices. FMM also has an advantage over cluster analysis in the latter's sensitivity to the metric of the variables involved. Taken together, these comparisons suggest FMM as a particularly useful statistical approach to addressing questions about dimensional versus dichotomous or polytomous latent structure. Simulation research also suggests FMM tends to provide more accurate results on taxonic structure than cluster analysis (Cleland, Rothschild, & Haslam, 2000).

To date, only one study has empirically investigated the question of whether virtue represents a distinct state of operating in the world. McGrath, Rashid, Park, and Peterson (2010) examined 83,576 U.S. residents who completed the VIA Inventory between 2002 and 2003. That study drew the conclusion that the underlying structure for the 24 VIA strengths was dimensional. However, the published study only reported results from taxometric analyses, so only dichotomous structure was considered as an alternative to dimensional structure. The present study instead applied FMM to the character strengths. This allowed for tests of structures involving more than two classes, to evaluate whether there are subtypes of virtuous character. This strategy is consistent with literature criticizing prior research with the VIA Classification that examines the strengths as individual elements rather than as part of a holistic approach to operating in the world (Fowers, 2008; Kristjánsson, 2010).

Method

Participants

The sample for this study was drawn from an archival data set of 2,452,519 adults who created a free account on the VIA Institute on Character website and completed the VIA-120 online with no missing data between 2013 and 2017. In instances where the same individual completed the instrument more than once, the date of completion was used to exclude all but the first administration.

For the present study, participation was limited to residents of the United States, for two reasons. If an international sample were used, the failure to find class structure in the data could reflect cultural differences in the understanding of what it means to be virtuous. Second, demographic statistics suggested a very skewed sample, so the decision was made to stratify the sample to more accurately reflect the adult population, and demographics would tend to vary across nations.

The data set included 813,453 participants who identified the U.S. as their nation of residence. Within this subset, the sample was 62.7% female and 37.3% male. The sample was much younger on average than the general American population, with a mean age of 34.1 (*SD* = 13.6) and 40.6% in their 20s. Those who reported educational level were also highly educated, with 59.8% reporting a bachelor's degree or higher. Ethnicity data were not collected.

To obtain a sample more consistent with the general U.S. adult population, a stratified sample of 10,000 cases was generated. Unfortunately, education level was missing for more than 90% of cases and age for 14.3%. Given the skew was worst for education, while gender was available for more than 99% of respondents, the decision was made to approximate 2010 U.S. Census data for these two variables. SAS unrestricted random sampling with replacement was used for the stratified sampling (SAS Institute, 2017). The final sample was allocated 50% to each gender, matched Census data exactly on educational distribution, and was more consistent with Census data on age than was the original sample. The comparison of the original and stratified samples to Census data is provided in the upper part of Table 2.

Measure

The VIA-120 is a face-valid self-report questionnaire using a five-point rating scale ranging from *very much like me* to *very much unlike me* to measure the extent to which various behavioral and self-descriptive statements relevant to the character strengths apply. The original included 10 items representing each strength; the VIA-120 items were the five items from each scale with the highest corrected item-total correlation in a sample of over 400,000 U.S. residents. All items on the original VIA Inventory were keyed positively; no justification has been provided for the lack of reverse keying. The VIA-IS has consistently demonstrated adequate internal consistency and test-retest reliability (Park, Peterson, & Seligman, 2004). In the stratified sample, Cronbach's alpha estimates ranged between .71 and .88 across scales with the exception of Teamwork (α = .54). Validity has been established through comparisons with appropriate criteria as well as ratings of character strengths by informants (Park et al., 2004; Ruch et al., 2010). One study found that only two of 24 scores correlated significantly with a measure of social desirability (Peterson & Seligman, 2004).

Table 2. Comparison of original and stratified U.S. Samples to 2010 census data.

Variable	Census	Stratified Sample	Original Sample
Education			
No HS degree	14.40	14.00	1.69
HS graduate	28.50	28.00	2.14
Some college	21.30	22.00	12.04
Associate degree	7.60	8.00	4.36
Bachelor's degree	17.70	18.00	29.67
Graduate work	10.40	10.00	50.10
Euclidean distance		1.14	51.69
Age			
20–24	9.62	17.27	24.37
25–29	9.75	15.89	16.20
30–34	8.93	9.41	13.10
35–39	8.93	7.62	10.48
40–44	9.34	7.20	8.88
45–49	10.16	8.55	8.15
50–54	9.89	11.31	7.16
55–59	8.65	10.26	5.68
60–64	7.28	6.70	3.48
65+	17.45	5.80	2.51
Euclidean distance		15.68	23.18
Gender			
Male	48.40	50.00	37.34
Female	51.60	50.00	62.67
Euclidean distance		2.26	15.65
Key Scales			
Gratitude		3.86	3.94
Kindness		4.12	4.22
Love		3.92	4.05
Creativity		3.86	3.82
Curiosity		3.87	3.89
Learning		3.94	3.64
Perseverance		3.62	3.83
Prudence		3.71	3.70
Self-Regulation		3.17	3.22

HS = high school. Values for demographic variables are percents, values for key scales are means. Euclidean distances are based on comparison with U.S. Census data.

Procedure

Literature suggests that the results of classification analyses can be rendered inaccurate by the inclusion of indicators that do not reflect the latent class structure, and that analysis is improved by limiting the analysis to those variables most indicative of that structure (Raftery & Dean, 2006). Given the focus on virtue, preliminary analyses were conducted to identify which of the 24 character strengths best identified each of the three empirically reliable virtues: Caring, Inquisitiveness, and Self-Control. Specifically, confirmatory factor analysis (CFA) was used to identify a subset of the strengths that demonstrated good fit for purposes of locating participants on the three latent factors.

Once this subset was identified, FMMs – or more precisely, latent profile analyses, since the strength scores were treated as dimensional indicators – were estimated beginning with one class, and incrementing the number of classes until it exceeded the number of indicators. All analyses were conducted using Mplus version 8 (Muthén & Muthén, 1998–2017). Default options for FMM with Mplus were used in all analyses, except that the number of random starts was increased as necessary to achieve replication of the log of the likelihood estimate and bootstrap draws associated with the bootstrapped likelihood ratio test described below. In particular, reliance on default settings meant that variables were assumed to be orthogonal (conditionally independent) within classes.

For CFAs, the set of dimensional goodness of fit indicators provided by Mplus was reviewed. These included three information criteria: Akaike (AIC; Akaike, 1973), Bayesian (BIC; Schwarz, 1978), and sample-size adjusted BIC (SABIC; Yang, 2006). Because values for these statistics are not constrained or standardized, there is no absolute standard for fit. Instead, values are compared across nested models, with lower values suggesting less loss of information, i.e., better fit. It also included the root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis Index (TLI), and standardized root mean square residual (SRMR). Following Hu and Bentler (1999), a value close to .95 for the TLI and CFI, .08 for SRMR, and .06 for RMSEA were treated as evidence of good fit in CFA.

The three information criteria were also used to evaluate the FMMs. Entropy, an indicator of precision in classification, was also reviewed, with values close to 1 (the maximum value) considered evidence of well-specified classes. Mplus also generates three significance tests for FMMs: the Vuong-Lo-Mendell-Rubin likelihood ratio test (VLMR), the Lo-Mendell-Rubin adjusted likelihood ratio test (LMR), and the parametric bootstrapped likelihood ratio test (BLRT; Lo, Mendell, & Rubin, 2001; McLachlan & Peel, 2000; Vuong, 1989). Each evaluates whether the current model demonstrates significantly superior fit to a model with one fewer classes based on the log likelihood difference between models. The three tests are closely related statistically, and they tend to produce very similar results. In fact, the BLRT is based on the same test statistic value as the VLMR, but the probability of a Type I error is determined using bootstrapping rather than the hypothetical distribution for two times the log likelihood difference.

Nylund, Asparouhov, and Muthén (2007) compared the LMR, BLRT, AIC, and BIC as indicators of the accuracy of FMMs. They found the BIC superior to the AIC, particularly as sample size increased. However, the BLRT proved the best indicator across different types of latent class models. Finally, some authors have pointed to the importance of subjectively evaluating the fit of the models to the theoretical understanding of the construct before their acceptance, including the distribution of the cases to classes (e.g., Marsh, Lüdtke, Trautwein, & Morin, 2009).

Results

The indicators used to specify the three latent virtues in the CFA were based on findings reported by McGrath et al. (2018). They reviewed 12 exploratory factor analyses of the 24 character strengths, and found the same three-factor structure in each case. Seventeen strength scales were associated with loadings ≥ .40 on only one factor in ≥ 10 of the 12 samples. For Caring, the best indicators proved to be Forgiveness, Gratitude, Kindness, Leadership, Love, and Teamwork; for Inquisitiveness, they were Bravery, Hope, Creativity, Curiosity, Love of Learning, and Zest; and for Self-Control, Honesty, Modesty, Perseverance, Prudence, and Self-Regulation. Goodness of fit indices were consistently unacceptable (see Table 3). Subsequent analyses included subsets of the 17 indicators. The best combination was based on a mean loading across the 12 samples ≥ .60, no other mean loadings ≥ .40, and clear conceptual relationship to the factor. These criteria were met by three indicators for each factor: Gratitude, Kindness, and Love for Caring; Creativity, Curiosity, and Love of Learning for Inquisitiveness; and Perseverance, Prudence, and Self-Regulation for Self-Control. Fit indices consistently indicated the acceptability of this model (see Table 3). All standardized loadings for this model were ≥ .44. Correlations between factors varied between .27 (Inquisitiveness and Self-Control) and .56 (Caring and Inquisitiveness). These nine scales were then used as the basis for all FMMs. The lower part of Table 2 provides means for these nine scales comparing the stratified sample to the original sample. As a measure of the size of the differences, absolute *d* values were computed comparing the means for the two samples. All were small (< .20) with the exception of those for Perseverance (.29) and Learning (.36). The median value was .12.

Table 3 also provides the fit criteria for each of a series of FMMs estimating 1–10 classes. The pattern of results offers no clear evidence for one solution over any other. As the number of classes increased, the three information criteria consistently declined, suggesting a model with more classes than predictors was the best model. To highlight this issue further, Akaike and Schwarz weights were estimated for the AIC and BIC, respectively, using the following formula (Vandekerckhove, Matzke, & Wagenmakers, 2015):

$$w_{IC_i} = \frac{\exp(-.5*(IC_i - IC_{min}))}{\sum_1^k exp(-.5*(IC_i - IC_{min}))}$$

That is, the weight for each information criterion is based on its difference from the minimum information criterion value as a proportion of the total differences. The ratio of two weights can be interpreted as the relative likelihood that one model is accurate relative to another. What is striking is that the weight for every information criterion value except the smallest was zero, indicating a probability of zero that that model was correct relative to the model associated with the smallest AIC or BIC value. This might be taken to suggest the 10-class model as the best model, but the same pattern of all weights equaling zero except the last was replicated if the analysis had stopped at any number of classes less than 10.

Consistent with this finding, in every case significance tests indicated more polytomous models were superior, with the exception of the LMR and VLMR for the 4-class versus 3-class solutions. Entropy values were also consistently poor, in only one point achieving a minimally acceptable value of .80 (with rounding). The lack of consistency across fit statistics, with the failure of the BLRT to

Table 3. Fit statistics for confirmatory factor analyses and latent profile analyses.

	AIC	AIC *w*	BIC	BIC *w*	SABIC	CFI	TLI	RMSEA	SRMR	Entropy	VLMR	LMR	BLRT
CFA													
Initial Model	319,832.62		320,221.98		320,050.38	.75	.70	.13	.08				
Final Model	181,879.19		182,095.50		182,000.17	.95	.92	.07	.04				
FMM													
1 class	204,417.65	0.00	204,547.43	0.00	204,490.23								
2 classes	192,982.77	0.00	193,184.66	0.00	193,095.68					0.74	11,454.88	11,331.85	11,454.88
3 classes	189,085.30	0.00	189,359.30	0.00	189,238.54					0.75	3917.46	3875.39	3917.46
4 classes	186,216.69	0.00	186,562.79	0.00	186,410.25					0.74	2888.61*	2857.59*	2888.61
5 classes	184,122.39	0.00	184,540.59	0.00	184,356.27					0.77	2114.30	2091.59	2114.30
6 classes	182,571.19	0.00	183,061.50	0.00	182,845.40					0.75	1571.20	1554.32	1571.20
7 classes	181,494.16	0.00	182,056.56	0.00	181,808.69					0.76	1097.04	1085.25	1097.04
8 classes	180,451.13	0.00	181,085.64	0.00	180,805.99					0.78	1063.03	1051.61	1063.03
9 classes	179,591.80	0.00	180,298.42	0.00	179,986.99					0.80	879.32	869.88	879.32
10 classes	179,136.11	1.00	179,914.83	1.00	179,571.62					0.79	725.55	717.75	725.55

*Not significant. All other significance tests were significant at $p \leq .001$.

Note. AIC = Akaike information criterion; AIC *w* = Akaike weight; BIC = Bayesian information criterion; BIC *w* = Schwarz weight; SABIC = Sample-size adjusted BIC; CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation; SRMR = standardized root mean square residual; VLMR = Vuong-Lo-Mendell-Rubin test value; LMR = Lo-Mendell-Rubin test value; BLRT = parametric bootstrapped test value.

identify a stopping point in particular, suggests none of the class-based models fit the data adequately.

Despite the lack of a clear 'winner,' a case could be made for two outcomes. The lack of significance for two tests associated with the four-class model would suggest an advantage to the three-class model. The nine-class model was associated with the largest value for entropy, though it was still barely .80. Table 4 provides the means and probability of membership for each class in these two models.

The three-class model suggests cluster 3 members are higher on average than either other cluster on all nine variables. This could suggest a class of individuals who are generally virtuous (Exceptional Virtue). Class 2 members are higher on all strengths reflecting Caring and Inquisitiveness, but generated the poorest scores of all three groups on measures of Self-Control (Caring and Inquisitive). Class 1 means varied only between 3.02 and 3.74, suggesting a moderate level of virtue across the board (Typical Virtue).

The results for the nine-class solution are more difficult to interpret because of their greater complexity. Figure 1 provides graphs of the means for each class. The matrix of means is also repeated in Table 4 only including relative outliers, i.e., means > 4.0 or < 3.0, to simplify the presentation. Based on the extreme means, the classes could be given the following descriptors:

- Class 1: Kind
- Class 2: Caring
- Class 3: Information Gatherers
- Class 4: Inquisitive
- Class 5: Poor Goal-Seekers
- Class 6: Persevering
- Class 7: Very Caring/Inquisitive
- Class 8: Caring/Inquisitive
- Class 9: Exceptional

The models can also be evaluated based on their consistency with expectations based on the conceptual understanding of virtue. For example, the three-class solution suggested 53% of participants belonged in the exceptional category. In contrast, the two categories in the nine-class solution that were associated with higher than average scores across the board included only 13% of respondents. The latter would seem a more accurate representation of reality than the former. On the other hand, the three-class solution would seem to be a more intuitive model for virtue classification. Furthermore, three of the nine classes combined accounted for < 10% of participants, a finding that suggests the model was generating trivial groups.

Two sets of supplementary FMMs were conducted to explore for alternative factors that could have generated weak evidence for class structure. First, concerns have been raised that FMM fit indicators are sensitive to larger sample size, with the result that in large samples each increase in the number of classes results in better fit (e.g., Marsh et al., 2009). A random subsample of 1,000 cases was drawn from the sample without replacement, and again evaluated for one to ten classes. Again, Akaike and Schwarz weights were all zero except those for the ten-class solution, all entropy values were < .80 except one, all BLRTs were significant ($p < .001$), and most VLMR and LMR tests were significant. The only model for which there was evidence was the six-class solution (entropy = .80, and VLMR and LMR for the seven-class solution not significant), failing to replicate the conclusion drawn from the full sample.

Second, it has been suggested that measurement error can affect FMM results (Meyer & Morin, 2016). To test this possibility, scores were generated for the three factors (Caring, Inquisitiveness, and Self-Control) using the nine indicators, and FFMs retaining 1–4 classes were computed. Again, all Akaike and Schwarz weights were 0 except those for the 4-class solution, no entropy score exceeded .76, and all significance tests were significant ($p < .001$). It is noteworthy that for the three-class solution, the mean scores for all three factors were highest in class 3 and lowest in class 1. Marsh et al. (2009) have suggested that solutions in which the relative elevation of means across classes is the same across all indicators, i.e., where the differentiation of classes is purely quantitative, raise doubts about the appropriateness of a class-based model. Neither sample size nor measurement error seems to be adequate as an alternative explanation for the weak evidence for class structure. This conclusion essentially replicates conclusion drawn previously by McGrath et al. (2010) using taxometric analysis that suggested the absence of a categorical latent structure for the strengths.

Discussion

In summary, the results offer limited evidence of meaningful classes of individuals in terms of their level of virtue. Of course, this study is not without its limitation. Advocates of moral philosophy or theology could object to self-reported character strengths as an evidence base for conclusions about concepts as complex and nebulous as virtue, similar to Walsh's (2015) discussion of the nature of wisdom. Informant ratings of virtue could provide a very different context for the

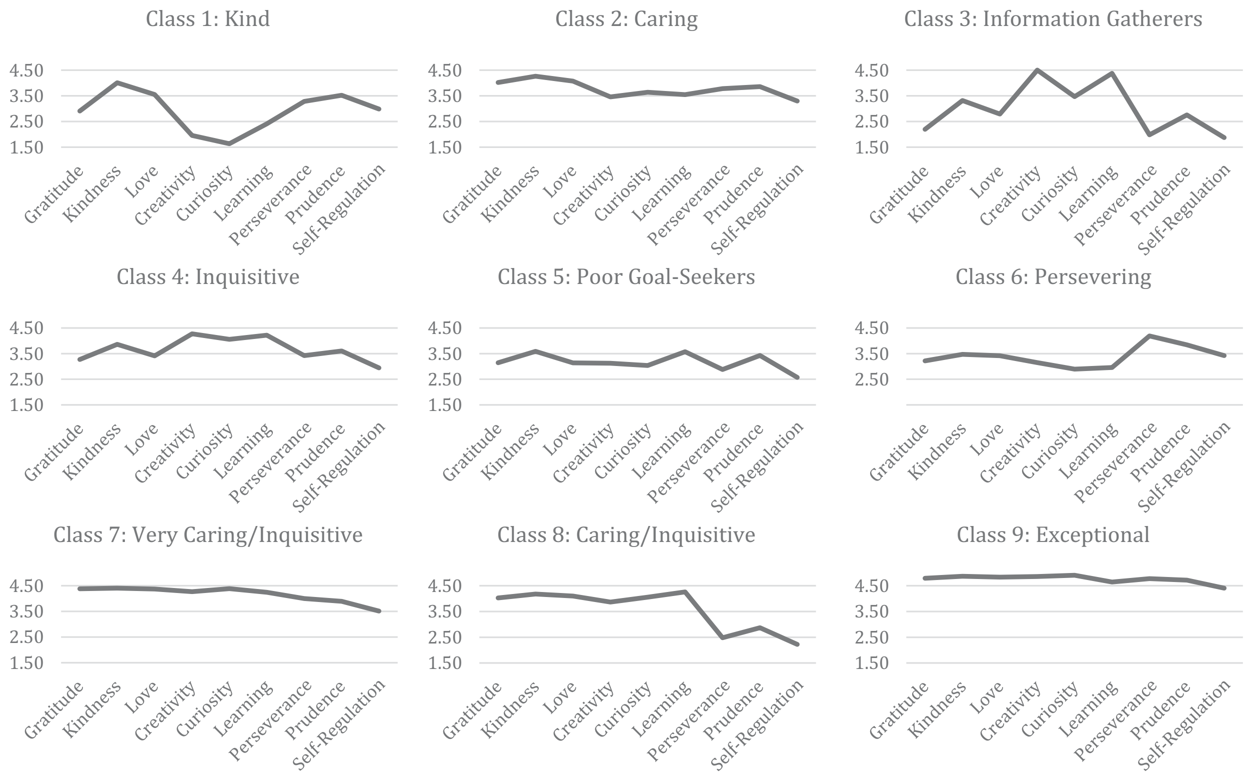


Figure 1. Means for the nine-class solution. Strengths include gratitude, kindness, love, creativity, curiosity, learning, perseverance, prudence, and self-regulation.

Table 4. Means for three-class and nine-class solutions.

	1	2	3	4	5	6	7	8	9
3 classes									
Gratitude	3.36	3.41	4.30						
Kindness	3.74	3.92	4.39						
Love	3.49	3.58	4.30						
Creativity	3.10	4.00	4.10						
Curiosity	3.02	3.86	4.24						
Learning	3.17	4.17	4.12						
Perseverance	3.61	2.87	4.00						
Prudence	3.69	3.22	3.97						
Self-Regulation	3.08	2.52	3.53						
p(class)	.20	.27	.53						
9 classes									
Gratitude	2.91	4.02	2.19	3.27	3.14	3.22	4.38	4.03	4.79
Kindness	4.01	4.26	3.32	3.86	3.59	3.47	4.41	4.18	4.87
Love	3.55	4.08	2.79	3.42	3.14	3.42	4.37	4.10	4.84
Creativity	1.96	3.46	4.50	4.27	3.13	3.15	4.27	3.87	4.86
Curiosity	1.63	3.64	3.47	4.06	3.04	2.90	4.39	4.06	4.91
Learning	2.41	3.55	4.38	4.22	3.57	2.96	4.25	4.26	4.64
Perseverance	3.28	3.78	1.98	3.42	2.88	4.19	4.00	2.48	4.78
Prudence	3.52	3.86	2.76	3.60	3.43	3.84	3.89	2.87	4.72
Self-Regulation	2.99	3.30	1.87	2.95	2.58	3.42	3.51	2.22	4.41
p(class)	.01	.23	.02	.16	.09	.06	.30	.09	.04
9-class extremes									
Gratitude	2.91	4.02	2.19				4.38	4.03	4.79
Kindness	4.01	4.26					4.41	4.18	4.87
Love		4.08	2.79				4.37	4.10	4.84
Creativity	1.96		4.50	4.27			4.27		4.86
Curiosity	1.63			4.06		2.90	4.39	4.06	4.91
Learning	2.41		4.38	4.22		2.96	4.25	4.26	4.64
Perseverance			1.98		2.88	4.19		2.48	4.78
Prudence			2.76					2.87	4.72
Self-Regulation	2.99		1.87	2.95	2.58			2.22	4.41

evaluation of a taxon of virtue, and would be a worthwhile source of data for a replication.

Second, the use of traditional psychometric scales that assume a monotonic relationship between scores and the underlying construct is similarly problematic in the context of virtue measurement, where there is a strong tradition suggesting a 'golden mean' that represents an optimal state between excess and deficit (Fowers, 2008; Schwartz & Sharpe, 2006). Even Peterson (2006) and Seligman (2014), the original authors of the VIA-IS, have expressed support for the hypothesis of an optimal level for each strength.

Third, the VIA model has been criticized for its bias towards Western cultural norms and ideals in its conceptualizations of character and virtue (Held, 2005; Kristjánsson, 2010). The VIA model represents only one possible perspective on the nature of character and virtue, and the relationship between those two concepts. Other perspectives could conceivably lead to very different conclusions about the latent nature of virtue.

Fourth, the use of a self-selected Internet sample raises concerns about the external validity of the study. In particular, the respondents needed to have sufficient resources to have access to the Internet, and needed to have researched the concept of character strengths sufficiently to approach the sites through which data were collected. Furthermore, respondents who accessed the VIA Institute on Character website may have been biased in their responding by their interest in these topics. Finally, some traditions suggest true virtuousness is a rare state. If that is the case, the procedures used in studies such as this one may simply be incapable of detecting a state with a very low base rate.

With these caveats in mind, the present study raises reasonable concerns about traditional perspectives on what should be the goals of efforts to encourage virtue or moral development. It suggests the possibility that ancient concept of 'the virtuous person' as a distinct class may be a cultural myth rather than an acute observation of reality. Virtue may not be a state we achieve, one that has associated with it some degree of permanency. Instead, the present findings would suggest it may be more accurate to think in terms of more or less virtuousness. It is also consistent with concluding our capacity to demonstrate excellence in character will vary across situations and contexts. Finally, these findings offer a framework for a new perspective on thinking about one of the longstanding concepts in the field of virtue theory, often called the unity or reciprocity of the virtues (Vaccarezza, 2017). Specifically, they would suggest a reformulation of the concept not as the simultaneous *presence* of multiple virtues in an individual, but as co-occurring high *levels* of the virtues.

These findings do not necessarily support the purely situationist model of virtue suggested by Doris (1998). Even if virtues are better seen as dimensions than as categories, it can still be hypothesized that some people will consistently behave more virtuously or wisely than others will when they are placed in the same situation. That is, findings about quantitative versus qualitative structure to the virtues does not have implications for whether virtues are states or traits.

The findings are more consistent with Vranas' (2009) claim that evaluations of individuals as organically good, bad, or indeterminate are inherently faulty. However, where Vranas suggests virtue is undetectable, the present results suggest what may be undetectable is a distinct class of virtuous people.

The perception of a class of individuals who have achieved a permanent state of virtuous character distinct from that of the great majority of people may provide a social ideal that many consider worth pursuing. However, this study raises the possibility it misrepresents what is possible, and could even encourage complacency once people believe they have achieved a state of relatively virtuous character, or have been identified by others as exemplifying such a state. Rather than seeing it as a status achieved, perhaps it would be better to think of virtue as an ideal that must be continuously pursued, an aspiration rather than a destination.

Acknowlegement

We are grateful to the VIA Institute on Character for access to the data used in this study.

Disclosure statement

The second author is a Senior Scientist for the VIA Institute on Character, which holds the copyright to the VIA Inventory of Strengths.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principles. In B. N. Petrov & F. Csaski (Eds.), *Second international symposium of information theory* (pp. 267–281). Budapest, Hungary: Academia Kiado.

Beauchaine, T. P. (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology*, *15*, 501–527.

Beauchaine, T. P., & Beauchaine, R. J., III. (2002). A comparison of maximum covariance and *k*-means cluster analysis in classifying cases into known taxon groups. *Psychological Methods*, *7*, 245–261.

Biswas-Diener, R. (2006). From the equator to the North Pole: A study of character strengths. *Journal of Happiness Studies*, *7*, 293–310.

Campbell, J. (1949). *The hero with a thousand faces*. Princeton, NJ: Princeton University Press.

Cleland, C. M., Rothschild, L., & Haslam, N. (2000). Detecting latent taxa: Monte Carlo comparison of taxometric, mixture model, and clustering procedures. *Psychological Reports*, *87*, 37–47.

Dahlsgaard, K., Peterson, C., & Seligman, M. E. P. (2005). Shared virtue: The convergence of valued human strengths across culture and history. *Review of General Psychology*, *9*, 203–213.

DiStefano, C. (2012). Cluster analysis and latent class clustering techniques. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 645–666). New York, NY: Guilford.

Doris, J. (1998). Persons, situations and virtue ethics. *Noûs*, *32*, 504–530.

Fowers, B. J. (2008). From continence to virtue: Recovering goodness, character unity, and character types for positive psychology. *Theory & Psychology*, *18*, 629–653.

Han, H., Kim, J., Jeong, C., & Cohen, G. L. (2017). Attainable and relevant moral exemplars are more effective than extraordinary exemplars in promoting voluntary service engagement. *Frontiers in Psychology*, *8*, 283.

Hatzimalonas, M. R. (2018). Philotimo: Vices and virtues of a moral archetype. In N. L. Brown, T. Lomas, & F. J. Eiroa-Orosa (Eds.), *The Routledge international handbook of critical positive psychology* (pp. 474–487). New York, NY: Routledge/Taylor & Francis.

Held, B. S. (2005). The 'virtues' of positive psychology. *Journal of Theoretical and Philosophical Psychology*, *25*, 1–34.

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

Kristjánsson, K. (2010). Positive psychology, happiness, and virtue: The troublesome conceptual issues. *Review of General Psychology*, *14*, 296–310.

Lenzenweger, M. F., McLachlan, G., & Rubin, D. B. (2007). Resolving the latent structure of schizophrenia endophenotypes using expectation-maximization-based finite mixture modeling. *Journal of Abnormal Psychology*, *116*, 16–29.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767–778.

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, *16*, 191–225.

Maslow, A. H. (1998). *Toward a psychology of being*. New York, NY: Wiley. (Originally published 1968).

McGrath, R. E. (2014). Character strengths in 75 nations: An update. *Journal of Positive Psychology*, *10*, 41–52.

McGrath, R. E. (2015). Integrating psychological and cultural perspectives on virtue: The hierarchical structure of character strengths. *Journal of Positive Psychology*, *10*, 407–424.

McGrath, R. E. (2016). Measurement invariance in translations of the VIA inventory of strengths. *European Journal of Psychological Assessment*, *32*, 187–194.

McGrath, R. E. (2017). *Technical report: The VIA assessment suite for adults: Development and initial evaluation*. Cincinnati, OH: VIA Institute on Character. Retrieved from http://www.viacharacter.org/www/Portals/0/VIA%20Technical%20Report%202017-Assessment%20Suite.pdf

McGrath, R. E., Greenberg, M. J., & Hall-Simmonds, A. (2018). Scarecrow, Tin Woodsman, and Cowardly Lion: The three-factor model of virtue. *Journal of Positive Psychology*, *13*, 373–392.

McGrath, R. E., Rashid, T., Park, N., & Peterson, C. (2010). Is optimal functioning a distinct state? *The Humanistic Psychologist*, *38*, 159–169.

McGrath, R. E., & Walters, G. D. (2012). Taxometric analysis as a general strategy for distinguishing categorical from dimensional latent structure. *Psychological Methods*, *17*, 284–293.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Meyer, J. P., & Morin, A. J. S. (2016). A person-centered approach to commitment research: Theory, research, and methodology. *Journal of Organizational Behavior*, *37*, 584–612.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*, 159–179.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Nietzsche, F. (1978). *Thus spoke Zarathustra*. W. Kaufman, Trans. New York, NY: Penguin. (Original work published 1891).

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling*, *14*, 535–569.

Park, N., Peterson, C., & Seligman, M. E. P. (2004). Strengths of character and well-being. *Journal of Social and Clinical Psychology*, *23*, 603–619.

Peterson, C. (2006). The Values in Action (VIA) classification of strengths. In M. Csikszentmihalyi & L. Csikszentmihalyi (Eds.), *A life worth living: Contributions to positive psychology* (pp. 29–48). New York, NY: Oxford University Press.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Washington, DC/New York, NY: American Psychological Association/Oxford University Press.

Raftery, A. E., & Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association*, *101*, 168–178.

Rank, O. (1989). *Art and artist: Creative urge and personality development*. New York, NY: W. W. Norton. (Originally published 1932).

Ruch, W., Proyer, R. T., Harzer, C., Park, N., Peterson, C., & Seligman, M. P. (2010). Values in Action Inventory of Strengths (VIA-IS): Adaptation and validation of the German version and the development of a peer-rating form. *Journal of Individual Differences*, *31*, 138–149.

Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). *Introduction to the taxometric method: A practical guide*. Mahwah, NJ: Lawrence Erlbaum.

SAS Institute. (2017). *SAS 9.4 language reference: Concepts* (6th ed. ed.). Cary, NC: Author.

Schwartz, B., & Sharpe, K. E. (2006). Practical wisdom: Aristotle meets positive psychology. *Journal of Happiness Studies*, *7*, 377–395.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Seligman, M. E. P. (2014). Chris Peterson's unfinished masterwork: The real mental illnesses. *Journal of Positive Psychology*, *10*, 3–6.

Tonidandel, S., & Overall, J. E. (2004). Determining the number of clusters by sampling with replacement. *Psychological Methods*, *9*, 238–249.

Vaccarezza, M. S. (2017). The unity of the virtues reconsidered: Competing accounts in philosophy and positive psychology. *Review of Philosophy and Psychology*, *8*, 637–651.

Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford library of psychology: The Oxford handbook of computational and mathematical psychology* (pp. 300–319). New York, NY: Oxford University Press.

Vranas, P. B. (2009). Against moral character evaluations: The undetectability of virtue and vice. *Journal of Ethics*, *13*, 213–233.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica : Journal of the Econometric Society*, *57*, 307–333.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.

Walsh, R. (2015). What is wisdom? Cross-cultural and cross-disciplinary syntheses. *Review of General Psychology*, *19*, 278–293.

Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis*, *50*, 1090–1104.