Psychometric Properties of Measures of Team Diversity With Likert Data

Educational and Psychological
Measurement

2015, Vol. 75(3) 512–534

© The Author(s) 2014

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164414541275

epm.sagepub.com

Lifang Deng¹, George A. Marcoulides², and
Ke-Hai Yuan³

Abstract

Certain diversity among team members is beneficial to the growth of an organization. Multiple measures have been proposed to quantify diversity, although little is known about their psychometric properties. This article proposes several methods to evaluate the unidimensionality and reliability of three measures of diversity. To approximate the interval scale required by the measures of diversity, a transformation on the Likert-item scores is proposed. Ridge maximum likelihood is used to deal with the issue of small sample size, and methods for evaluating the significance of the difference of two reliability estimates with correlated samples are also developed. Results with a real data set on entrepreneurial teams indicate that different measures of diversity may correspond to significantly different estimates of reliability. Results also indicate that diversity measures obtained with the transformed data tend to be more unidimensional than their counterparts obtained from Likert data. However, diversity measures obtained from Likert data tend to yield greater reliability estimates. Among the three examined measures of diversity, the standard deviation is found to yield greater and more efficient reliability estimates than the others and is thus recommended.

Keywords

unidimensionality, reliability, normal-curve transformation, ridge structural equation modeling

¹Beihang University, Beijing, China

²University of California, Santa Barbara, CA, USA

³University of Notre Dame, Notre Dame, IN, USA

Corresponding Author:

Ke-Hai Yuan, University of Notre Dame, 123a Haggar Hall, Notre Dame, IN 46556, USA.

Email: kyuan@nd.edu

Introduction

The compositional diversity of team members within an organization has been shown to affect the performance and growth of the organization (Harrison & Klein, 2007; Van Knippenberg, De Dreu, & Homan, 2004). Among various kinds of diversity (e.g., demographic, informational, experiential, or personality attributes), not all have been determined to be beneficial to the growth of an organization. Some researchers have indicated that it is merely the differences between team members in terms of their skill level, knowledge, and perspectives that are needed to foster creativity and innovation (e.g., Guzzo & Shea, 1992). However, the findings in the extant literature have not been consistent (Jackson, Joshi, & Erhardt, 2003; Stewart, 2006; Van Knippenberg et al., 2004; Webber & Donahue, 2001) and indicate that our understanding of diversity and its role is still relatively limited. To facilitate a better understanding of diversity, Harrison and Klein (2007) classified diversity into three distinctive types: *separation*, *variety*, and *disparity*. Separation is for the difference in position or opinion among team members; variety is used to describe diversity in expertise, knowledge, or experience; and disparity refers to inequality in status or resources held among team members. Such a classification allows researchers to identify different roles of different types of diversity.

A variety of measures have also been proposed to quantify different types of diversity among individuals within a team. According to Harrison and Klein (2007), separation should be measured by either the standard deviation or the average of Euclidean distances, variety should be measured by the so-called Blau's (1977) index or entropy (Teachman, 1980), and disparity should be measured by the coefficient of variation or the ratio of the average of the absolute differences over the mean. Harrison and Klein (2007) also discussed the type of scales required by each of these diversity measures and emphasized that measuring separation requires the observed data on team members to be at the interval scale, whereas measuring disparity requires the observed data to be at the ratio scale. However, in the study of human behavior within the fields of education, management, psychology, and related social and behavioral sciences, it is extremely difficult to obtain data at the ratio or even interval scales. What are typically obtained are data collected from a survey using questionnaires that are commonly only ordinal or Likert type. Nevertheless, researchers still regularly apply procedures that require interval-scale data to ordinal data. For example, factor analysis is commonly applied to Likert data for item selection or scale development (Raykov & Marcoulides, 2011). Although such a practice may still yield interpretable results, a better method is to factor analyze the polychoric correlation matrix (Babakus, Ferguson, & Jöreskog, 1987). Given that the observed values in Likert data used to determine the above-mentioned diversity measures are somewhat arbitrary, we propose to transform them to avoid the arbitrariness. The transformation is based on threshold values under the normal curve, parallel to those used in estimating polychoric correlations (Olsson, 1979). We can call it the normal-curve (NC) transformation. Although the transformed data are still limited in number of values, we argue that they are closer to the conditions required by diversity

measures than the commonly used Likert data. To see the effect of the transformation, we will study the psychometric properties of several diversity measures when applied to both Likert and transformed data.

Unidimensionality and reliability are probably the two most basic psychometric properties one has to consider for any scale or instrument. Unidimensionality implies that the statistical dependence among the items can be accounted for by a single underlying latent trait, and reliability informs about the degree to which the observed individual differences are indicative of true individual differences on the latent dimension of interest. Measures of diversity, especially those for measuring separation with Likert data, are also subject to such properties if they aim to properly capture any underlying trait. In particular, when measurements in a scale are not unidimensional, the empirical meaning of the scale will be different from the meaning assigned to it, which will create interpretational confounding (e.g., Anderson & Gerbing, 1982; Bagozzi, 1980; Burt, 1973, 1976). Reliability is equally important because, for measurements with a low reliability index, the observed values of the obtained measurements can be mostly due to random errors. Additionally, because the value of the determined reliability index sets a bound on validity (Allen & Yen, 1979), a high reliability (index) is a necessary condition for high validity (Raykov & Marcoulides, 2011). We hope that by studying the unidimensionality and reliability of different measures of diversity, the inconsistent findings obtained to date on the roles of diversity can be better understood.

The methodological development presented in this article was motivated by the need to study the psychometric properties of diversity measures based on 13 Likert items administered to entrepreneurial teams. Because the number of teams plays the role of sample size, which is not sufficiently large, a method to deal with the issue of small sample sizes was also needed especially when using factor analysis to evaluate the unidimensionality of the diversity measures. For such a purpose, we make use of the ridge maximum likelihood (ML) method originally developed in Yuan and Chan (2008). This method has been shown to yield more accurate parameter estimates than the normal-distribution-based maximum likelihood (NML) even for normally distributed data. We also develop methods for evaluating the significance of the difference of two reliability estimates with correlated samples. This enables us to determine whether different measures of diversity correspond to significantly different reliability estimates. If different diversity measures yield significantly different reliability estimates, then it is better to use the one that corresponds to the greatest reliability.

In the next section, the methodological components for studying the unidimensionality and reliability of different measures of diversity are given, including the formulations of different diversity measures, the NC transformation, ridge ML, and standard error (SE) for difference of reliability estimates. A real data set with Likert scale and its analysis are presented in the following section. We conclude with a discussion and recommendations. It is important to note that our focus is on the psychometric properties (unidimensionality and reliability) of different measures of

diversity, not on interrater reliability issues (for further details on interrater reliability, see Algina, 1978; Schuster & Smith, 2002; Shrout & Fleiss, 1979).

Methodology

This section first introduces the three diversity measures that will be used in the analysis of the real data. Then, the NC transformation is described. Ridge ML for factor analysis is reviewed next. Formulas for SE of the difference of two reliability estimates are developed at the end of this section. These measures and techniques will be used to analyze the real data in the subsequent section.

Diversity Measures

Let x_{ijk} be the score of person k on item j within team i , $k=1, 2, \dots, n_i$; $j=1, 2, \dots, p$; $i=1, 2, \dots, N$. Three measures of diversity derived from x_{ijk} will be studied. These are the average of absolute distances (aad) among team members,

$$\text{aad}_{ij} = \frac{2}{n_i(n_i - 1)} \sum_{k_1=1}^{n_i-1} \sum_{k_2=k_1+1}^{n_i} |x_{ijk_1} - x_{ijk_2}|; \quad (1)$$

the average of absolute deviations from the mean (aadm) of team members,

$$\text{aadm}_{ij} = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} |x_{ijk} - \bar{x}_{ij}|, \quad (2)$$

where $\bar{x}_{ij} = \sum_{k=1}^{n_i} x_{ijk} / n_i$; and the standard deviation (sd) among team members,

$$\text{sd}_{ij} = \left[\frac{1}{n_i - 1} \sum_{k=1}^{n_i} (x_{ijk} - \bar{x}_{ij})^2 \right]^{1/2}. \quad (3)$$

Two measures for separation were recommended by Harrison and Klein (2007). One is the standard deviation in which the denominator is n_i instead of $n_i - 1$. Another is the square root of the average of the squared Euclidean distances $(x_{ijk_1} - x_{ijk_2})^2$, in which $k_1 = k_2$ is not distinguished from $k_1 \neq k_2$. According to Biemann and Kearney (2010), these measures may contain substantial bias due to including terms that are obviously 0 or without correcting for the loss of degrees of freedom. The diversity measure in (1) only includes the absolute distances for different team members, and degrees of freedom loss are accounted for in (2) and (3). Parallel to the average Euclidean distance (aed) in Harrison and Klein (2007) or Biemann and Kearney (2010), we define aed_{ij} as

$$\text{aed}_{ij} = \left[\frac{2}{n_i(n_i - 1)} \sum_{k_1=1}^{n_i-1} \sum_{k_2=k_1+1}^{n_i} (x_{ijk_1} - x_{ijk_2})^2 \right]^{1/2}.$$

Because aed_{ij} is proportional to sd_{ij} (see Hays, 1981) and any results of reliability and unidimensionality analysis of aed_{ij} would be identical to those of sd_{ij} , we do not separately examine aed_{ij} in this article.

As we were not able to locate any references in the literature in which the $aadm_{ij}$ in (2) or an index that is proportional to $aadm_{ij}$ has been proposed to measure diversity, $aadm_{ij}$ can be regarded as a new measure. The psychometric properties of the three measures, aad , $aadm$, and sd , will be examined through real data analysis in the following section.

Quantities in the form of the average of absolute distances (e.g., aad) are not presented as stand-alone measures in either Harrison and Klein (2007) or Biemann and Kearney (2010). Instead, they are divided by the team mean score for measuring disparity. Another measure for disparity recommended by Harrison and Klein (2007) is the coefficient of variation. Since these measures require the observed data to possess the properties of ratio scale, they may not be applicable to Likert data and will not be studied in this article. Similarly, variety will not be measured through Likert data and neither do we study Blau's index or the entropy.

Normal-Curve Transformation

Since all three measures of diversity (aad , $aadm$, sd) are obtained by arithmetic operations, they are ideally applicable to data that are of interval scale (Harrison & Klein, 2007). However, as indicated previously, measurements in the social and behavioral sciences are typically Likert or ordinal scale. To approximate interval scales, we propose a transformation to Likert data in this subsection.

With a total of $N_l = \sum_{i=1}^N n_i$ individual observations and c categories for a given item, let $\hat{q}_l$ be the proportion of observations¹ for category l . Following the convention of polychoric correlations (Olsson, 1979), we may assume that, for each observed x_{ijk} , there is an underlying continuous variable $z_{ijk} \sim N(0, 1)$ such that $x_{ijk} = l$ whenever z_{ijk} belongs to the interval $(h_{l-1}, h_l]$, where $h_0 < h_1 < \dots < h_c$ are threshold values to be estimated. This implies that the probability for $x_{ijk} = l$ is given by

$$q_l = \Phi(h_l) - \Phi(h_{l-1}), \quad l = 1, 2, \dots, c,$$

where $\Phi(\cdot)$ is the cumulative distribution function of $z \sim N(0, 1)$, with $h_0 = -\infty$ and $h_c = \infty$. Thus, the marginal maximum likelihood estimate of h_l is given by

$$\hat{h}_l = \Phi^{-1}\left(\sum_{t=1}^l \hat{q}_t\right), \quad l = 1, 2, \dots, c-1. \quad (4)$$

Based on this underlying NC assumption, we propose a transformation to the Likert x_{ijk} by

$$y_{ijk} = (\hat{h}_{l-1} + \hat{h}_l)/2 \quad \text{if } x_{ijk} = l, \quad l = 1, 2, \dots, c. \quad (5)$$

Notice that there are only $c - 1$ finite values of $\hat{h}_l$ in (4), and we cannot use $\hat{h}_0 = -\infty$ or $\hat{h}_c = \infty$ because they will result in $y_{ijk} = -\infty$ when $x_{ijk} = 1$ or $y_{ijk} = \infty$ when $x_{ijk} = c$. We further propose to use

$$\hat{h}_0 = \Phi^{-1}(.5/N_t) \quad \text{and} \quad \hat{h}_c = \Phi^{-1}(1 - .5/N_t). \quad (6)$$

The proposed values in (6) are equivalent to assigning a value of .5 to cells with zero number of observations in the analysis of contingency tables, because we can think of an extra category $x_{ijk} = 0$ below $x_{ijk} = 1$ and another extra category $x_{ijk} = c + 1$ above $x_{ijk} = c$, and both had zero number of observations. The proposed values in (6) are also similar to the so-called continuity correction in applying the central limit theorem to categorical data (Feller, 1945), where a step of .5 is used when jumping from 1 to the next whole number.

Notice that the correction in (6) is for y_{ijk} to avoid being $-\infty$ or ∞ whenever $x_{ijk} = 1$ or c . If the nominal number of categories is c but only $c - 1$ or fewer number of categories are observed, we may simply treat the unobserved categories in the middle as having probability of zero by just applying the correction to the end points of x_{ijk} .

We need to note that the transformed y_{ijk} do not possess the property of interval scales, although they avoid the arbitrary nature of Likert data that assign consecutive whole numbers to ordered categories. Closely related to polychoric correlation, the rationale of the transformation in (5) depends heavily on the assumption of a normal curve underlying the observed frequencies. If the NC assumption holds, the y_{ijk} obtained by the NC transformation determined by equations (4), (5), and (6) is simply the middle point of the interval z_{ijk} belongs, and thus represents the best prediction of the true value of z_{ijk} in the sense of smallest absolute mean difference.

Applying each of the three diversity measures, aad_{ij} , $aadm_{ij}$, and sd_{ij} , to the transformed y_{ijk} yields three more measures of diversity. In the next section, their reliability and unidimensionality are examined, and the results are contrasted with those obtained based on Likert data.

Ridge Maximum Likelihood for Factor Analysis With Small Sample Sizes

As indicated in the previous section, the number of teams, N , plays the role of sample size when evaluating the psychometric properties of the diversity measures aad , $aadm$, and sd . Since it can be expensive to have a large N , we use ridge ML for factor analysis of the diversity measures in (1) to (3) when studying their unidimensionality. Unless all the n_i are sufficiently large, the diversity measures in (1) to (3) cannot be regarded as normally distributed. As such, we expect ridge ML to work better than NML when factor analyzing the diversity measures.

Let $\mathbf{S}$ be a sample covariance matrix of size p , and we are interested in modeling $\mathbf{\Sigma} = E(\mathbf{S})$ by a confirmatory factor model

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Lambda}\Phi\mathbf{\Lambda}' + \mathbf{\Psi}, \quad (7)$$

where $\mathbf{\Lambda}$ is a factor loading matrix, $\mathbf{\Phi}$ is a factor correlation matrix, and $\mathbf{\Psi}$ is a diagonal matrix of measurement errors/uniquenesses. The widely used NML procedure for covariance structure analysis is to minimize

$$F_{ML}(\mathbf{S}, \mathbf{\Sigma}(\boldsymbol{\theta})) = \text{tr}[\mathbf{S}\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})] - \log |\mathbf{S}\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})| - p$$

for parameter estimation. Let $a > 0$ be a small number and $\mathbf{S}_a = \mathbf{S} + a\mathbf{I}$, with $\mathbf{I}$ being the identity matrix. The ridge ML developed in Yuan and Chan (2008) is to estimate $\boldsymbol{\theta}_a$ by minimizing $F_{ML}(\mathbf{S}_a, \mathbf{\Sigma}(\boldsymbol{\theta}_a))$, and let the estimates be denoted by $\hat{\boldsymbol{\theta}}_a$. The corresponding estimates $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ are obtained by subtracting a from each of the elements of $\hat{\boldsymbol{\theta}}_a$ corresponding to the diagonal elements of $\mathbf{\Psi}$, leaving the other elements of $\hat{\boldsymbol{\theta}}_a$ unchanged. Standard errors of $\hat{\boldsymbol{\theta}}$ are obtained by a sandwich-type covariance matrix, which accounts for the unknown underlying population distribution of the involved diversity measure. As for overall model evaluation, Yuan and Chan (2008) showed that, unless $a = 0$, $T_{ML} = (N - 1)F_{ML}(\mathbf{S}_a, \mathbf{\Sigma}(\boldsymbol{\theta}_a))$ does not asymptotically follow the nominal chi-square distribution χ^2_{df} even if data are normally distributed. They developed a rescaled statistic T_{RML} and an adjusted statistic T_{AML} . Parallel to the development for NML in Satorra and Bentler (1994), T_{RML} asymptotically follows a distribution whose mean equals df , and T_{AML} asymptotically follows a distribution whose mean and variance equal those of the approximating distribution. Since the details of ridge ML have already been described in Yuan and Chan (2008), no further elaboration is given here. Our purpose is to apply ridge ML to evaluate the unidimensionality of each of the three measures of diversity in (1) to (3) and to determine whether the corresponding sample covariance matrix can be reasonably fitted by a one-factor model. Following the recommendation of Yuan and Chan (2008), $a = p/N$ is used in applying the ridge ML.

In order to fully justify applying a factor analysis to each of the diversity measures, we do not need to assume that each of aad_{ij} , aadm_{ij} , or sd_{ij} is identically distributed across $i = 1, 2, \dots, N$. The development in Lee and Shi (1998) implies that the vector $\mathbf{d}_i = (\text{aad}_{i1}, \text{aad}_{i2}, \dots, \text{aad}_{ip})'$ does not need to have the same population covariance as i varies. Since for both reliability and unidimensionality the analysis is based on the sample covariance matrix $\mathbf{S}$ of the corresponding diversity measures with the assumption $E(\mathbf{S}) = \mathbf{\Sigma}$, our study of the psychometric properties of $\mathbf{d}_i$ is for the population represented by the sample $\mathbf{d}_i$, $i = 1, 2, \dots, N$. We will further discuss this point in the concluding section.

Standard Error for Difference of Two Reliability Estimates With Correlated Samples

Among the many available estimates of reliability for equally weighted composite scores, coefficient alpha is most widely used in practice even though it can over- or underestimate the population reliability (Raykov, 1997). Another popular estimate is coefficient omega defined through the factor loadings and error variances by fitting

the sample covariance matrix to a one-factor model (McDonald, 1999). Both are applicable when evaluating the reliability of the different diversity measures. Our interest is whether different diversity measures will yield significantly different reliability estimates. Thus, we need to have an estimate of the SE of the difference of two estimates of alpha or omega. When the two estimates are independent, the variance of the difference of the two estimates is simply the summation of the variances of the two estimates of alpha or omega. However, with respect to the three diversity measures, the variance or SE of the difference of two estimates of alpha or omega depends on their correlation. Since the SE for the difference of two reliability estimates with correlated samples will facilitate comparison of reliabilities in other contexts, and the literature to date does not contain such a development, we provide more details for obtaining consistent SEs of the difference of two estimates of alpha and omega, respectively. We also present the necessary notation and formulas for calculating the SEs. The complete details leading to the calculation formulas are given in Appendices A and B.

Let $\mathbf{S} = (s_{jk})$ be a sample covariance matrix of size p , and $\mathbf{s} = \text{vech}(\mathbf{S})$ be the vector by stacking the elements in the lower-triangular part of $\mathbf{S}$. Then, with $p^* = p(p+1)/2$, $\mathbf{s}$ is a vector of $p^* \times 1$, and the sample coefficient alpha is given by

$$\hat{\alpha} = g(\mathbf{s}) = \frac{p}{p-1} \left(1 - \sum_{j=1}^p s_{jj} / \sum_{j=1}^p \sum_{k=1}^p s_{jk} \right) = \frac{p}{p-1} \left(1 - \frac{\mathbf{a}'\mathbf{s}}{\mathbf{b}'\mathbf{s}} \right),$$

where $\mathbf{a}$ is a $p^* \times 1$ vector whose elements are 1 corresponding to s_{jj} and 0 elsewhere; and $\mathbf{b}$ is also a $p^* \times 1$ vector whose elements are 1 corresponding to s_{jj} and 2 corresponding to s_{jk} when $j \neq k$. For example, at $p=3$, $\mathbf{s} = (s_{11}, s_{21}, s_{31}, s_{22}, s_{32}, s_{33})'$, $\mathbf{a} = (1, 0, 0, 1, 0, 1)'$, and $\mathbf{b} = (1, 2, 2, 1, 2, 1)'$. We need to have the Jacobian matrix or the matrix of derivatives of $g(\mathbf{s})$ with respect to the elements of $\mathbf{s}$, and it is given by

$$\dot{g}(\mathbf{s}) = -\frac{p}{p-1} \left[\frac{1}{\mathbf{b}'\mathbf{s}} \mathbf{a} - \frac{\mathbf{a}'\mathbf{s}}{(\mathbf{b}'\mathbf{s})^2} \mathbf{b} \right].$$

With $\mathbf{s}_1 = \text{vech}(\mathbf{S}_1)$ and $\mathbf{s}_2 = \text{vech}(\mathbf{S}_2)$ from two correlated samples, standard error for $\hat{\alpha}_2 - \hat{\alpha}_1 = g(\mathbf{s}_2) - g(\mathbf{s}_1)$ also involves the variance-covariance matrices of $\mathbf{s}_1$ and $\mathbf{s}_2$. Denote these by $\mathbf{\Gamma}_{11} = \text{Var}(\sqrt{n}\mathbf{s}_1)$, $\mathbf{\Gamma}_{22} = \text{Var}(\sqrt{n}\mathbf{s}_2)$, and $\mathbf{\Gamma}_{12} = \text{Cov}(\sqrt{n}\mathbf{s}_1, \sqrt{n}\mathbf{s}_2)$, where $n = N - 1$. These are consistently estimated by their sample counterparts, with details given in Appendix A. With the introduced notation, the result given in Appendix B implies that $\sqrt{n}[(\hat{\alpha}_2 - \alpha_2) - (\hat{\alpha}_1 - \alpha_1)]$ is asymptotically normally distributed with mean zero and variance consistently estimated by

$$\hat{\tau}_\alpha^2 = \dot{g}'(\mathbf{s}_1) \hat{\mathbf{\Gamma}}_{11} \dot{g}(\mathbf{s}_1) + \dot{g}'(\mathbf{s}_2) \hat{\mathbf{\Gamma}}_{22} \dot{g}(\mathbf{s}_2) - 2\dot{g}'(\mathbf{s}_1) \hat{\mathbf{\Gamma}}_{12} \dot{g}(\mathbf{s}_2). \quad (8)$$

It follows from (8) that the SE of $(\hat{\alpha}_2 - \hat{\alpha}_1)$ is consistently estimated by $\hat{\tau}_\alpha / \sqrt{n}$, which will be used in the next section when evaluating the significance of the

difference $(\hat{\alpha}_2 - \hat{\alpha}_1)$. Confidence interval (CI) for $\alpha_2 - \alpha_1$ with confidence level $(1 - 2\beta)$ can be obtained as

$$[\hat{\alpha}_2 - \hat{\alpha}_1 - c_\beta \hat{\tau}_\alpha / \sqrt{n}, \hat{\alpha}_2 - \hat{\alpha}_1 + c_\beta \hat{\tau}_\alpha / \sqrt{n}],$$

where c_β is the critical value corresponding to probability $1 - \beta$ under the standard normal curve.

We next consider the sample coefficient omega, which is defined through the estimates of a one-factor model. With p items, the covariance structure of the one-factor model can be represented by (7), where $\mathbf{\Lambda} = \boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_p)'$ is a vector of factor loadings, and $\boldsymbol{\Psi} = \text{diag}(\psi_{11}, \psi_{22}, \dots, \psi_{pp})$ is a diagonal matrix of error variances. Let $\hat{\boldsymbol{\theta}} = (\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p, \hat{\psi}_{11}, \hat{\psi}_{22}, \dots, \hat{\psi}_{pp})'$ be the ridge ML estimates for the one-factor model. Then the sample coefficient omega is given by

$$\hat{\omega} = h(\hat{\boldsymbol{\theta}}) = \frac{(\mathbf{1}'_p \hat{\boldsymbol{\lambda}})^2}{(\mathbf{1}'_p \hat{\boldsymbol{\lambda}})^2 + \text{tr}(\hat{\boldsymbol{\Psi}})},$$

where $\mathbf{1}_p$ is a vector of p 1s. With two covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$, the ridge tuning parameters a_1 and a_2 can be different. We set them equal ($a_1 = a_2 = a = p/N$) in our study and denote $\mathbf{S}_{a1} = \mathbf{S}_1 + a\mathbf{I}$ and $\mathbf{S}_{a2} = \mathbf{S}_2 + a\mathbf{I}$. Let the parameter estimates by minimizing $F_{ML}(\mathbf{S}_{a1}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_{a1}))$ and $F_{ML}(\mathbf{S}_{a2}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_{a2}))$ be denoted as $\hat{\boldsymbol{\theta}}_{a1}$ and $\hat{\boldsymbol{\theta}}_{a2}$, respectively. We need to introduce additional notation for presenting the SE of $\hat{\omega}_2 - \hat{\omega}_1 = h(\hat{\boldsymbol{\theta}}_2) - h(\hat{\boldsymbol{\theta}}_1)$.

Let $\text{vec}(\boldsymbol{\Sigma})$ be the vector of stacking all the columns of $\boldsymbol{\Sigma}$ and $\boldsymbol{\sigma} = \text{vech}(\boldsymbol{\Sigma})$. Then there exists a $p^2 \times p^*$ matrix $\mathbf{D}_p$ such that $\text{vec}(\boldsymbol{\Sigma}) = \mathbf{D}_p \text{vech}(\boldsymbol{\Sigma})$, and $\mathbf{D}_p$ is called the duplication matrix (e.g., Schott, 2005). Notice that the covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ in fitting $\mathbf{S}_1$ and $\mathbf{S}_2$ are the same, the difference between fitting the two samples are in parameter estimates. One is $\hat{\boldsymbol{\theta}}_1$ and the other is $\hat{\boldsymbol{\theta}}_2$, these are obtained by subtracting a from $\hat{\boldsymbol{\theta}}_{a1}$ and $\hat{\boldsymbol{\theta}}_{a2}$ corresponding to each error variance, respectively. Let the Jacobian matrices of $\boldsymbol{\sigma}(\boldsymbol{\theta}) = \text{vech}[\boldsymbol{\Sigma}(\boldsymbol{\theta})]$ and $h(\boldsymbol{\theta})$ be denoted by $\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}) = \partial \boldsymbol{\sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'$ and $\dot{h}(\boldsymbol{\theta})$ (see Yuan & Bentler, 2002), respectively; $\mathbf{W}(\boldsymbol{\theta}) = \mathbf{2}^{-1} \mathbf{D}_p' [\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] \mathbf{D}_p$ and $\hat{\mathbf{C}}_{aj} = \mathbf{W}(\hat{\boldsymbol{\theta}}_{aj}) \dot{\boldsymbol{\sigma}}(\hat{\boldsymbol{\theta}}_{aj}) [\dot{\boldsymbol{\sigma}}'(\hat{\boldsymbol{\theta}}_{aj}) \mathbf{W}(\hat{\boldsymbol{\theta}}_{aj}) \dot{\boldsymbol{\sigma}}(\hat{\boldsymbol{\theta}}_{aj})]^{-1}$. Then Appendix B shows that $\sqrt{n}[(\hat{\omega}_2 - \hat{\omega}_1) - (\omega_2 - \omega_1)]$ is asymptotically normally distributed with mean zero and variance can be consistently estimated by

$$\hat{\tau}_\omega^2 = \dot{h}'(\hat{\boldsymbol{\theta}}_1) \hat{\boldsymbol{\Sigma}}_{11} \dot{h}(\hat{\boldsymbol{\theta}}_1) + \dot{h}'(\hat{\boldsymbol{\theta}}_2) \hat{\boldsymbol{\Sigma}}_{22} \dot{h}(\hat{\boldsymbol{\theta}}_2) - 2 \dot{h}'(\hat{\boldsymbol{\theta}}_1) \hat{\boldsymbol{\Sigma}}_{12} \dot{h}(\hat{\boldsymbol{\theta}}_2), \quad (9)$$

with $\hat{\boldsymbol{\Sigma}}_{jk} = \hat{\mathbf{C}}'_{aj} \hat{\Gamma}_{jk} \hat{\mathbf{C}}_{ak}$, $j, k = 1, 2$. The result in (9) allows us to evaluate the significance of $(\hat{\omega}_2 - \hat{\omega}_1)$. Alternatively, we can obtain a $(1 - 2\beta)$ -level confidence interval (CI) for $(\omega_2 - \omega_1)$ as

$$[\hat{\omega}_2 - \hat{\omega}_1 - c_\beta \hat{\tau}_\omega / \sqrt{n}, \hat{\omega}_2 - \hat{\omega}_1 + c_\beta \hat{\tau}_\omega / \sqrt{n}].$$

Before ending this section, we note that the validity of SEs for $\hat{\alpha}_2 - \hat{\alpha}_1$ and $\hat{\omega}_2 - \hat{\omega}_1$ in this subsection does not need the normality assumption.

Psychometric Analysis of Diversity Measures With Entrepreneurial Teams

The data are part of a longitudinal study examining the impact of team attributes and team process on team performance (Deng, Ye, & Xie, 2013). Participants for the study are members nested within teams, which are distributed across provinces in the well-developed Eastern part of China and include Beijing. Because diversity is known to affect team performance, 13 items measuring diversity were administered to each team member starting from the first wave. The English version of the 13 diversity items are included in Appendix C, and each participant was asked to endorse each item using a 5-point Likert scale (1 = *strongly disagree*, 2 = *somewhat disagree*, 3 = *neutral/no opinion*, 4 = *somewhat agree*, 5 = *strongly agree*). In the design, the first five items are about the information/background of team members and are used to measure *information diversity*. The last eight items are about their opinions and measure *underlying diversity*. Following the design, we separate the five information-diversity items from the eight underlying-diversity items when studying their psychometric properties of reliability and unidimensionality. It is also worthy to note that, although Items 4, 6, 7, 8, 11, and 12 are phrased in the opposite direction from the other seven items, they do not affect the values of aad_{ij} , $aadm_{ij}$, or sd_{ij} whether they are reversed or not in the analysis, since each of the diversity measures uses absolute values of centralized scores or score differences.

There are a total of four waves of data in the longitudinal study of Deng et al. (2013). However, the majority of the participants showed little change in their answers to the 13 diversity items across the waves. Consequently, our analysis uses data from only the first wave. It should also be noted that there were many teams in which team members provided identical answers when endorsing each of the 13 items (resulting in a diversity value of 0 on all items), and these data are not included in our analysis. In summary, the study in this section is based on $N_t = 177$ individual participants from $N = 52$ teams, with the number of participants in a team ranging from 2 to 5.

As described in the second section, diversity measures aad_{ij} , $aadm_{ij}$, and sd_{ij} are obtained based on the Likert data and NC-transformed data for each of the 13 diversity items, respectively. These are also referred to as samples in the following discussion.

Distribution Properties

Before evaluating the reliability and unidimensionality of these measures, it is informative to check their distribution characteristics. In particular, we want to know whether the NC transformation has any effect on the distribution of the diversity

Table 1. Sample Skewness and Excess Kurtosis of the Three Measures of Diversities: Average of Absolute Distance (aad) Among Team Members, Average of Absolute Deviation From the Mean (aadm) of Team Members, and Standard Deviation (sd) Among Team Members.

Item	Sample skewness					
	Likert data			Transformed data		
	aad	aadm	sd	aad	aadm	sd
1	1.640	1.725	1.294	.832	.859	.363
2	.827	1.024	.655	.618	.780	.385
3	1.028	1.225	.694	.681	.814	.375
4	.425	.778	.122	.505	.859	.202
5	1.057	1.418	.616	1.189	1.493	.707
6	.724	.812	.534	.705	.826	.485
7	.547	.646	.403	.588	.758	.317
8	.951	1.266	.613	.834	1.122	.492
9	.614	.874	.329	.349	.535	.106
10	.603	.818	.269	.543	.773	.219
11	.262	.386	-.034	.302	.434	-.011
12	.863	1.158	.587	.897	1.192	.583
13	1.693	1.903	1.357	1.666	1.880	1.297
aave	.864	1.079	.577	.747	.948	.426

Item	Sample excess kurtosis					
	Likert data			Transformed data		
	aad	aadm	sd	aad	aadm	sd
1	2.187	2.588	1.074	-.071	.039	-.984
2	.175	1.133	-.515	-.654	-.162	-1.032
3	.588	1.246	-.339	-.356	-.021	-.803
4	.278	.913	-.080	.394	1.092	-.015
5	2.272	3.279	.966	2.962	3.751	1.689
6	-.669	-.411	-.844	-.672	-.345	-.936
7	-.242	.032	-.272	-.065	.371	-.412
8	1.339	2.851	.241	.954	2.044	.110
9	-.300	.456	-.874	-.972	-.522	-1.313
10	-.252	.244	-.625	-.339	.220	-.719
11	-.047	.304	-.484	-.036	.310	-.470
12	.839	2.501	-.122	.810	2.172	-.144
13	3.240	3.907	2.257	3.223	3.917	2.145
aave	.956	1.528	.669	.885	1.151	.829

measures. Table 1 contains the sample marginal skewness and excess kurtosis of each diversity measure for each of the 13 items. The absolute averages (aave) of the sample skewness and excess kurtosis across the 13 items are also reported at the bottom of the table. According to Table 34C of Pearson and Hartly (1954), at sample size

Table 2. Test Statistics by Fitting One-Factor Model to Each of the Three Measures of Information and Underlying Diversities: Average of Absolute Distance (aad) Among Team Members, Average of Absolute Deviation From the Mean (aadm) of Team Members, and Standard Deviation (sd) Among Team Members.

(a) Information diversity (5 items, $df = 5$)

		Likert data		Transformed data	
		Statistic	p Value	Statistic	p Value
aad	T_{ML}	8.450	.133	5.299	.380
	T_{RML}	8.872	.114	5.578	.349
	T_{AML}	5.160	.151	3.587	.343
	df_{AML}	2.908		3.216	
aadm	T_{ML}	8.663	.123	5.394	.370
	T_{RML}	8.240	.143	5.388	.370
	T_{AML}	4.606	.179	3.548	.360
	df_{AML}	2.794		3.293	
sd	T_{ML}	6.039	.302	3.770	.583
	T_{RML}	9.397	.094	5.711	.335
	T_{AML}	5.612	.131	3.580	.331
	df_{AML}	2.986		3.134	

(b) Underlying diversity (8 items, $df = 20$)

		Likert data		Transformed data	
		Statistic	p Value	Statistic	p Value
aad	T_{ML}	24.335	.228	28.521	.098
	T_{RML}	31.084	.054	28.682	.094
	T_{AML}	6.893	.177	7.546	.205
	df_{AML}	4.435		5.262	
aadm	T_{ML}	24.193	.234	28.972	.088
	T_{RML}	35.471	.018	29.692	.075
	T_{AML}	9.685	.108	8.822	.180
	df_{AML}	5.461		5.943	
sd	T_{ML}	13.196	.869	17.696	.607
	T_{RML}	30.545	.061	28.402	.100
	T_{AML}	6.958	.183	7.806	.208
	df_{AML}	4.556		5.497	

50, sample skewness is statistically significant at 2% or 10% level if its absolute value is greater than .787 or .533, respectively. At $N=52$, these critical values are slightly smaller. It is clear that multiple entries of sample skewness in the top panel of Table 1 are greater than .787. This is because each of the diversity measures is obtained using absolute values or square root of a summation of squared deviations, and such kinds of measures tend to have longer right tail. The values of sample skewness in Table 1 suggest that NC transformation does make the resulting diversity measures less skewed on average, although not all the values following the

transformation become smaller. Comparing among the three diversity measures, the values of sample skewness corresponding to sd_{ij} are uniformly the smallest while those corresponding to $aadm_{ij}$ are uniformly the largest, suggesting that different diversity measures have different distributional shapes.

The lower panel of Table 1 contains the marginal sample excess kurtosis of each of the diversity measures. According to Pearson and Hartly (1954, Table 34C), at $N=50$, sample kurtosis is significantly different from that of a normal distribution (whose excess kurtosis equals 0) at 2% or 10% level if its value is outside the interval $[-1.05, 1.88]$ or $[-.85, .99]$. At $N=52$, the end values of these intervals slightly move to the center. Similar to skewness, multiple entries of excess sample kurtosis are outside the two intervals. While the kurtosis values of aad_{ij} and $aadm_{ij}$ following the NC transformation become smaller on average, the average kurtosis of sd_{ij} following the NC transformation becomes greater. None of the diversity measures enjoys uniformly smallest excess kurtosis although the absolute average for sd_{ij} with the Likert data is the smallest.

The results in Table 1 suggest that, on average, sd has the smallest skewness and kurtosis with either the Likert data or the NC-transformed data. We note that the sample skewness and excess kurtosis for the 13th item are still significant at level .02. Thus, NML-based SEs (see, e.g., Van Zyl, Neudecker, & Nel, 2000) for reliability estimates are not valid even when sample size is large, and SEs based on the sandwich-type covariance matrix are needed. Similarly, we have to rely on the rescaled or adjusted statistics when evaluating the unidimensionality of the diversity measures using factor analysis.

Unidimensionality

Because the first five items are designed to measure information diversity and the last eight items are designed to measure underlying diversity, we would like to see whether some or all of the diversity measures on the first five or last eight items follow a one-factor model. If some or all of them follow a one-factor model, then we may choose a measure that is most reliable in future applications. If none of them follow a one-factor model, then we may need to further study the dimensionality of these diversity measures to better understand their factor structure as well as their relationship with the content of the items that these diversity measures are derived. Only after the factor structures of the aad_{ij} , $aadm_{ij}$, or sd_{ij} are well understood can we make better use of these diversity measures.

Since $N=52$ plays the role of sample size, which may not be sufficiently large for factor analysis, we use ridge ML for more reliable parameter estimates and overall model evaluation. Following the recommendation of Yuan and Chan (2008), the ridge parameter is chosen as $a=p/N=5/52$ when studying the five items of information diversity, and $a=8/52$ when studying the eight items of underlying diversity. Fitting the one-factor model to the original Likert data as well as the NC-transformed data with five and eight items by ridge ML, respectively, the rescaled and adjusted

Table 3. Ridge ML Estimates of Factor Loadings and Error Variances by Fitting One-Factor Model to Each of the Three Measures of Information (Items 1-5) and Underlying (Items 6-13) Diversities With Likert Data: Average of Absolute Distance (aad) Among Team Members, Average of Absolute Deviation From the Mean (aadm) of Team Members, and Standard Deviation (sd) Among Team Members.

θ	aad			aadm			sd		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
λ_1	.294	.175	1.676	.307	.185	1.657	.218	.108	2.013
λ_2	.408	.160	2.547	.375	.164	2.284	.353	.109	3.253
λ_3	.799	.184	4.348	.760	.193	3.946	.609	.109	5.593
λ_4	.148	.100	1.482	.154	.101	1.525	.121	.072	1.678
λ_5	.409	.126	3.249	.391	.134	2.926	.314	.081	3.858
ψ_{11}	.565	.148	4.484	.537	.147	4.296	.303	.072	5.579
ψ_{22}	.371	.107	4.383	.323	.097	4.335	.210	.063	4.832
ψ_{33}	.049	.218	.668	.065	.213	.756	.032	.098	1.316
ψ_{44}	.424	.102	5.089	.399	.110	4.491	.218	.049	6.379
ψ_{55}	.500	.160	3.737	.487	.165	3.542	.271	.077	4.765

θ	aad			aadm			sd		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
λ_6	.425	.105	4.059	.459	.108	4.237	.317	.074	4.293
λ_7	.400	.085	4.679	.251	.114	2.207	.315	.061	5.199
λ_8	.400	.139	2.871	.203	.156	1.300	.340	.081	4.208
λ_9	.241	.063	3.816	.184	.084	2.180	.202	.050	4.030
λ_{10}	.265	.077	3.452	.392	.100	3.931	.208	.055	3.820
λ_{11}	.247	.060	4.130	.177	.091	1.934	.224	.046	4.918
λ_{12}	.315	.131	2.411	.215	.145	1.480	.271	.081	3.358
λ_{13}	.358	.084	4.271	.542	.126	4.287	.256	.055	4.661
ψ_{66}	.326	.087	5.507	.240	.091	4.345	.185	.044	7.732
ψ_{77}	.186	.045	7.518	.244	.053	7.449	.102	.025	10.363
ψ_{88}	.268	.082	5.162	.331	.084	5.751	.133	.041	6.995
ψ_{99}	.285	.058	7.585	.270	.066	6.370	.160	.030	10.471
$\psi_{10, 10}$	.286	.057	7.708	.180	.058	5.756	.148	.028	10.634
$\psi_{11, 11}$	.160	.040	7.740	.170	.036	9.007	.085	.020	11.729
$\psi_{12, 12}$	.353	.064	7.960	.331	.076	6.401	.193	.035	9.905
$\psi_{13, 13}$	.621	.231	3.348	.441	.157	3.791	.322	.117	4.074

test statistics, T_{RML} and T_{AML} , together with their associated p values, are obtained and reported in Table 2. The degrees of freedom for the adjusted statistic, df_{AML} , are also included to better understand the value of T_{AML} . The statistic T_{ML} is reported as well for comparison purpose.

The statistics T_{RML} and T_{AML} for information diversity in the top panel of Table 2 suggest that, except the fit for the measure sd with the Likert data being marginal, other samples are well fitted by the one-factor model. The results also suggest that

Table 4. Ridge ML Estimates of Factor Loadings and Error Variances by Fitting One-Factor Model to Each of the Three Measures of Information (Items 1-5) and Underlying (Items 6-13) Diversities With NC-Transformed Data: Average of Absolute Distance (aad) Among Team Members, Average of Absolute Deviation From the Mean (aadm) of Team Members, and Standard Deviation (sd) Among Team Members.

θ	aad			aadm			sd		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
λ_1	.192	.141	1.362	.178	.144	1.236	.187	.095	1.978
λ_2	.300	.132	2.279	.266	.127	2.087	.269	.091	2.968
λ_3	.675	.206	3.283	.682	.228	2.991	.484	.107	4.539
λ_4	.192	.129	1.485	.196	.126	1.557	.154	.096	1.598
λ_5	.317	.122	2.604	.306	.126	2.438	.244	.080	3.065
ψ_{11}	.583	.113	6.022	.577	.111	6.046	.323	.055	7.663
ψ_{22}	.311	.078	5.206	.288	.070	5.519	.168	.046	5.719
ψ_{33}	.052	.236	.625	.013	.269	.407	.054	.085	1.765
ψ_{44}	.520	.138	4.472	.489	.147	3.994	.267	.068	5.350
ψ_{55}	.439	.139	3.856	.434	.143	3.696	.230	.066	4.968

θ	aad			aadm			sd		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
λ_6	.473	.116	4.094	.510	.120	4.255	.346	.080	4.356
λ_7	.490	.118	4.132	.379	.146	2.590	.367	.078	4.723
λ_8	.416	.166	2.505	.261	.184	1.417	.344	.096	3.562
λ_9	.305	.100	3.048	.257	.107	2.409	.267	.076	3.512
λ_{10}	.348	.117	2.984	.438	.133	3.292	.274	.080	3.420
λ_{11}	.333	.086	3.865	.266	.115	2.323	.304	.065	4.688
λ_{12}	.356	.155	2.291	.281	.170	1.646	.310	.094	3.288
λ_{13}	.386	.089	4.345	.508	.123	4.124	.279	.056	4.978
ψ_{66}	.402	.114	4.898	.306	.114	4.027	.227	.056	6.778
ψ_{77}	.310	.083	5.575	.368	.095	5.500	.169	.043	7.566
ψ_{88}	.409	.105	5.336	.458	.105	5.807	.205	.052	6.926
ψ_{99}	.606	.104	7.315	.553	.116	6.115	.341	.054	9.227
$\psi_{10, 10}$	.539	.101	6.843	.419	.107	5.350	.282	.052	8.443
$\psi_{11, 11}$	.332	.084	5.791	.337	.077	6.396	.175	.042	7.797
$\psi_{12, 12}$	.527	.098	6.964	.488	.107	6.003	.282	.051	8.555
$\psi_{13, 13}$	.724	.262	3.354	.595	.208	3.609	.376	.131	4.031

the fit to each of the three diversity measures with the NC-transformed data is a lot better than its counterpart obtained with the Likert data.

The statistic T_{AML} in the lower panel of Table 2 suggests that the fit of the one-factor model to each of the samples with underlying diversity is reasonable. However, the statistic T_{RML} suggests that the fit is marginal, although the p values corresponding to T_{RML} under the transformed data are uniformly larger. Similar to

the results displayed in the top panel of Table 2, all the samples under NC transformation are fitted by the one-factor model uniformly better according to T_{AML} . However, unlike in the top panel where sd_{ij} is fitted by the one-factor model least well, $aadm_{ij}$ in the lower panel is fitted by the one-factor model least well.

It is interesting to note that some of the statistics T_{RML} in Table 2 are multiple times of the corresponding T_{AML} and so are their corresponding degrees of freedom. This is because the measures of diversity are not normally distributed. In particular, when data are far from symmetrically distributed, T_{AML} may differ substantially from T_{RML} , and df_{AML} automatically accounts for the value of T_{AML} due to certain distributional characteristics of the sample. The statistic T_{ML} is very close to T_{RML} for some samples and is quite different from T_{RML} for other samples. This is expected because their difference also depends on the distribution of the sample.

In summary, there exist differences among the test statistics regarding the unidimensionality of the three diversity measures. But the differences are not substantial. The results in Table 2 suggest that the fit with NC-transformed data is substantially better than with Likert data. The difference in p values between T_{RML} and T_{AML} on each sample is consistent with the literature when they are applied to NML (Satorra & Bentler, 1994). In particular, Bentler and Yuan (1999), Fouladi (2000), Nevitt and Hancock (2004), and Savalei (2010) found that T_{RML} tends to reject correct models too often at smaller sample sizes; and results in Nevitt and Hancock (2004) and Savalei (2010) indicate that Type I errors of T_{AML} tend to be lower than the nominal level.

Table 3 contains the ridge ML estimates of factor loadings and error variances for the three diversity measures with Likert data. Like the test statistics in Table 2, there exist noticeable differences among parameter estimates. For example, parameter estimates for λ_8 , λ_{11} , and λ_{12} with $aadm$ in the lower panel of Table 3 are not statistically significant at the .05 level, whereas they are significant with aad and sd . Other noticeable patterns include (a) estimates of factor loadings and error variances with the measure sd for information diversity are uniformly the smallest and (b) estimates for error variances with sd for underlying diversity are uniformly the smallest. In particular, for all the 13 items, the z -statistics for sd in Table 3 are uniformly the largest, implying that parameter estimates under sd tend to be more efficient.

Estimates of factor loadings and error variances with NC-transformed data are reported in Table 4, where again there exist noticeable differences among parameter estimates across the three samples. Error variance estimates for the measure sd with underlying diversity are still uniformly the smallest, but the pattern with information diversity is not so clear. Again, the z -statistics with the measure sd are uniformly the largest across the 13 items. Comparing Tables 3 and 4, except for λ_{11} , whose estimate with $aadm$ in Table 3 is nonsignificant but significant in Table 4, the transformation does not change the significance status of other parameter estimates across the two tables. However, fluctuations exist in parameter estimates due to the transformation, some estimates in Table 3 are slightly greater while others in Table 4 are slightly greater.

Table 5. Estimates of Reliabilities Alpha and Omega for Three Measures of Information and Underlying Diversities: Average of Absolute Distance (aad) Among Team Members, Average of Absolute Deviation From the Mean (aadm) of Team Members, and Standard Deviation (sd) Among Team Members.

(a) Reliability α applied to sample covariance matrix

Information diversity (5 items)	Likert data			Transformed data		
	$\hat{\alpha}$	SE	z	$\hat{\alpha}$	SE	z
aad	.641	.092	6.988	.556	.101	5.500
aadm	.644	.097	6.615	.554	.101	5.488
sd	.669	.078	8.610	.603	.095	6.333
aadm – aad	.003	.014	.232	–.002	.014	–.166
sd – aad	.028	.028	.989	.047	.028	1.666
sd – aadm	.025	.041	.606	.049	.040	1.229

Underlying diversity (8 items)	Likert data			Transformed data		
	$\hat{\alpha}$	SE	z	$\hat{\alpha}$	SE	z
aad	.738	.074	9.953	.714	.080	8.969
aadm	.723	.080	9.015	.701	.084	8.303
sd	.774	.063	12.356	.752	.068	11.004
aadm – aad	–.016	.020	–.769	–.014	.020	–.693
sd – aad	.036	.021	1.686	.038	.023	1.665
sd – aadm	.051	.039	1.324	.052	.040	1.296

(b) Reliability ω following ridge ML

Information diversity (5 items)	Likert data			Transformed data		
	$\hat{\omega}$	SE	z	$\hat{\omega}$	SE	z
aad	.689	.077	8.973	.596	.080	7.483
aadm	.686	.086	7.969	.595	.083	7.159
sd	.716	.064	11.272	.632	.077	8.248
aadm – aad	–.003	.014	–.255	–.001	.016	–.036
sd – aad	.027	.027	1.018	.036	.033	1.103
sd – aadm	.031	.039	.784	.037	.047	.780

Underlying diversity (8 items)	Likert data			Transformed data		
	$\hat{\omega}$	SE	z	$\hat{\omega}$	SE	z
aad	.739	.078	9.520	.715	.081	8.845
aadm	.727	.073	9.923	.705	.080	8.851
sd	.774	.065	11.846	.751	.070	10.761
aadm – aad	–.012	.025	–.490	–.010	.021	–.488
sd – aad	.035	.022	1.582	.036	.023	1.542
sd – aadm	.047	.038	1.236	.046	.039	1.196

Reliability

Table 5 contains the estimates of alpha and omega for both information diversity and underlying diversity. Their SEs and corresponding z -statistics are also reported in the table. The differences of reliability estimates for either alpha or omega, together with their SEs and corresponding z -statistics are reported as well. The results in Table 5 suggest that both $\hat{\alpha}$ and $\hat{\omega}$ with the Likert data are uniformly greater than those with the transformed data, so are the corresponding z -statistics. Except for aad with underlying diversity, all the other $\hat{\omega}$ are greater than the corresponding $\hat{\alpha}$.

Among the three diversity measures (aad, aadm, sd), sd always corresponds to the largest $\hat{\alpha}$ and $\hat{\omega}$ with either the Likert or the transformed data, and for either information or underlying diversity. Except for the information diversity scale with Likert data in the top left portion of Table 5, aadm always corresponds to the smallest $\hat{\alpha}$ and $\hat{\omega}$. However, the largest z -statistic for reliability difference is always between sd and aad (sd – aad). This is because the SEs corresponding to the difference of the estimates between sd and aad tend to be smaller than those between sd and aadm, due to different correlations between the estimates of reliability.

Although none of the differences in reliability estimates are significant at the level of .05, three differences of $\hat{\alpha}$ are at the level of .1, corresponding to those between sd and aad for information diversity with transformed data ($z = 1.666$), and for underlying diversity with both Likert and NC-transformed data ($z = 1.686, 1.665$). Two differences in $\hat{\omega}$ corresponding to those between sd and aad for underlying diversity are also marginal ($z = 1.582, 1.542$). As there are only $N = 52$ independent teams in the analysis, we would expect that the difference of reliability estimates between sd and aad to become more pronounced and significant with a larger N .

Most $\hat{\omega}$ are greater than $\hat{\alpha}$ in Table 5. There are also exceptions, for example, the measure sd with the NC-transformed sample for underlying diversity. Such observed differences are expected since the items may not be essentially tau-equivalent or literally unidimensional (see, e.g., Raykov, 1997).

Discussion and Conclusion

In this article, we described methodology for studying unidimensionality and reliability of diversity measures with Likert data. Using some real data, the analyses indicate that the reliability estimates corresponding to sd are the greatest. The z -statistics for the reliability estimates corresponding to sd are the largest, and so are the corresponding z -statistics for estimates of factor loading and error variances. The SEs for these estimates are also the smallest. These indicate that the diversity measure sd tends to yield more efficient parameter estimates than both aad and aadm. With respect to unidimensionality, there is little difference in test statistics across the three diversity measures. Thus, among the three measures of diversity, sd is the preferred measure. Comparing between the NC-transformed data and the Likert data, the transformed data are closer to the underlying values if the NC assumption holds. The transformed data are less skewed on average; diversity measures with the transformed data also

tend to be more unidimensional than those with Likert data. However, reliability estimates following the transformed data tend to be smaller than those following the Likert data. It would appear that some additional studies are needed to further examine the merit of the transformed data versus Likert data.

The implication in studying the reliability and unidimensionality of the diversity measures is that there exists a model in which each measure (aad_{ij} , $aadm_{ij}$, or sd_{ij}) is linearly related to a latent diversity trait plus an error or uniqueness term. In contrast, models in studying rater reliability assume that the observed scores are linearly related to some underlying traits. Clearly, both kinds of models are hypothetical and are motivated by the needs to study reliability and/or unidimensionality of the corresponding measurements. The obtained results in this study indicate that each of the three diversity measures is fitted by the one-factor model reasonably well, and each subscale defined by these measures also has a decent reliability. More research is clearly needed as to whether the model behind the diversity measures aad , $aadm$, or sd is more plausible or that behind the original x_{ijk} or y_{ijk} is more plausible. This issue might be best addressed through an analysis in which many different sets of real data are examined rather than through analytical or simulation studies.

Appendix A

This appendix gives the formula for calculating consistent estimates $\hat{\Gamma}_{jk}$ of

$$\Gamma_{jk} = \text{Cov}(\sqrt{ns_j}, \sqrt{ns_k}), \quad j, k = 1, 2.$$

Let the two different measures of diversity be denoted by $\mathbf{d}_{i1}$ and $\mathbf{d}_{i2}$, each is a $p \times 1$ vector, $i = 1, 2, \dots, N = n + 1$. Let $\bar{\mathbf{d}}_j$ be the sample mean of the j th diversity measure, and $\mathbf{t}_{ij} = \text{vech}[(\mathbf{d}_{ij} - \bar{\mathbf{d}}_j)(\mathbf{d}_{ij} - \bar{\mathbf{d}}_j)']$. Then a consistent estimate of Γ_{jk} is given by

$$\hat{\Gamma}_{jk} = \frac{1}{N} \sum_{i=1}^N (\mathbf{t}_{ij} - \bar{\mathbf{t}}_j)(\mathbf{t}_{ik} - \bar{\mathbf{t}}_k)',$$

where $\bar{\mathbf{t}}_j$ and $\bar{\mathbf{t}}_k$ are the vectors of sample means of $\mathbf{t}_{ij}$ and $\mathbf{t}_{ik}$, respectively.

Appendix B

Asymptotic Distributions of $\hat{\alpha}_2 - \hat{\alpha}_1$ and $\hat{\omega}_2 - \hat{\omega}_1$

This appendix shows that both $\hat{\alpha}_2 - \hat{\alpha}_1$ and $\hat{\omega}_2 - \hat{\omega}_1$ are asymptotically normally distributed and gives the formulas for calculating consistent estimates of their variances.

In the Methodology section we have introduced the notation $\hat{\alpha}_1 = g(\mathbf{s}_1)$ and $\hat{\alpha}_2 = g(\mathbf{s}_2)$. Their population counterparts are given by $\alpha_1 = g(\boldsymbol{\sigma}_1)$ and $\alpha_2 = g(\boldsymbol{\sigma}_2)$. It follows from standard asymptotics that

$$\begin{aligned} \sqrt{n}[(\hat{\alpha}_2 - \alpha_2) - (\hat{\alpha}_1 - \alpha_1)] &= \sqrt{n}[g(\mathbf{s}_2) - g(\boldsymbol{\sigma}_2)] - \sqrt{n}[g(\mathbf{s}_1) - g(\boldsymbol{\sigma}_1)] \\ &= \dot{g}'(\boldsymbol{\sigma}_2)\sqrt{n}(\mathbf{s}_2 - \boldsymbol{\sigma}_2) - \dot{g}'(\boldsymbol{\sigma}_1)\sqrt{n}(\mathbf{s}_1 - \boldsymbol{\sigma}_1) + o_p(1), \end{aligned} \quad (\text{B1})$$

where $o_p(1)$ denotes a term that converges to 0 in probability when $n \rightarrow \infty$. According to the central limit theorem, $\sqrt{n}(\mathbf{s}_1 - \boldsymbol{\sigma}_1)$ and $\sqrt{n}(\mathbf{s}_2 - \boldsymbol{\sigma}_2)$ are jointly asymptotically normally distributed with asymptotic variance-covariance matrices $\boldsymbol{\Gamma}_{11}$, $\boldsymbol{\Gamma}_{22}$, and $\boldsymbol{\Gamma}_{12}$. It follows from (B1) that

$$\sqrt{n}[(\hat{\alpha}_2 - \alpha_2) - (\hat{\alpha}_1 - \alpha_1)] \xrightarrow{\mathcal{L}} N(0, \tau_\alpha^2), \quad (\text{B2})$$

where

$$\tau_\alpha^2 = \dot{g}'(\boldsymbol{\sigma}_1)\boldsymbol{\Gamma}_{11}\dot{g}(\boldsymbol{\sigma}_1) + \dot{g}'(\boldsymbol{\sigma}_2)\boldsymbol{\Gamma}_{22}\dot{g}(\boldsymbol{\sigma}_2) - 2\dot{g}'(\boldsymbol{\sigma}_1)\boldsymbol{\Gamma}_{12}\dot{g}(\boldsymbol{\sigma}_2).$$

For two estimates $\hat{\omega}_1 = h(\hat{\boldsymbol{\theta}}_1)$ and $\hat{\omega}_2 = h(\hat{\boldsymbol{\theta}}_2)$ of omega, there exists

$$\begin{aligned} \sqrt{n}[(\hat{\omega}_2 - \omega_2) - (\hat{\omega}_1 - \omega_1)] &= \sqrt{n}[h(\hat{\boldsymbol{\theta}}_2) - h(\boldsymbol{\theta}_2)] - \sqrt{n}[h(\hat{\boldsymbol{\theta}}_1) - h(\boldsymbol{\theta}_1)] \\ &= h(\boldsymbol{\theta}_2)\sqrt{n}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) - h(\boldsymbol{\theta}_1)\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) + o_p(1). \end{aligned} \quad (\text{B3})$$

The development in Yuan and Chan (2008) implies that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) = \sqrt{n}(\hat{\boldsymbol{\theta}}_{aj} - \boldsymbol{\theta}_{aj}) = \mathbf{C}'_{aj}\sqrt{n}(\mathbf{s}_j - \boldsymbol{\sigma}_j) + o_p(1), \quad j = 1, 2, \quad (\text{B4})$$

where

$$\mathbf{C}_{aj} = \mathbf{W}(\boldsymbol{\theta}_{aj})\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}_{aj})[\dot{\boldsymbol{\sigma}}'(\boldsymbol{\theta}_{aj})\mathbf{W}(\boldsymbol{\theta}_{aj})\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}_{aj})]^{-1}.$$

Combining (B3) and (B4) yields

$$\sqrt{n}[(\hat{\omega}_2 - \omega_2) - (\hat{\omega}_1 - \omega_1)] \xrightarrow{\mathcal{L}} N(0, \tau_\omega^2),$$

where

$$\tau_\omega^2 = \dot{h}'(\boldsymbol{\theta}_1)\boldsymbol{\Omega}_{11}\dot{h}(\boldsymbol{\theta}_1) + \dot{h}'(\boldsymbol{\theta}_2)\boldsymbol{\Omega}_{22}\dot{h}(\boldsymbol{\theta}_2) - 2\dot{h}'(\boldsymbol{\theta}_1)\boldsymbol{\Omega}_{12}\dot{h}(\boldsymbol{\theta}_2)$$

with $\boldsymbol{\Omega}_{jk} = \mathbf{C}'_{aj}\boldsymbol{\Gamma}_{jk}\mathbf{C}_{ak}$.

Appendix C

Thirteen Items for Measuring Team Diversity

The first five items are for measuring information diversity, and the last eight items are for measuring underlying diversity. Participants were asked to endorse each of the items using a 5-point Likert-type scale.

1. Overall, the ages of team members are widely distributed.
2. Overall, team members have diverse background and training.
3. Overall, knowledge and specialty of team members are complementary.

4. Overall, team members have similar social experience (reversed).
 5. Overall, team members have different expertise.
 6. Overall, team members have the same value regarding entrepreneurial development (reversed).
 7. If starting a new business, team members will aim to achieve the same goal (reversed).
 8. If starting a new business, team members will have the same ambition (reversed).
 9. Overall, team members have different personality.
 10. Overall, team members have different working styles.
 11. Overall, team members are on the same page regarding the goal of the team (reversed).
 12. Overall, team members have the same consensus regarding the focus of the team (reversed).
 13. Team members have different entrepreneurial philosophy.
-

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported in part by a grant from the National Natural Science Foundation of China (71002023) and a grant from China Scholarship Council.

Note

1. We omit the subscripts i , j , and k from q and h to simplify the notation.

References

- Algina, J. (1978). Comment on Bartko's "On various intraclass correlation reliability coefficients." *Psychological Bulletin*, 85, 135-138.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.
- Anderson, J. C., & Gerbing, D. W. (1982). Some methods for respecifying measurement models to obtain unidimensional construct measurement. *Journal of Marketing Research*, 19, 453-460.
- Babakus, E., Ferguson, J. C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, 24, 222-228.
- Bagozzi, R. P. (1980). *Causal models in marketing*. New York: John Wiley.

- Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, 34, 181-197.
- Biemann, T., & Kearney, E. (2010). Size does matter: How varying group sizes in a sample affect the most common measures of group diversity. *Organizational Research Methods*, 13, 582-599.
- Blau, P. M. (1977). *Inequality and heterogeneity*. New York, NY: Free Press.
- Burt, R. S. (1973). Confirmatory factor-analysis structures and the theory construction process. *Sociological Methods & Research*, 2, 131-187.
- Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research*, 5, 3-52.
- Deng, L., Ye, S., & Xie, L. (2013). A longitudinal study of team trait combinations, team process and performance. *Journal of Management Science* (manuscript under review).
- Feller, W. (1945). On the normal approximation to the binomial distribution. *Annals of Mathematical Statistics*, 16, 319-329.
- Fouladi, R. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling*, 7, 356-410.
- Guzzo, R. A., & Shea, G.P. (1992). Group performance and intergroup relations in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 3, 2nd ed., pp. 269-313). Palo Alto, CA: Consulting Psychologists Press.
- Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*, 32, 1199-1228.
- Hays, W. L. (1981). *Statistics* (3rd ed.). New York, NY: Holt, Rinehart & Winston.
- Jackson, S. E., Joshi, A., & Erhardt, N. L. (2003). Recent research on team and organizational diversity: SWOT analysis and implications. *Journal of Management*, 29, 801-830.
- Lee, S.-Y., & Shi, J.-Q. (1998). Analysis of covariance structures with independent and non-identically distributed observations. *Statistica Sinica*, 8, 543-557.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Nevitt, J., & Hancock, G. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research*, 39, 439-478.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44, 443-460.
- Pearson, E. S., & Hartly, H. O. (1954). *Biometrika tables for statisticians* (Vol. 1). London, England: Biometrika Trust.
- Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, 32, 329-353.
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Routledge.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Savalei, V. (2010). Small sample statistics for incomplete nonnormal data: Extensions of complete data formulae and a Monte Carlo comparison. *Structural Equation Modeling*, 17, 241-264.

- Schott, J. (2005). *Matrix analysis for statistics* (2nd ed.). New York, NY: John Wiley.
- Schuster, C., & Smith, D. A. (2002). Indexing systematic rater agreement with a latent-class model. *Psychological Methods*, 3, 384-395.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Stewart, G. L. (2006). A meta-analytic review of relationships between team design features and team performance. *Journal of Management*, 32, 29-54.
- Teachman, J. D. (1980). Analysis of population diversity. *Sociological Methods & Research*, 8, 341-362.
- Van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, 89, 1008-1022.
- Van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, 65, 271-280.
- Webber, S. S., & Donahue, L. M. (2001). Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis. *Journal of Management*, 27, 141-162.
- Yuan, K.-H., & Bentler, P. M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika*, 67, 251-259.
- Yuan, K.-H., & Chan, W. (2008). Structural equation modeling with near singular covariance matrices. *Computational Statistics & Data Analysis*, 52, 4842-4858.