ISSN-L 1015-5759 • ISSN-Print 1015-5759 • ISSN-Online 2151-2426

Official Organ of the European Association of Psychological Assessment

european journal of psychological assessment

www.hogrefe.com/journals/ejpa

Abstracted/Indexed in:
Current Contents/
Social & Behavioral Sciences
Social Sciences Citation Index (SSCI)
Social Scisearch
PsycINFO
Psychological Abstracts
PSYNDEX
ERIH
Scopus

Edited by
Matthias Ziegler

HOGREFE

Original Article

Measurement Invariance in Translations of the VIA Inventory of Strengths

Robert E. McGrath

School of Psychology, Fairleigh Dickinson University, Teaneck, NJ, USA

Abstract. The VIA Classification of Character Strengths and Virtues (Peterson & Seligman, 2004) has been an influential contribution to the study of prosocial traits, and provided the basis for the VIA Inventory of Strengths (VIA-IS). Inherent to the Classification is the assumption that the character strengths included in the model are cross-culturally relevant. The emergence of a latent trait model for the VIA Classification from exploratory factor analytic research and the availability of data from translated versions of the VIA-IS provides a basis for evaluating this assumption. A sample of 15,540 individuals from 16 nations who completed the VIA-IS online was used to evaluate measurement equivalence. Multigroup confirmatory factor analysis and a relatively new statistical procedure, alignment analysis, were used to evaluate configural, metric, and scalar invariance across translations of the instrument. Consistent support was found for configural and metric invariance, and scalar invariance was also demonstrated under a number of circumstances. The findings lend support to the cross-cultural relevance of the VIA Classification of Character Strengths and Virtues as well as to existing translations of the VIA-IS.

Keywords: virtues, character strengths, multigroup confirmatory factor analysis, cross-cultural relevance

Peterson and Seligman (2004) made a very influential contribution to the study of positive social functioning with the development of their model of strengths and virtues, sometimes called the VIA Classification of Character Strengths and Virtues. Based on a review of documents from eight moral and religious traditions that have had an enduring impact on the understanding of socially valued behavioral patterns (e.g., Athenian Greece and Confucianism), they found six concepts that they considered cross-culturally recognized virtues: Wisdom and Knowledge, Courage, Humanity, Justice, Temperance, and Transcendence (Dahlsgaard, Peterson, & Seligman, 2005). They also identified 24 character strengths. The list of strengths and their descriptions was the product of an intensive 3-year process with input from more than 50 experts on positive and negative functioning, extensive brainstorming, reviews of historical lists of virtues, and examination of popular literature and media (N. Mayerson, personal communication, June 23, 2011). These were conceptualized as more personal prosocial trait variables, and were therefore seen as amenable to psychological measurement. Peterson and Seligman also developed an instrument called the VIA Inventory of Strengths (VIA-IS), a 240-item self-report measure comprised of 10-item scales representing each of the strengths, that is currently available through the VIA Institute on Character.

One of Peterson and Seligman's (2004) most interesting innovations was the modeling of the virtues as hierarchically encompassing subsets of the character strengths. For example, the Wisdom and Knowledge virtue subsumed five character strengths, including Creativity, Curiosity, and Love of Learning; Capacity to Love and Be Loved and Kindness were classified as elements of Humanity, while Fairness was characterized as a more specific manifestation of Justice. Though this hierarchical structure was based on conceptual considerations, it is reminiscent of latent measurement modeling. It is not surprising to find then that a number of studies have since been conducted evaluating whether the 24 strengths measured by the VIA-IS collapse into six factors as the VIA model would suggest.

To date, at least eight studies have been conducted evaluating the latent dimensional structure of the 24 VIA-IS scales using exploratory factor analytic techniques (Brdar & Kashdan, 2010; Littman-Ovadia & Lavy, 2012; Macdonald, Bore, & Munro, 2008; McGrath, 2014a; Peterson, Park, Pole, D'Andrea, & Seligman, 2008; Peterson & Seligman, 2004; Ruch et al., 2010; Shryack, Steger, Krueger, & Kallie, 2010; Singh & Choubisa, 2010). These have consistently found the original VIA Classification does not provide an accurate representation of their latent structure, which is not surprising given the Classification was derived conceptually. Instead, research has consistently identified 3–5 factors, with certain commonalities across solutions despite marked differences in the samples and factor analytic techniques used.

The most extensive of these analyses was completed by McGrath (2014a), involving almost 460,000 US residents, and using several different strategies for estimating the

DOI: 10.1027/1015-5759/a000248

number of factors, factor extraction, and factor rotation. The best solution for the existing VIA-IS scales suggested five latent factors, which was also the most common number retained in other factor analytic studies. Interpersonal Strengths overlapped primarily with Peterson and Seligman's Humanity and Justice virtues. Emotional Strengths did not correspond well with any of the six virtues, but a similar factor had emerged in prior studies (e.g., Ruch et al., 2010). Strengths of Restraint primarily encompassed strengths Peterson and Seligman associated with Courage and Temperance. Theological Strengths corresponded well with the Transcendence virtue, while Intellectual Strengths was consistent with the Wisdom and Knowledge virtue. The names of these factors were drawn from previous factor analytic studies, to highlight the degree of consistency in findings across studies.

The emergence of some consistency in exploratory factor analytic studies provides a basis for evaluating another element of the VIA Classification. An important claim underlying the VIA Classification is the cross-cultural generalizability of the strengths. Though many of the data sources and experts that contributed to the identification of the 24 character strengths were reflective of Western cultures, the strengths were intended to be widely if not universally recognized as valuable cultural concepts. The strengths were thought to be sufficiently important for optimal social and personal functioning that all cultures would demonstrate some analog. Subsequent research has in fact suggested that the strengths listed by Peterson and Seligman are recognized across a variety of cultures, including many that still rely on oral traditions (e.g., Biswas-Diener, 2006; McGrath, 2014b).

Based on the assumed cross-cultural familiarity of the VIA character strengths, a number of researchers requested permission from the VIA Institute to translate the VIA-IS. Permission is granted so long as the research team meets the following conditions: (1) items and instructions must be translated as necessary to maintain the psychological meaning of the items; (2) the research team must back-translate the items to English to verify equivalent meaning, and modify translations as necessary to assure rough equivalence in meaning; (3) an academic psychologist from the target culture or a person who has completed a VIA training course to familiarize themselves with the VIA model must be involved in the translation process; and (4) the translation is considered conditional until at least 300 cases have been gathered and adequate reliability has been demonstrated for each of the 24 scales. So far, the VIA-IS has been translated from English into 20 languages. These translations are all freely available for completion at the website of the VIA Institute (http://www.viacharacter.org).

The collection of data using these translations creates an opportunity to evaluate the validity of the very important assumption of cross-cultural consistency in the understanding of the character strengths. Specifically, multigroup confirmatory factor analysis has become a popular tool for evaluating whether the same latent structure is applicable to culturally distinct subgroups of individuals. If so, the findings support the conclusion that the constructs underlying responses to the variables are equivalent across groups, and that they have the same meaning across groups. The purpose of the present study was therefore to evaluate measurement invariance across samples from different countries who completed different translations of the VIA-IS.

Method

Participants

Cases were excluded if they omitted even one of the 240 items. In addition, upon completion they received immediate feedback about their results. This level of compliance suggests a fairly strong investment in completing the instrument, probably indicating a desire to obtain accurate feedback in most cases.

For purposes of this study, the VIA Institute provided data from individuals who completed the VIA-IS online at the VIA website between 2010 and 2012. There were 15 nations for which a sufficient number of individuals had completed a translated version of the VIA-IS to be included in this study: Denmark, the Netherlands, France, Germany, Israel, Italy, Japan, Korea, Portugal, Brazil, Spain, Sweden, Turkey, mainland China (simplified Chinese), and Hong Kong (traditional Chinese). A sample of 1,000 cases was made available for each of the translations, except Danish ($N = 999$), Dutch ($N = 827$), and Turkish ($N = 714$). A sample of 1,000 English-speaking US residents served as the reference group for this study, for a total sample size of 15,540. Table 1 provides demographic information for the sample as a whole.

Given the group sizes, it is not surprising to find significant differences across the 16 samples on every variable in Table 1 ($p < .001$). The mean η^2 value across one-way analyses of variance comparing the groups on the 24 character strengths was .08, with a range of [.03, .11]. The η^2 value for age was .19, while Cramér's V was .19 for gender and .31 for education. All of these effects can be considered small- to medium-sized.

Statistical Analysis

Two statistical approaches were used in this study. The first was multigroup confirmatory factor analysis (MGCFA), which is the standard approach to evaluating measurement invariance across multiple groups. The second approach is a relatively new one that complements traditional MGCFA, called factor alignment analysis (Asparouhov & Muthén, 2014). Where MGCFA is often used to evaluate overall model fit across multiple groups, alignment analysis offers a tool for identifying which parameters in the model demonstrate equivalence across which groups. This will be described in more detail below. All statistical analyses were conducted using Mplus version 7.11 (Muthén & Muthén, 2012).

Table 1. Descriptive statistics for the entire sample

Variable	*N*	*M*	*SD*	%
Gender				
Female	8,712			56.67
Male	6,662			43.33
Education				
< HS graduate	462			5.69
HS graduate	888			10.93
Some college	2,178			26.81
Associate degree	834			10.26
Bachelor's degree	1,856			22.84
Graduate degree	1,907			23.47
Age	13,248	33.61	11.97	
Beauty	15,540	3.68	0.67	
Bravery	15,540	3.61	0.60	
Creativity	15,540	3.67	0.72	
Curiosity	15,540	3.88	0.60	
Fairness	15,540	3.98	0.55	
Forgiveness	15,540	3.60	0.66	
Gratitude	15,540	3.84	0.62	
Honesty	15,540	3.92	0.52	
Hope	15,540	3.71	0.67	
Modesty	15,540	3.37	0.64	
Humor	15,540	3.73	0.66	
Kindness	15,540	3.90	0.57	
Leadership	15,540	3.75	0.59	
Love	15,540	3.91	0.59	
Learning	15,540	3.75	0.66	
Judgment	15,540	3.87	0.57	
Perspective	15,540	3.74	0.58	
Perseverance	15,540	3.68	0.69	
Prudence	15,540	3.56	0.60	
Self-regulation	15,540	3.41	0.61	
Social IQ	15,540	3.77	0.58	
Spirituality	15,540	3.34	0.87	
Teamwork	15,540	3.78	0.58	
Zest	15,540	3.63	0.65	

Notes. HS = High school; Social IQ = Social Intelligence.

Given small- to medium-sized differences across the groups in demographic variables, additional analyses were conducted to evaluate whether these differences could account for any of the findings reported below. Specifically, scores on the 24 strength scales were regressed onto age and gender (education was omitted because of cultural differences in educational systems and excessive missing data). Residuals were then used in the same MGCFA analyses described below. The results were almost exactly the same as those described in the Results section, so it was concluded that the instances of nonequivalence described below could not be accounted for by demographic variations across the groups.

The following goodness-of-fit statistics were used in this study. The chi-square test is provided though it is often misleading as an index of model fit. A nonsignificant outcome would indicate convergence between the sample covariance matrix and the covariance matrix estimated from the model. In practice, though, the power resulting from large sample sizes means the test is almost always significant even if fit is good by other standards (Bentler & Bonnet, 1980). Accordingly, the chi-square statistic was used to evaluate relative fit across translations, not overall fit.

The root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and comparative fit index (CFI) are dimensional indicators of model goodness of fit scaled to range from 0 to 1. The RMSEA indicates how well the model estimates the covariance matrix under optimal conditions, with values of .07 or less considered desirable (Steiger, 2007). The RMSEA is particularly popular because it allows for the computation of a confidence interval. The SRMR similarly evaluates the degree of discrepancy between model estimates and actual covariances, with values of .08 or less considered desirable (Hu & Bentler, 1999). The CFI indicates the degree to which the estimated model is superior to a model in which all measured variables are assumed to be uncorrelated. Values of .95 or greater are considered indicative of good fit (Hu & Bentler). The final two indices used in this study, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), are unconstrained in value, and there is no value that can be considered a benchmark for adequate fit. However, AIC and BIC values for different models can be compared, with smaller values indicating better fit.

Results

Preliminary Analyses

Before conducting cross-country analyses, several models were evaluated in terms of their fit to the US group only (see the Preliminary Analyses section of Table 3). The first model (M0) represented the five-factor model of the 24 character strengths McGrath (2014a) derived using exploratory factor analytic methods. Table 2 summarizes the factors and the strength scales that were primarily associated with each factor in that earlier study. In the M0 model, each scale was associated with the factor listed in Table 2, and unstandardized factor loadings were fixed to 1.0 for the scale that was most strongly related to each factor. Covariances between factors were allowed to vary freely, but all residual covariances between scales were fixed to zero.

As expected given the restricted set of loadings in this model compared to exploratory factor analysis, M0 demonstrated inadequate fit; only the SRMR met the benchmark mentioned above. Because alignment does not allow cross-loadings, the model was loosened by freeing residual covariances based on modification indices for M0. The first iteration freed error covariances associated with modification indices of 40 or greater. This model also did not achieve acceptable fit. Subsequent models increased the number of covariances to be estimated by freeing those covariances associated with modification indices that were > 30, then > 20, and finally > 10. Only this last model met

Table 2. Factor-scale associations in the present study

Factor/scale	Structure matrix loading
Interpersonal	
Fairness	.82
Kindness	.75
Teamwork	.74
Modesty	.58
Leadership	.73
Forgiveness	.62
Emotional	
Social IQ	.78
Humor	.65
Bravery	.73
Creativity	.64
Perspective	.74
Restraint	
Prudence	.72
Perseverance	.72
Self-regulation	.67
Judgment	.68
Honesty	.65
Theological	
Zest	.80
Hope	.77
Gratitude	.74
Spirituality	.59
Love	.63
Intellectual	
Learning	.78
Beauty	.63
Curiosity	.72

Note. Structure matrix loadings are from McGrath (2014a).

acceptable criteria for fit. It involved estimating 126 covariances that had been set to zero in the M0 model, of which 25 (20%) involved two scales already linked by a shared factor. This model was used for the subsequent analyses.

MultiGroup Confirmatory Factor Analyses

Mplus was used to evaluate three models involving increasingly stringent definitions of measurement invariance (Bontempo & Hofer, 2007). The first model assumed configural invariance, which is a precondition for cross-country generalizability. Configural invariance occurs when the same factor loadings are estimated in each country (i.e., the linkages between the variables and the factors are the same in each country), but the factor loadings are estimated separately in each country. In addition, the intercepts for the equations that result from regressing the observed variable on the factors are also allowed to vary across countries (note that since cross-loadings were prohibited, these regression equations always involved a single factor as the predictor).

The second model evaluated metric invariance, in which factor loadings were assumed invariant across groups. Metric invariance is considered the necessary minimum condition for concluding that items are being interpreted equivalently across groups, because it suggests the relationship between the factor and the observed variable is the same across all groups. The third model evaluated scalar invariance, which requires equivalence in both the factor loadings and intercepts. Scalar invariance, which is considered evidence of strong measurement invariance, suggests that values on the manifest variables can be directly compared across groups: differences between scores from two groups mean the same thing as differences between scores involving two members of the same group. Results for the three models may be found in Table 3.

As could be expected, the configural invariance model fit the data the best of the three. The RMSEA value was still low, though it exceeded the desired value of .07; the SRMR and CFI values were acceptable; and the minimum for the AIC was associated with this model.

However, there are several indicators that the metric invariance model involved relatively little change in fit. The RMSEA value actually declined by .003, the SRMR value increased by only .023, and the CFI value dropped by .005. All three of these change statistics are within the range that have been suggested as indicative of equivalent fit (Cheung & Rensvold, 2002). Though the AIC value increased slightly, the BIC value for the metric model was the lowest for any of the three models.

In contrast, the AIC and BIC values for the scalar model were higher than for either of the other models. None of the other fit statistics met their benchmark values, and the change in RMSEA, CFI, and SRMR is all consistent with a decline in fit. However, all three were still near the criterion for adequate fit. The results provide reasonable evidence for equivalence across translations in factor loadings, but there does seem to be some instances where the scale intercepts differ.

The chi square for each model can be decomposed into the contribution for each group. This makes it possible to use the chi square to evaluate relative fit of the model across the translations. The values below the horizontal line in Table 3 provide each group's relative contribution to problems in fit. For example, the chi-square value for Brazil in the configural model is 7% of the total chi-square value for the configural model (15,977.95). Since the chi-square value is increased by poor fit, larger values here would indicate a greater contribution to poor fit in the model as a whole. Since the model was developed to fit the US sample, it is not surprising to find fit was optimized in this group. What is noteworthy is the absence of any country that accounted for a sizeable portion of the overall poor fit of the model. Across the three models, no country except France in the metric model accounted for more than 8% of the total chi-square value: the chi square for any country was at worst twice that for the US sample. To put this in perspective, a value of .08 means the chi square for that country was around 1,200, the same chi square found when only modification indices > 20 were loosened in the US sample. Another way to think about this is that if the model were equally appropriate to all 16 countries, these proportions would have been expected to be .06–.07, hardly any

Table 3. Multigroup confirmatory factor analysis results

Model	X^2	*df*	RMSEA	CI	SRMR	CFI	AIC	BIC
Preliminary analyses								
M0	3,837.34	242	.12	[.12, .13]	.08	.75	32,751.85	33,154.29
MI > 40	1,914.76	203	.09	[.09, .10]	.07	.88	30,907.27	31,501.10
MI > 30	1,510.47	183	.09	[.08, .09]	.06	.91	30,542.98	31,234.97
MI > 20	1,206.63	165	.08	[.08, .08]	.06	.93	30,275.14	31,055.47
MI > 10	696.68	116	.07	[.07, .08]	.05	.96	29,863.19	30,884.00
MGCFA								
Configural	15,977.95	1,856	.09	[.09, .09]	.05	.95	430,263.35	455,726.45
Metric	17,495.37	2,141	.09	[.09, .09]	.08	.94	431,210.77	454,493.29
Scalar	28,582.36	2,426	.11	[.10, .11]	.09	.90	441,727.76	462,829.69

Country	Configural	Metric	Scalar
USA	.04	.04	.04
Brazil	.07	.07	.06
Denmark	.05	.05	.07
France	.06	.09	.07
Germany	.07	.07	.06
Hong Kong	.08	.08	.07
Israel	.06	.06	.05
Italy	.06	.06	.05
Japan	.07	.07	.08
Korea	.08	.07	.07
Mainland China	.08	.07	.09
Netherlands	.05	.05	.05
Portugal	.06	.06	.06
Spain	.06	.06	.06
Sweden	.06	.06	.07
Turkey	.04	.04	.05

Notes. CI = RMSEA confidence interval from Mplus. M0 is the baseline model derived from McGrath (2014a). MI = modification indices; MGCFA = multigroup confirmatory factor analysis. Values in the lower half of the table (e.g., .06, .09, and .07 for France) represent the proportion of the total chi-square value for that model accounted for by each country.

difference from what was actually observed. When these analyses were replicated applying the M0 model to the 16 countries (so the model was not optimized for the US sample), only two of the proportionate chi-square contributions reached .08, and some countries demonstrated chi-square values lower than the US.

Given its status as the only Muslim country in the sample, the relatively low contribution of Turkey to the chi-square values is particularly striking. With the exception of the aberrant value for France in the metric model, the countries where the US-derived model seemed to fit most poorly were restricted to East Asia: Hong Kong, Japan, Korea, and mainland China.

Based on evidence of at least metric invariance, another confirmatory factor analysis was conducted for the entire sample without consideration of nationality for purposes of estimating factor covariances. Fit statistics for this overall model were consistent with those for the metric invariance model (RMSEA = .09, SRMR = .05, CFI = .96). Factor correlations are provided in Table 4 and are quite large even though McGrath (2014a) found substantial agreement between oblique and orthogonal rotations in the study on which this model was based.

Table 4. Correlations between factors in the total sample

	Interpersonal	Emotional	Restraint	Theological
Emotional	.76			
Restraint	.83	.81		
Theological	.81	.87	.75	
Intellectual	.63	.74	.61	.78

Alignment Analysis

Factor alignment analysis was specifically developed for Mplus. The procedure begins by estimating a configural model in which all loadings and intercepts are allowed to vary freely across groups, with factor score means set to 0 and factor score variances to 1 for all groups. Mplus then loosens these restrictions on the means and variances as necessary to allow for increasing convergence (alignment) in the factor loadings and intercepts until a model is obtained that is as close to scalar invariance as possible. The result is a model that is heuristically developed to maximize equivalence in the intercepts and factor loadings

Table 5. Alignment analysis results

	Intercepts			Loadings		
Factor/scale	Fit function	Var	# Invariant groups	Fit function	Var	# Invariant groups
Interpersonal						
Fairness	−41.71	0.003	11	−28.15	0.001	15
Kindness	−56.78	0.003	13	−32.27	0.001	15
Teamwork	−72.57	0.002	12	−40.69	0.001	16
Modesty	−38.75	0.011	9	−26.21	0.004	14
Leadership	−38.14	0.004	12	−23.77	0.001	15
Forgiveness	−37.43	0.008	11	−23.33	0.002	15
Emotional						
Social IQ	−32.39	0.007	12	−36.35	0.001	16
Humor	−30.90	0.014	8	−38.50	0.001	16
Bravery	−67.09	0.003	14	−28.02	0.003	13
Creativity	−59.19	0.001	15	−24.92	0.004	13
Perspective	−52.10	0.011	8	−19.69	0.001	16
Restraint						
Prudence	−55.23	0.013	7	−27.72	0.002	16
Perseverance	−49.29	0.000	16	−50.70	0.001	16
Self-Regulation	−12.31	0.009	8	−27.17	0.002	15
Judgment	−67.77	0.004	11	−34.89	0.009	12
Honesty	−56.37	0.007	7	−33.50	0.002	15
Theological						
Zest	−44.33	0.015	9	−29.82	0.001	16
Hope	−41.72	0.005	12	−24.60	0.001	16
Gratitude	−37.50	0.005	11	−42.86	0.001	15
Spirituality	−77.20	0.039	5	−37.46	0.009	15
Love	−58.30	0.002	14	−22.20	0.005	14
Intellectual						
Learning	−73.54	0.004	15	−37.71	0.005	15
Beauty	−58.48	0.033	6	−12.39	0.004	16
Curiosity	−22.61	0.016	8	−40.72	0.000	16

Notes. Var = variance of group estimates; Social IQ = Social Intelligence.

across groups. However, the alignment analysis can identify which parameters (factor loadings and intercepts) in which groups differ significantly from the other groups. That is, it provides complementary information to the multigroup confirmatory factor analysis. Where the latter gives you overall information about how well the model applies to the groups, alignment analysis tells you specifically which parameters in which groups are most inconsistent with the overall findings.

Initially, all factor score means were allowed to vary from 0 once the configural model was estimated. However, the results were poorly identified, which the developers noted as a common problem for alignment analysis when all factor means are allowed to vary (Asparouhov & Muthén, 2014). This problem was resolved by fixing the factor means for Turkey to 0, chosen because the average of the initial factor mean estimates for the Turkish group was closer to 0 than that of any other group. The number of random starts was also increased to 100 to achieve replication of global minima for the loss function.

Table 5 provides results from the alignment analyses. Three statistics are provided concerning the degree of equivalence demonstrated in intercept and loading estimates for each of the 24 strengths. Just as the chi-square results for the MGCFA models were decomposed by group, Mplus reports the fit function decomposed by manifest variable. In this case, smaller fit function values are indicative of greater equivalence. The variance of the parameter estimates across groups similarly indicates the degree of convergence, with smaller values indicating less variability from one nation to another. Finally, the number of invariant groups identifies the number of groups for which parameters were consistent. In this case higher values are better. Ideally this number would equal 16, indicating perfect consistency across the groups.

Consistent with earlier findings that the evidence for metric equivalence is stronger than that for scalar invariance, the intercept fit values and variances are consistently larger than those for the loadings and the number of invariant groups is consistently smaller, suggesting more variability across groups in intercept estimates. However, even for the loadings the fit function and variance values were not large in an absolute sense, supporting the earlier conclusion that the data approach but do not achieve scalar equivalence.

The results for Spirituality were particularly marked. This is interesting given evidence that Spirituality demonstrates the most variability both across and within countries (McGrath, 2014b). This instance will be used in the Discussion section to demonstrate some of the implications of these findings.

In contrast to the intercept statistics, fit function values for the loadings are substantially smaller, and not one of the variances exceeds .01. Perhaps most telling is that loadings approached complete invariance across groups for most of the strength scales: only five of 24 scales did not achieve invariance in at least 15 groups.

Discussion

The purpose of this study was to evaluate whether different translations of the VIA-IS, completed by respondents in different countries, demonstrated measurement invariance. This was evaluated by first developing a latent variable model for the VIA-IS that was consistent with standards for adequate fit in a sample of US residents who completed the inventory in English. This model was then applied simultaneously to 15 other countries to evaluate configural, metric, and scalar equivalence.

What are the implications of configural, metric, and scalar equivalence? Configural equivalence indicates the same placement of scales within factors is meaningful across the translations. Metric equivalence implies the impact of those factors on the scale scores is also consistent. Scalar equivalence implies differences in the scores across translations is a function of differences in the factor scores. Metric equivalence is the minimum condition for concluding that respondents across translations are understanding the meaning and implications of the items in similar ways. Both scalar and metric invariance are needed to assume that the scores on the same scale are comparable across translations, though only metric invariance is needed if the goal is to compare respondents across translations on differences in scores (e.g., on rank orderings).

The findings for Spirituality and Beauty can be used to demonstrate the implications. Both scales demonstrated equivalent loadings in almost every group. What this suggests is that respondents across most translations interpreted the items about spirituality in the context of a cluster of strengths reflecting what might be called a sense of grace or transcendence, while the items comprising the Appreciation of Beauty scale were seen as more reflective of intellectual interests. However, the intercept for the equation describing these strengths as a function of those latent variables had to vary from translation to translation in order to achieve adequate fit. That finding implies scores on the Spirituality scale varied across translations for reasons other than the sense of transcendence, and the appreciation of beauty was influenced by other culturally variations than general intellectualism. Accordingly, Spirituality and Beauty scores can vary for respondents from different cultures for reasons that are not well understood through this model. That said, it is still reasonable to conclude that if a person demonstrates a higher score for Spirituality over Beauty in the Japanese group, that ordering occurs for the same reasons it would in a member of the US group.

The findings provide support for the presence of at least configural and metric equivalence across nations and translations. The evidence for scalar equivalence is weaker, but still approaches acceptable levels in a number of instances. For example, most of the scales comprising the Interpersonal factor achieved intercept invariance across most translations (with Modesty as an outlier) as well as loading invariance. In fact, only four of the 24 scales achieved intercept invariance in less than half of the groups (Prudence, Honesty, Spirituality, and Beauty). These findings suggest that in most cases responses across translations demonstrated a common underlying structural model, and in many instances direct comparison of scores is also reasonable. The results in Table 5 provide guidance on which scales are most similar in their interpretation across translations, with Perseverance, Creativity, and Learning in particular standing out from the rest.

The primary potential limitation of this study has to do with the method of data collections. Questions may be raised about the adequacy of unfiltered Internet samples, though several studies have now concluded that this tends not to be a more serious issue than is true for other common methods of data collection (Gosling, Vazire, Srivastava, & John, 2004; Weigold, Weigold, & Russell, 2013). For example, ready access to the Internet is not necessarily typical of all the cultures represented in this study, but the same bias would be evident if data collection were based largely on college students. A second issue that potentially limits the generalizability of the findings is the large number of residual covariances that were allowed to vary to achieve adequate model fit. Since the selection was based on modification indices for this sample, the selection could have been different with a different sample. The likelihood of this problem is somewhat mitigated by the size of the current sample, however.

Given the degree of support for at least metric invariance and even scalar invariance, it is important to keep in mind that this study could have underestimated the degree to which the 24 strengths are recognizable across cultures. Most translations were created by research teams working independently of each other, and they did not account for cross-cultural differences in how the strengths tend to manifest themselves. Furthermore, while data collection via the Internet assured some homogeneity in terms of access to technology across the groups, the groups also demonstrated substantial differences across descriptive variables. Taken together, the findings from this study are consistent with the cross-cultural relevance of the VIA Classification of character strengths.

Acknowledgments

I am grateful to the VIA Institute on Character for access to the data used in the conduct of this study.

References

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling*. Advance online publication. doi: 10.1080/10705511.2014.919210

Bentler, P. M., & Bonnet, D. C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Biswas-Diener, R. (2006). From the equator to the North Pole: A study of character strengths. *Journal of Happiness Studies, 7*, 293–310. doi: 10.1007/s10902-005-3646-8

Bontempo, D. E., & Hofer, S. M. (2007). Assessing factorial invariance in cross-sectional and longitudinal studies. In A. D. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology* (pp. 153–175). Oxford, UK: Oxford University Press.

Brdar, I., & Kashdan, T. B. (2010). Character strengths and well-being in Croatia: An empirical investigation of structure and correlates. *Journal of Research in Personality, 44*, 151–154.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.

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. doi: 10.1037/1089-2680.9.3.203

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*, 93–104.

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.

Littman-Ovadia, H., & Lavy, S. (2012). Character strengths in Israel: Hebrew adaptation of the VIA Inventory of Strengths. *European Journal of Psychological Assessment, 28*, 41–50.

Macdonald, C., Bore, M., & Munro, D. (2008). Values in Action scale and the big 5: An empirical indication of structure. *Journal of Research in Personality, 42*, 787–799.

McGrath, R. E. (2014a). Scale- and item-level factor analysis of the VIA Inventory of strengths. *Assessment, 21*, 4–14.

McGrath, R. E. (2014b). Character strengths in 75 nations: An update. *Journal of Positive Psychology, 10*, 41–52. doi: 10.1080/17439760.2014.888580

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

Peterson, C., Park, N., Pole, N., D'Andrea, W., & Seligman, M. E. P. (2008). Strengths of character and posttraumatic growth. *Journal of Traumatic Stress, 21*, 214–217.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A classification and handbook*. Washington, DC: American Psychological Association.

Ruch, W., Proyer, R. T., Harzer, C., Park, N., Peterson, C., & Seligman, M. E. 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.

Shryack, J., Steger, M. F., Krueger, R. F., & Kallie, C. S. (2010). The structure of virtue: An empirical investigation of the dimensionality of the Virtues in Action Inventory of Strengths. *Personality and Individual Differences, 48*, 714–719.

Singh, K., & Choubisa, R. (2010). Empirical validation of Values in Action-Inventory of Strengths (VIA-IS) in Indian context. *National Academy of Psychology India Psychological Studies, 55*, 151–158.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*, 893–898.

Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods, 18*, 53–70.

Date of acceptance: October 12, 2014
Published online: May 21, 2015

Robert McGrath

School of Psychology
Fairleigh Dickinson University
Teaneck, NJ 07555
USA
E-mail mcgrath@fdu.edu

