Gaining from discretization of continuous data: The correspondence analysis biplot approach

Se-Kang Kim¹ · Craig L. Frisby²

Published online: 7 November 2018
© Psychonomic Society, Inc. 2018

Abstract

According to Stevens's classification of measurement, continuous data can be either ratio or interval scale data. The relationship between two continuous variables is assumed to be linear and is estimated with the Pearson correlation coefficient, which assumes normality between the variables. If researchers use conventional statistics (*t* test or analysis of variance) or factor analysis of correlation matrices to study gender or race differences, the data are assumed to be continuous and normally distributed. If continuous data are discretized, they become ordinal; thus, discretization is widely considered to be a downgrading of measurement. However, discretization is advantageous for data analysis, because it provides interactive relationships between the discretized variables and naturally measured categorical variables such as gender and race. Such interactive relationship information between categories is not available with the ratio or interval scale of measurement, but it is useful to researchers in some applications. In the present study, Wechsler intelligence and memory scores were discretized, and the interactive relationships were examined among the discretized Wechsler scores (by gender and race). Unlike in previous studies, we estimated category associations and used correlations to enhance their interpretation, and our results showed distinct gender and racial/ethnic group differences in the correlational patterns.

Keywords Discretization of continuous data · Correspondence analysis · Biplot · Wechsler adult intelligence and memory scales

For traditional multivariate analytic methods (e.g., multivariate analysis of variance [MANOVA] or maximum likelihood estimation of factor analysis models), the data are typically regarded as random samples from a population that is normally distributed. If the population is assumed to be normally distributed, the level of measurement is at least an interval scale, and data are assumed to be continuous. According to Pearson (1904), when correlations are estimated between continuous variables, any pair of variables to be correlated are assumed to be (bivariate) normal, and he called this correlation “normal correlation” (Nishisato, 2007, p. 19). Following this principle, any factor-analytic approaches are based on “normal” correlation matrices, and if the multivariate normality assumption is violated among the variables, the validity of the factor analysis results would be questionable.

In terms of Stevens's classification of measurement (Stevens, 1951), continuous data are at least interval-scale data. If continuous data are discretized into several ordered categories, the discretized data have an ordinal scale of measurement. Thus, discretization is considered to be a downgrading of measurement, because it transforms ratio or interval scale data into ordinal scale data, which is against the object of scaling (Nishisato, 2007). This is because ratio or interval scale data include more numeric information than do ordinal scale data (Stevens, 1951). However, Nishisato (1980, 1994, 1999, 2007) argues that discretization, despite its degrading of measurement, is advantageous in terms of data analysis. That is, discretized continuous data are amenable to many types of correspondence analysis (e.g., simple, multiple, or canonical correspondence analysis; Greenacre, 2016), and they allow estimation of interactive relationships between categories, a technique that is not available in the analysis of continuous data but is important to those who want to study interactive relationships. For example, using correspondence analysis (CA), Greenacre (2016) discretized ages (16–75 or more) into seven age groups (16–24, 25–34, . . . , 65–74, 75+) and examined the interactive relationships between age groups, gender categories, and self-assessed health categories

✉ Se-Kang Kim
sekim@fordham.edu

¹ Fordham University, Bronx, NY, USA

² University of Missouri, Columbia, MO, USA

(see Greenacre, 2016, pp. 122–123). Thus, the sacrifice of fine information (by age discretization) provides substantial gains for studying interactive relationships between Gender $\times$ Age Groups $\times$ Self-Perceived Health conditions.

Numerous studies have discretized continuous data in order to create categories (e.g., Jacoby & Matell, 1971; Johari & Sclove, 1976; Shaw, Huffman, & Haviland, 1987; Stefansky & Kaiser, 1973). Cox (1957) examined information loss due to discretization (of continuous data) and studied the effects for two through six intervals (e.g., ordered categories) in discretized continuous variables. In a recovery study, Green and Rao (1970) recommended generating at least six, but preferably eight, ordered categories from continuous data. Later, Lehmann and Hurlbert (1972) concluded that the use of two and three ordered categories may be appropriate in the discretization of continuous data.

Cognitive ability scores obtained from standardized tests have traditionally been measured on an interval scale. The nature and extent of racial/ethnic group differences in mean performance on individual and group-administered tests of cognitive abilities have been well researched in the psychological literature (Gottfredson, 2005; Jensen, 1998; Lynn, 2006; Mackintosh, 1998; Rushton & Jensen, 2005). One approach is simply to compare the mean scores of general or specific cognitive ability measures for different racial/ethnic groups (e.g., Black, Hispanic, and White) in order to test Spearman's hypothesis, which was designed to evaluate racial/ethnic subgroup differences in average performance on cognitive tests (e.g., Jensen, 1980, 1985, 1992; Rushton & Jensen, 2005; Spearman, 1927). Also, factor-analytic methods have been used (e.g., multigroup confirmatory factor analysis) to examine factor loadings in combination with mean score differences in cognitive abilities among racial/ethnic groups (e.g., Ashton & Lee, 2005; Colom & Lynn, 2004; Dolan, 2000; Dolan & Hamaker, 2001; Frisby & Beaujean, 2015; Harrington, 2009; Jensen, 1992).

Racial/ethnic group and gender differences in IQ scores

Differences between Black and White groups

Weiss, Saklofske, Holdnack, and Prifitera (2015) reviewed the literature documenting measurement invariance for intelligence tests such as the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III) for adult racial/ethnic groups. They reported that Black and White groups differed by approximately one standard deviation (equivalent to 15 IQ points) on IQ tests (Jensen, 1998; Weiss, 2003), and that such tests do not show evidence of statistical bias (Brown, Reynolds, & Whitaker, 1999; Jensen, 1980; Reynolds & Kaiser, 1990; Reynolds & Suzuki, 2013). However, more recent research

has also shown that the Black/White gap in full-scale IQ (FSIQ) scores generally tends to be larger for adults than for children and adolescents. Although the mean FSIQ score for Whites remains consistent across scores on the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV) and WAIS-IV (approximately 103), the mean FSIQ on the WISC-IV for African Americans is approximately 3 points higher than the WAIS-IV FSIQ mean for African Americans. Because of these slight differences, the Black/White FSIQ gap for adults (approximately 14.5 points on the WAIS-IV) is 3 points higher than the Black/White FSIQ gap for children and adolescents (approximately 11.5 points on the WISC-IV; Weiss, Chen, Harris, Holdnack, & Saklofske, 2010).

In terms of Level I (simple registering, storage, and retrieval of stimuli involving no intentional transformation of information prior to output) and Level II (complex generalization, transfer, and verbal mediation of stimuli prior to output) cognitive abilities, Vernon (1981) found that Blacks and Whites showed statistically significant differences in Level II abilities. Also, factor-analytic methods have been used to examine factor loadings combined with mean score differences in cognitive abilities (Chen, West, & Sousa, 2006; Dolan, 2000; Dolan & Hamaker, 2001; Harrington, 2009; Jensen, 1992); specifically, researchers examined the factor loadings on general and specific cognitive abilities and then correlated these loadings with the mean score differences between Blacks and Whites.

Differences between Hispanic and White groups

The mean FSIQ scores of American Hispanics are typically positioned between those of Blacks and Whites (Neisser et al., 1996). When White/Hispanic mean differences in FSIQ were measured in standard deviation units, Gottfredson (1998) estimated this difference to be approximately 0.5 standard deviations, Sackett and Wilk (1994) estimated the difference to be in the approximate range of 0.6–0.8 standard deviation units, and Herrnstein and Murray (1994) estimated the difference to be approximately 0.5–1.0 standard deviation units. A meta-analytic study estimated this difference to be about 0.7 standard deviations (favoring Whites; Roth, Bevier, Bobko, Switzer, & Tyler, 2001). As an attempt to understand factor structure differences between Hispanics and Whites on cognitive ability test data, Jensen (1973) found that White and Mexican-American pupils scored farthest apart on measures of memory. This finding was confirmed by Weiss et al. (2015) using standardization data from the WISC-V. To examine Hispanic/White differences, Hartmann, Kruuse, and Nyborg (2006) reported that differences were found not only in general ability but also in specific cognitive abilities. Although the analytic methods in Jensen's (1973) and Hartmann et al.'s (2006) studies differed, the studies resulted in similar conclusions.

Gender differences

Gender differences on various types of cognitive ability scores have been studied mainly using meta-analytic techniques. Hyde (1981) reanalyzed the well-established studies on cognitive gender differences by Maccoby and Jacklin (1974), focusing specifically on verbal ability, quantitative ability, and visual-spatial ability. Gender differences were minor in their analyses. Similarly, Hyde and Linn (1988) conducted a meta-analysis of 165 studies and found that females scored higher (with a mean effect size of + 0.11) in verbal ability performance. However, they concluded that the difference was so small that gender differences in verbal ability no longer exist. Feingold (1988) also found that gender differences are disappearing in cognitive ability. Voyer, Voyer, and Bryden (1995) also conducted a meta-analysis of 286 effect sizes from a variety of spatial ability measures. They concluded that gender differences (favoring males) were significant in spatial abilities, but that the magnitudes of gender differences were tending to decrease.

Shortcomings of previous research

From this brief literature review, three conclusions are evident: (1) The size of mean differences in the cognitive test performance of nationally representative samples of racial/ethnic groups varies as a function of mean score group comparisons in FSIQ; (2) the relative advantage that one racial/ethnic group displays over another group differs as a function of the specific type of cognitive ability; and (3) there is no strong evidence for gender differences in cognitive abilities, with the exception of spatial abilities (e.g., Hedges & Nowell, 1995).

However, racial/ethnic group and gender differences in cognitive abilities were investigated separately. As far as we know, no previous studies have examined the interaction of racial/ethnic and gender groups with specific cognitive ability levels (high, medium, low) in a single study. However, a few studies have examined either Age $\times$ Race interaction in cognitive abilities or Gender $\times$ Race interaction with body and head size in IQ (Chen, Kaufman, & Kaufman, 1994; Jensen & Johnson, 1994). Previous studies have mainly examined mean/effect size or factor-loading differences (e.g., Hedges & Nowell, 1995; Hyde & Linn, 1988). For the present study, we examined the correlational patterns between Gender $\times$ Racial/Ethnic Groups and specific achievement levels in various cognitive abilities, using the biplot procedure within the context of CA (Beh & Lombardo, 2014; Greenacre, 2016).

An example illustrating the value of CA

Sourial et al. (2010) demonstrated the value of using CA to uncover hidden relationships among categorical variables,

using a hypothetical example of questionnaire data and the primary language spoken by 1,000 residents in each of five countries (Canada, USA, England, Italy, and Switzerland). Each of the 1,000 respondents from each country reported one among five languages spoken (English, French, Spanish, German, or Italian). These data can be displayed as a contingency table of frequencies, in which the country of origin constitutes the row variable and the language spoken represents the column variable.

One method of evaluating the relationship between country of origin (the row variable) and language spoken (the column variable) would be to conduct a chi-square test, which might indeed reveal that a statistically significant relationship exists. However, such an approach would only reveal that a statistically significant relationship exists between the row and column variables, but not *which* response categories are most closely related. Furthermore, two or more categorical variables involved in a chi-square test could easily violate the independence assumption between categories, which would render the results invalid. CA can be used for the analysis of multiple categorical variables without requiring the independence assumption among categories, thus accommodating any type of categorical variable, regardless of whether it is binary, ordinal, or nominal (Sourial et al., 2010).

The purpose of CA is to graphically represent contingency-table data in terms of the distance between individual row and column categories and the distance to the average row and column categories, respectively, in a two-dimensional space. The farther away from the origin (e.g., the intersection of two dimensions) a response category is along a particular dimension, the greater is its importance for that dimension.

From a visual inspection of the categories represented in a two-dimensional space, the Sourial et al. (2010) CA analysis was able to shed light on several important findings. First, the most important difference (or largest deviation) from row/column independence on the first dimension in the sample was between Italy (row variable), the Italian language (column variable), and the other countries and languages. That is, an unusually large proportion of persons living in Italy spoke Italian, relative to other languages. In addition, Switzerland (row variable) and German (column variable) were farthest away from the centroid on Dimension 2. That is, an unusually large proportion of persons living in Switzerland spoke German, relative to other languages. Second, the degree to which categories cluster can be used as a useful guideline to visually interpreting relationships between the row and column categories. Sourial et al.'s visual plot clearly showed that the country of Italy clustered closely with the Italian language and that the country of Switzerland clustered most closely with the German language, and to a lesser extent with the French language. The country of Canada was most closely associated with the English language, and to a lesser extent with the French language.

Simple visual inspection to approximate category associations is easy to perform, but without relying on any numerical estimates, an interpretation solely based on visual inspection could be arbitrary and misleading. Therefore, in the present study we first visually inspected category associations in order to approximate a global picture of the salient relationships among categories of the row and column variables, and then estimated the visually inspected associations numerically with correlations utilizing the CA biplot paradigm.

Introducing correspondence analysis biplot (CAB)

In this study, we neither tested Spearman's hypothesis nor examined factor loadings to study group differences. Instead, we utilized CAB to study the associations between gender racial/ethnic groups and three performance levels (low, medium, and high), in terms of FSIQ and various cognitive abilities. The CAB approach introduced in the present study was combined with CA and the biplot technique: Here, biplot mapping was used to visually depict associations among the categories of the row and column variables, and then CA was used to quantify such visually inspected associations with correlations. The detailed procedures will be described in the Method section.

Biplot

The *biplot* is a statistical approach that was first introduced by Gabriel (1971) in terms of a principal component analysis (PCA) of continuous data (e.g., Gabriel & Odoroff, 1990; Gower & Hand, 1996; Legendre & Gallagher, 2001). Biplot jointly displays rows and columns in order to visually approximate their relationships in a two-dimensional plane constructed from a pair of dimensions (Gabriel, 1971; Gabriel & Odoroff, 1990; Gower, Lubbe, & le Roux, 2011; Greenacre, 2010). The prefix “bi” in the word “biplot” does not necessarily stand for two dimensions, but simply for rows and columns. However, the first two dimensions are usually used to construct a biplot map, because they account for the largest amount of variance. A biplot is analogous to a multivariate version of a scatterplot. In an ordinary scatterplot, the horizontal axis and the vertical axis represent two variables (say, X and Y), whereas in a biplot, the horizontal axis represents the first dimension and the vertical axis represents the second dimension. This biplot paradigm has also been used for analyzing categorical data and has been associated with CA (Beh & Lombardo, 2014; Blasius & Greenacre, 2014; Bradu & Gabriel, 1978; Gabriel, 2002; Greenacre, 1993, 2010, 2016; Le Roux & Rouanet, 2004; Ter Braak, 1983; Ter Braak & Verdonschot, 1995). For example, Greenacre (2016) conducted a CA on health data and biplotted age levels (categories of a

row age variable) and perceived health categories (categories of a column health variable) in a two-dimensional map constructed with the first two dimensions, and they then visually inspected the variables' relationships.

Testing the statistical significance of dimensions

To construct a two-dimensional biplot, it is necessary to test whether the dimensions included in a biplot are statistically significant. Otherwise, any statistics (e.g., category correlations) estimated from a biplot that consists of nonsignificant dimensions would be false. Therefore, we conducted a permutation test to examine the statistical significance of the principal inertias (or eigenvalues) for the dimensions; these detailed procedures will also be included in the Method section.

Correspondence analysis

To examine the relationships between gender as this variable interacts with racial/ethnic groups and categories of performance levels in cognitive abilities, cognitive ability performance needed to be discretized as high, medium, and low. CA is analogous to PCA of a two-way contingency table. The row and column categories in a two-way table work as the input variables in PCA. As PCA estimates loadings for the individual input variables on orthogonal dimensions, CA also estimates the coordinates for row and column categories on orthogonal dimensions. The category coordinates in CA are analogous to the variable loadings in PCA. However, unlike PCA, CA estimates separately the dimensions from row categories and the dimensions from column categories, while dimensionality is kept the same for the row and column categories (e.g., Greenacre, 2016; Nishisato, 1980, 1994); as a result, row category coordinates derive from the row dimensions, and column category coordinates from the column dimensions.

Row and column coordinates To estimate the row and column coordinates, *singular value decomposition* (SVD) of a residual data matrix was conducted. The residual data matrix $\mathbf{S}$ included only information about the category associations and was defined as

$$\mathbf{S} = \mathbf{D}_r^{-1/2} (\mathbf{P} - \mathbf{rc}^T) \mathbf{D}_c^{-1/2} \quad (1)$$

where $\mathbf{P} = \mathbf{N}/n_{++}$; the total count $n_{++} = \sum_i \sum_j n_{ij}$, where n_{ij} refers to the frequency count in the row i ($i = 1, 2, \dots, I$) and column j ($j = 1, 2, \dots, J$); $\mathbf{rc}^T$ = the products of the row and column masses, which are the expected relative frequencies when the null hypothesis of independence is true; the row masses are defined as $\mathbf{r} = (n_{1+}/n_{++}, \dots, n_{I+}/n_{++})^T$, and the column masses are defined as $\mathbf{c} = (n_{+1}/n_{++}, \dots, n_{+J}/n_{++})^T$, where n_{i+} and n_{+j} are the marginal sums of row i and column j , respectively; and

$\mathbf{D}_r = \text{diag}(\mathbf{r})$ and $\mathbf{D}_c = \text{diag}(\mathbf{c})$. In the residual matrix $\mathbf{S}$, $\mathbf{D}_r^{-1/2}$ and $\mathbf{D}_c^{-1/2}$ function as weights. In the equation, $(\mathbf{P} - \mathbf{rc}^T)$ is the core residual part of interest, because it carries only row–column associations and is directly related to the size of the total inertia (or variance). The SVD of $\mathbf{S}$ is $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, by constraining $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$, where $\mathbf{U}$ = a left singular (or eigen-) vector matrix; $\mathbf{V}^T$ = a transposed right singular (or eigen-) vector matrix; and $\mathbf{\Sigma}$ = the diagonal matrix of singular values in descending order: $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_K \geq 0$, and $K = \text{rank}(\mathbf{S})$ or the maximum dimensionality, which equals $\min(I - 1, J - 1)$ (I = number of rows and J = number of columns).

Standard and principal coordinates CA estimates two types of coordinates: standard and principal. The row standard coordinates are defined as $\Phi = \{\phi\} = \mathbf{D}_r^{-1/2}\mathbf{U}$, and the column standard coordinates as $\Gamma = \{\gamma\} = \mathbf{D}_c^{-1/2}\mathbf{V}$. Numerically, the principal coordinates are equal to the standard coordinates multiplied by singular values; thus, the principal coordinates $\mathbf{F}$ of rows and $\mathbf{G}$ of columns are defined as $\mathbf{F} = \Phi\mathbf{\Sigma}$ and $\mathbf{G} = \Gamma\mathbf{\Sigma}$. The standard coordinates Φ or Γ are scaled to have a weighed mean of 0 and a variance (or standard deviation) of 1 in dimension k , such as $\Gamma\mathbf{D}_c = \Phi\mathbf{D}_r = \mathbf{0}$ and $\Gamma\mathbf{D}_c\Gamma^T = \Phi\mathbf{D}_r\Phi^T = \mathbf{I}$. In our biplot, the column standard coordinates Γ are set to have mean 0 and unit length 1 in each dimension, so that reference can be made to each row profile being at the center of gravity of the associated column vectors.

Points and vectors in a biplot plane To enhance differences between the principal and standard coordinates in a biplot plane, the principal coordinates of rows $\mathbf{F}$, or $\{f_{ik}\}$ ($i = i^{\text{th}}$ row and $k = k^{\text{th}}$ dimension), are usually depicted as points, and the standard coordinates of columns Γ , or $\{\gamma_{jk}\}$ ($j = j^{\text{th}}$ column and $k = k^{\text{th}}$ dimension), as vectors. In a biplot, the row points are viewed as projections of the row profiles onto the best-fitting plane, whereas the columns are viewed as projections onto the same plane of the unit profiles $[1, 0, 0, \dots, 0]$, $[0, 1, 0, \dots, 0]$, $\dots$, and so on. For example, a row profile $[p_1, p_2, \dots, p_{30}]$, with 30 elements of proportions (p_j) adding up to 1, can be expressed as $p_1[1, 0, \dots, 0] + p_2[0, 1, 0, \dots, 0] + \dots + p_{30}[0, 0, \dots, 1]$. Note that we analyze a 6×30 table. It follows that the row profiles are weighted averages of the column points, the weights being the profile elements. The weighted average property makes a biplot so useful because the row profile points and column vector points lie in the same space, with the column vector points defining the most extreme profile points possible.

Visual inspection of category relationships in a biplot One can visually inspect the relationships between row and column categories with ease. If imaginary lines from the origin (0, 0) are drawn to two categories, one could easily approximate the strength and magnitude of their association by the angle between

the categories. According to trigonometry, if the angle between f_{ik} and ϕ_{jk} is close to zero, the correlation would be close to +1; if the angle is close to 90° , the correlation would be close to 0; if the angle is close to 180° , the correlation would be close to -1; if the angle is between 0° and 90° , the correlation would be between 0 and +1; and if the angle is between 90° and 180° , the correlation would be between 0 and -1. The appealing feature in a biplot is the ability to visually approximate the category associations without consulting any numerical statistics. However, to quantify the visually approximate category associations, we will estimate the category associations using numerical correlations, to enhance their interpretation.

Gaining from CA of discretized continuous data

For cognitive ability measurements, racial/ethnic group or gender differences in continuous measures such as the mean scores or factor loadings on cognitive ability measures are based on the ratio or interval scale of measurement, and they cannot assess associations between naturally measured gender or racial/ethnic categories and discretized levels (high, medium, or low) of cognitive ability measures. However, examining the relationships with discretized (specific) ability levels helps us understand certain attributes of the cognitive abilities maintained in different gender or racial/ethnic groups, which cannot be pursued in an analysis of continuous data. The discretization of cognitive ability could be arbitrary, but a researcher should prudently discretize continuous data using standard deviation units. For example, one can assign those who score one standard deviation above the mean as “high”; those within one standard deviation above or below the mean as “medium”; and those one standard deviation below the mean as “low.” As Nishisato argued (1999, 2007), richer and more meaningful information could then be obtained, such as an understanding of relationships with gender as this variable interacts with different racial/ethnic groups and specific levels of cognitive ability. Such three-way interactive relationship information (Gender $\times$ Racial/Ethnic Groups $\times$ Levels of Cognitive Ability) is not available without discretization of continuous data.

Summary of the present study

CA is a multivariate graphical technique designed to explore the relationships among categorical variables (e.g., Beh & Lombardo, 2014; Greenacre, 2016; Kim et al., 2016; Sourial et al., 2010). The outcome from CA is a graphical display of the rows and columns of a contingency table that is designed to permit visualization of the salient relationships among the variable responses in a low-dimensional space. However, CA does not estimate the visualized relationships with correlations, and in the present study we estimated correlations between the categories displayed in the space.

To do so, we discretized continuous Wechsler intelligence and memory scores into three levels (high, medium, and low), according to the suggestion by Lehmann and Hurlbert (1972), and then investigated the interaction of racial/ethnic group differences with gender (Gender $\times$ Races) and their association with specific cognitive ability levels. Specifically, we planned to study the magnitudes and directions of these variables' correlations with general ability, specific cognitive abilities, and memory abilities. Several steps are needed in order to conduct a CA biplot analysis as illustrated in the present study:

- (1) Discretize continuous cognitive ability scores into three mutually exclusive achievement-level categories: high, medium, and low;
- (2) construct a two-way contingency table with the interaction of gender and racial/ethnic group as the row categories (gender–race) and discretized ability levels as the column categories;
- (3) Conduct CA with the constructed two-way table in order to estimate the row and column dimensions;
- (4) Test the statistical significance of the dimensional eigenvalues or principal inertias;
- (5) Construct a two-dimensional map with statistically significant dimensions and biplot the row and column categories in the map, in order to visually approximate the category associations;
- (6) Finally, estimate the category associations with correlations, to enhance their interpretation.

Method

Participants

A total of 900 participants between the ages 16 of 70 comprised the standardization sample of the WAIS-IV and Wechsler Memory Scale—Fourth Edition (WMS-IV) con-normed dataset (Wechsler, 2008, 2009). According to test manuals, the self-reported breakdown by race/ethnicity in the standardization dataset was 62.2% White ($n = 560$), 3.3% Asian ($n = 30$), 15.6% Black ($n = 140$), 16.3% Hispanic ($n = 147$), and 2.6% “Other” ($n = 23$). The gender breakdown was 50.3% female ($n = 453$), 49.7% male ($n = 447$). For the present study, only Black, Hispanic, and White subpopulation groups were analyzed.

Instruments

Wechsler Adult Intelligence Scale—Fourth Edition The WAIS-IV consists of ten core subtests from which users can derive a FSIQ score, as well as four index scales. The four

WAIS-IV index scales are Verbal Comprehension (*vc*), Perceptual Reasoning (*pr*), Working Memory (*wo*), and Processing Speed (*ps*). The *vc* index measures abstract verbal reasoning, comprehension and expression of increasingly difficult vocabulary words, and knowledge of general information acquired from the culture. The *pr* index measures spatial perception, visual abstract processing, nonverbal problem-solving, and nonverbal inductive reasoning. The *wo* index measures attention and concentration while mentally manipulating stimuli. The *ps* index measures visual perception and analysis, visual scanning speed, and visual working memory (Wechsler, 2008).

Wechsler Memory Scale—Fourth Edition The WMS-IV (Wechsler, 2009) is designed to assess a wide variety of memory abilities in individuals. The WMS-IV consists of ten subtests, which can be consolidated into five index scales. The five WMS-IV index scales are Immediate Memory (*im*), Delayed Memory (*dm*), Visual Working Memory (*wm*), Visual Memory (*vm*), and Auditory Memory (*am*). The *am* index measures the ability to retell brief stories or lists of word pairs immediately after hearing them. The *vm* index measures the ability to recall abstract designs (by drawing these immediately after presentation) or the configuration of picture cards on a grid (immediately after presentation). The *wm* index assesses the ability to remember a set of addition and subtraction rules, or the temporal order of visual objects, by arranging circles on a visual grid or arranging visual objects in the correct sequence. The *dm* index comprises all tests that require participants to recall or recognize stimuli after a time delay in which other interfering memory tasks are presented. The *im* index comprises all tests that require participants to recall stimuli immediately after presentation.

New aspects in the present study If we were to analyze this dataset in its original, continuous form, we could measure racial/ethnic or gender differences with FSIQ and index mean scores utilizing the *t* test or ANOVA procedures, assuming that the data are normally distributed. We would also need to check the normality of the FSIQ and index scores. Even if the normality assumption were met, we would not be able to clearly identify which gender or racial/ethnic group was most closely related to certain performance levels of cognitive abilities. To accomplish this objective, we discretized the FSIQ and index scores into high, medium, and low categories. If examinees' scores were at least one standard deviation (15 points) above the mean (100) of the entire standardization sample, they were assigned a prefix of “3,” which stood for the category of high performance. If examinees' scores were within one standard deviation above or below the mean, they were identified with the prefix “2,” which stood for the category of medium performance. If examinee scores were at least one standard deviation below the mean, they were identified

with the prefix “1,” which stood for the category of low performance. Thus, to analyze the data using the CAB procedure, we generated a 6 (two categories of the gender variable $\times$ three categories of the race/ethnicity variable) by 30 (three categories of performance levels of FSIQ and nine WAIS-IV and WMS-IV index score categories) contingency table. With the CAB results, first we visually inspected the locations of Gender $\times$ Racial/Ethnic group combinations along with 30 performance levels of the various cognitive abilities. This included FSIQ in a two-dimensional map, where the horizontal axis represented the first dimension and the vertical axis represented the second dimension. Then we estimated the correlations between Gender $\times$ Racial/Ethnic groups and their performance on discretized WAIS-IV index scores, WMS-IV index scores, and FSIQ, and also estimated the correlations of Gender $\times$ Racial/Ethnic groups with the first and second dimensions. The first dimension was assumed to represent a general ability factor for cognitive ability, and the second dimension was a group factor for various types of cognitive ability.

A hybrid approach: Combining interactive coding and stacking variables A multiway frequency table involves multiple variables with several categories, and the variables can be interactively coded in order to run CA (Greenacre, 2016). For example, we coded gender and racial/ethnic groups interactively, such as female and male Blacks, female and male Hispanics, and female and male Whites. However, when too many variables are involved, it is not practically meaningful to interactively code all variables for CA. In our contingency table, ten variables were involved in the columns (nine index variables + one FSIQ variable), where each variable consisted of three performance levels. If we had interactively coded them, there would be $3^{10} = 59,049$ combinations for CA. Thus, as an alternative procedure, we utilized the stacking approach (recommended by Greenacre, 2016) for the column variables. We stacked ten categorical variables. When the interactively coded rows and the stacked columns were combined, a 6 (Gender $\times$ Racial/Ethnic Group) $\times$ 30 (cognitive ability categories) table was formed. This hybrid approach that combined interactively coding and stacking not only simplified the estimation procedure but also reinforced our aim. That is, our aim was to study the associations between Gender $\times$ Racial/Ethnic group and performance levels on different cognitive ability variables.

A significance test for principal inertia We conducted a permutation test to determine the statistical significance of the principal inertias (e.g., Greenacre, 2016; Greenacre and Primmerio, 2013). For a permutation simulation, 10,000 random contingency tables are generated by series of permutations of the empirical data. A contingency table consists of the cross-tabulation matrix of row and column categories. There

are several steps in obtaining random contingency tables: (a) apply the raw data related to the (original) observed contingency table with one (interactively coded) variable from the rows and another variable (made from several stacked variables) from the columns; (b) independently permute the categories from either a row or a column variable of the created raw data; (c) cross-tabulate the permuted data in order to obtain a random contingency table; (d) repeat Steps a to c 10,000 times; (e) conduct CA of each random contingency table in order to estimate the random inertias (10,000 of them) for each dimension; (f) plot the 10,000 random inertias with a real, observed inertia in each dimension; and finally, (g) count random inertias larger than the (real) observed principal inertia in order to compute an empirical p value. For example, if fewer than 500 of the simulated principal inertias in a dimension were larger than the observed principal inertia, its p value would be less than .05, and the observed inertia of the dimension would thus be statistically significant.

Importance of estimating nonlinear category associations with linear correlations To quantify the visual approximation of the category associations, we needed to map the related, nonlinear chi-square statistics (estimated from CA) into the linear Euclidean-endowed space. The Euclidean distances can be estimated between categories using their coordinate values in a (Euclidean-endowed) biplot. Any two points in a biplot measured by a distance (Euclidean or chi-square) do not carry information about whether their relationships are close or distant until their distance is compared with the distance between another two points, since distances are not bounded. Moreover, distances do not inform us as to whether the category relationships are positive or negative. On the other hand, correlations are bounded between -1 and $+1$ and have a direction (negative or positive), and squared correlations can be interpreted as the shared variance between two categorical points. Therefore, we estimated correlations in order to enhance the interpretation of category associations.

Estimation of correlations with coordinates Let us assume a two-dimensional map (made from the first two dimensions) and the principal row category i and i' coordinates $\mathbf{f}_i = (f_{i1} \ f_{i2})$ and $\mathbf{f}_{i'} = (f_{i'2} \ f_{i'2})$. In Euclidean geometry, a scalar product between two vectors $\mathbf{f}_i$ and $\mathbf{f}_{i'}$ is denoted $\mathbf{f}_i^T \mathbf{f}_{i'}$, which is equal to the product of the lengths of the two vectors, multiplied by the cosine of the angle between them:

$$\mathbf{f}_i^T \mathbf{f}_{i'} = \|\mathbf{f}_i\| \cdot \|\mathbf{f}_{i'}\| \cdot \cos \theta \quad (2)$$

where $\|\mathbf{f}_i\|$ denotes the length of the vector $\mathbf{f}_i$. From Eq. 2,

$$\cos \theta_{ii'} = \frac{\mathbf{f}_i^T \mathbf{f}_{i'}}{\|\mathbf{f}_i\| \cdot \|\mathbf{f}_{i'}\|} \quad (3)$$

where $\cos \theta_{ii'}$ is a correlation estimate between rows i and i' . Similarly, in a given map, the correlation between columns j and j' is estimated by $\cos \theta_{jj'} = \gamma_j^T \gamma_{j'} / \|\gamma_j\| \cdot \|\gamma_{j'}\|$, where γ = standard column category coordinates, and the correlation between row i and column j is estimated by $\cos \theta_{ij} = \mathbf{f}_i^T \gamma_j / \|\mathbf{f}_i\| \cdot \|\gamma_j\|$ (for details, see Kim & Grochowalski, [in press](#)).

Results

Checking normality

First we checked the distributions of FSIQ and index scores for Blacks, Hispanics, and Whites. It is reasonable to expect that the separate index scores would not be normally distributed if

Fig. 1 Histograms of full-scale IQ scores for Blacks, Hispanics, and Whites, where the vertical axes represent frequencies and the horizontal axes, full-scale IQ score ranges.

the FSIQ scores, constructed using a linear combination of the index scores, were not normally distributed. As is shown in Fig. 1, the FSIQ scores for Blacks and Hispanics were positively skewed, and we also examined the distributions of the separate index scores, and consistent with our expectation, they were not normally distributed, either, indicating that any statistics (e.g., t test, ANOVA, or MANOVA) that assume the normality of scores should be used cautiously for examining FSIQ or index score differences across racial/ethnic groups.

Test significance of principal inertias (or dimensional eigenvalues)

We conducted CAB of a 6 (male and female Blacks, male and female Hispanics, and male and female Whites) $\times$ 30 (high, medium, and low levels of FSIQ + high, medium, and low levels of nine index scores) contingency table, and in total, five dimensions were extracted. The maximum dimensionality in our case was 5, or $\min(6 - 1, 30 - 1)$. Utilizing a permutation simulation, we estimated the empirical p values for the five principal inertias (or dimensional eigenvalues); they were .0000, .0000, .0140, .7783, and .9997, respectively.

The results showed that the first three dimensions were statistically significant at $\alpha = .05$. We also plotted the observed inertias, the mean inertias from 10,000 simulated inertias, and the 95th percentile from 10,000 simulated inertias over dimensions. As is shown in Fig. 2, the first three observed inertias were larger than the simulated inertias, but we used the only first two dimensions to construct a plane for our biplot analysis, because they accounted for the largest amount of variance

Fig. 2 Observed inertias, juxtaposed with the simulated inertias from 10,000 permutations. *Observed* = inertias from the observed data; *Sim. Mean* = mean of 10,000 simulated inertias with permutations; *Sim. 95%* = 95th percentile of the 10,000 simulated inertias with permutations.

(94% of total variance): 86% for the first dimension and 8% for the second dimension.

Visual inspection of category associations

The biplot aspect of CAB allowed us to visually inspect the category associations in the two-dimensional plane. As is shown in Fig. 3, all male and female Blacks and Hispanics were located to the left of the origin, and all male and female Whites were positioned to the right side of the origin. On the right side, all high- and medium-level performance on FSIQ and the different indices was positioned, whereas on the left side were all low levels of FSIQ and the indices. Visual inspection indicated that Blacks and Hispanics in general, irrespective of gender, were associated with low performance levels in FSIQ and the specific types of intelligence and memory ability. However, both male and female Whites were related to high- and medium-level performance in the same domains. The first dimension was labeled as the low versus medium/high performance dimension, and the second as a male versus female dimension. Although the category relationships were easy to approximate through visual inspection, these results did not provide numerical estimates for the category associations. As a result, we needed to estimate the (visually inspected) category associations through correlations.

Estimating category associations through correlations

Correlations of Gender $\times$ Racial/Ethnic groups with Dimensions 1 and 2 and FSIQ

First, we estimated the correlations of the female Black subgroup (fB), the female Hispanic subgroup (fH), and the female White subgroup (fW) with Dimension 1 (Dim 1), which was assumed to represent a general ability factor. These results are summarized in Table 1. The correlations were $r(\text{fB}, \text{Dim 1}) = .99$, $r(\text{fH}, \text{Dim 1}) = .93$, and $r(\text{fW}, \text{Dim 1}) = .86$. The correlations with Dimension 2 (Dim 2; which was assumed to represent a group factor for various types of cognitive abilities) were $r(\text{fB}, \text{Dim 2}) = 0$, $r(\text{fH}, \text{Dim 2}) = .27$, and $r(\text{fW}, \text{Dim 2}) = .50$. The correlational results imply that Dim 1 was substantially related with all the female groups, but female Whites were also substantially related with Dim 2. Second, we examined the correlations of male Blacks (mB), male Hispanics (mH), and male Whites (mW) with Dim 1. The correlations were $r(\text{mB}, \text{Dim 1}) = .92$, $r(\text{mH}, \text{Dim 1}) = .86$, and $r(\text{mW}, \text{Dim 1}) = .92$. The correlations with the group factor Dim 2 were $r(\text{mB}, \text{Dim 2}) = .20$, $r(\text{mH}, \text{Dim 2}) = .08$, and $r(\text{mW}, \text{Dim 2}) = .38$. The male correlational results imply that all male groups were highly related with Dim 1.

The correlations of the three performance levels (low, medium, and high) of FSIQ with female/male Blacks, female/male Hispanics, and female/male Whites were also summarized in Table 1.

Fig. 3 Biplot of Gender $\times$ Racial/Ethnic group and Wechsler index categories. f = female; m = male; B = Blacks; H = Hispanics; W = Whites; vc = WAIS-IV Verbal Comprehension; pr = WAIS-IV Perceptual Reasoning; wo = WAIS-IV Working Memory; ps = WAIS-

IV Processing Speed; im = WMS-IV Immediate Memory; dm = WMS-IV Delayed Memory; wm = WMS-IV Visual Working Memory; vm = WMS-IV Visual Memory; am = WMS-IV Auditory Memory. 1 = low performance, 2 = medium performance, and 3 = high performance.

Correlations of intelligence and memory indexes with Gender $\times$ Racial/Ethnic groups Blacks and Hispanics, irrespective of gender difference, were positively correlated with low performance (denoted as “1”) indexes, but negatively correlated with the medium and high indexes. Whites were generally positively related with either medium (denoted as “2”) or high (denoted as “3”) performance indexes, but a gender difference was manifest. Positive correlations with index categories equal to or larger than .50 are included in Table 2.

Table 1 Estimated correlations of female/male Blacks, Hispanics, and Whites with Dimensions 1 and 2 and FSIQ

Dim 1		Dim 2		FSIQ	High	Medium	Low
fB	.99	fB	0	fB	−.88	−1.00	.99
mB	.92	mB	.20	mB	−.63	−.84	.78
fH	.93	fH	.27	fH	−.94	−.69	.76
mH	.86	mH	.08	mH	−.84	−.96	.93
fW	.86	fW	.50	fW	.20	.48	−.40
mW	.92	mW	.38	mW	.78	.55	−.64

f = female; m = male; B = Blacks; H = Hispanics; W = Whites

Female Blacks, female Hispanics, and male Hispanics were related with both intelligence and memory ability, whereas male Blacks were related with only memory ability. Whites, regardless of gender, were highly related with either medium or high performance indexes, but female Whites were related with high levels of processing speed, immediate memory, delayed memory, and auditory memory, but male Whites were related with none of these.

Discussion

In the present study we introduced the correspondence analysis biplot procedure, which is a distribution-free method that estimates category associations with correlations. The results from multivariate statistics (e.g., MANOVA or factor analysis) operate under the assumption of normality of the data and should not be accepted as valid when the normality assumption is violated.

Since gender and racial/ethnic groups are naturally occurring categorical variables, this does not lend itself to linear relationships between these categorical variables and continuous

Table 2 Positive correlation results for female and male Black, Hispanic, and White subgroups with high (3), medium (2), and low (1) WAIS and WMS index scores

fB		mB		fH		mH		fW		mW	
vc1	.82	vc1	<.5	vc1	.97	vc1	.66	vc2	.73	vc3	.99
pr1	1.00	pr1	.88	pr1	.62	pr1	.98	pr2	1.00	pr3	1.00
wo1	.87	wo1	<.5	wo1	.95	wo1	.72	wo2	1.00	wo3	.98
ps1	.52	ps1	.89	ps3	.58	ps1	.71	ps3	.95	ps	<.5
im1	1.00	im1	.83	im1	.70	im1	.96	im3	.92	im	<.5
dm1	.90	dm1	1.00	dm	<.5	dm1	.98	dm3	.82	dm	<.5
wm1	.97	wm1	.96	wm	<.5	wm1	1.00	wm2	.99	wm3	1.00
vm1	1.00	vm1	.83	vm1	.71	vm1	.96	vm2	1.00	vm3	.98
am1	.72	am1	.98	am1	<.5	am1	.87	am3	.99	am	<.5

1 = low performance; 2 = medium performance; and 3 = high performance. f = female; m = male; B = Blacks; H = Hispanics; and W = Whites. Only correlation coefficients larger than 0.5 were included here and Black and Hispanic were positively correlated with all low performance levels, except “ps3” in female Hispanic. We boxed the gender differences appearing in racial/ethnic groups. vc = Verbal Comprehension; pr = Perceptual Reasoning; wo = Working Memory; ps = Processing Speed; im = Immediate Memory; dm = Delayed Memory; wm = Visual Working Memory; vm = Visual Memory; am = Auditory Memory.

variables. Special polyserial correlation estimates are exceptions, which are not usually available in popular statistical packages such as SPSS. However, even polyserial correlations cannot provide information about interactive relationships between Gender $\times$ Racial/Ethnic groups and cognitive ability levels. Thus, we utilized the CAB approach in order to meet our aims. To the best of our knowledge, our present study has been the first one to estimate correlations between gender interactions with racial/ethnic groups and discretized Wechsler intelligence and memory index scores.

Utilities of CAB over conventional approaches

Because gender and races are categorical variables, it may not be easy to estimate correlations between gender, race, and continuous scores. Usually the numbers of females and males (and different races, as well) are not matching, and estimating simple correlations between gender or different races and any cognitive subscale scores may not be easy. Even if the sample sizes are matched, a simple correlational approach cannot examine relationships between gender and race in terms of specific performance levels in cognitive ability, unlike the CAB results shown here.

Through visual inspection of Gender $\times$ Racial/Ethnic Group locations in a biplot (see Fig. 2), one can easily study group differences. Dimension 1 depicted low (Black and Hispanic) versus medium/high (White) performance levels, and Dimension 2 separated males (above the origin line) and females (below the origin line). The previous studies we reviewed had reported few gender differences (e.g., Voyer et al., 1995), except in spatial ability. In our results, however, gender differences were manifest across all racial/ethnic groups. First, in Dimension 2, differential correlational

patterns were found for gender across racial/ethnic groups (see Table 1). Second, there were gender differences in correlations with intelligence and memory index scores across races (see the boxed numbers in Table 2); (1) for Blacks, there were gender differences in the correlational patterns for verbal comprehension and working memory; (2) for Hispanics, gender differences were found in processing speed, delayed memory, and visual working memory; and (3) for Whites, gender differences were found in processing speed, immediate memory, delayed memory, and auditory memory. Unlike the previous gender difference studies, our results have provided richer information about gender differences over races. Such differential correlations were not found in any previous studies (e.g., Hartmann et al., 2006; Jensen, 1998). Future research will need to examine whether similar gender and racial/ethnic group differences exist (in correlational patterns) in a co-normed newer version of the Wechsler Adult Intelligence and Memory scales.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence*, 33, 431–444.
- Beh, E. J., & Lombardo, R. (2014). Correspondence analysis: Theory, practice and new strategies. West Sussex, UK: John Wiley & Sons, Ltd. doi:<https://doi.org/10.1002/9781118762875>.
- Blasius, J., & Greenacre, M. (Eds.). (2014). Visualization and verbalization of data. Boca Raton, FL: Chapman & Hall/CRC Press.
- Bradru, D., & Gabriel, K. R. (1978). The biplot as a diagnostic tool for models of two-way tables. *Technometrics*, 20, 47–68.

- Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since bias in mental testing. *School Psychology Quarterly*, 14, 208–238.
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41, 189–225.
- Chen, T.-H., Kaufman, A. S., & Kaufman, J. C. (1994). Examining the interaction of Age $\times$ Race pertaining to black–white differences at ages 15 to 93 on six horn abilities assessed by K-FAST, K-SNAP, and KAIT subtests. *Perceptual and Motor Skills*, 79, 1683–1690.
- Colom, R., & Lynn, R. (2004). Testing the developmental theory of sex differences in intelligence on 12–18 year olds. *Personality and Individual Differences*, 36, 75–82.
- Cox, D. R. (1957). Note on grouping. *Journal of the American Statistical Association*, 52, 543–547.
- Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, 35, 21–50.
- Dolan, C. V., & Hamaker, E. L. (2001). Investigating black-white differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC-R and K-ABC, and a critique of the method of correlated vectors. In F. Columbus (Ed.), *Advances in psychological research* (Vol. 6, pp. 31–60). Huntington, NY: Nova Science.
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, 43, 95–103. doi:<https://doi.org/10.1037/0003-066X.43.2.95>
- Frisby, C. L., & Beaujean, A. (2015). Testing Spearman's hypothesis using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence*, 51, 79–97.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58 (3), 453–467.
- Gabriel, K. R. (2002). Goodness of fit biplots and correspondence analysis. *Biometrika*, 89, 423–436.
- Gabriel, K. R., & Odoroff, C. I. (1990). Biplots in biomedical research. *Statistics in Medicine*, 9, 469–485.
- Gottfredson, L. (1998). Reconsidering fairness: A matter of social and ethical priorities. *Journal of Vocational Behavior* 33, 293–319.
- Gottfredson, L. S. (2005). Implications of cognitive differences for schooling within diverse societies. In C. L. Frisby & C. R. Reynolds (Eds.), *Comprehensive handbook of multicultural school psychology* (pp. 517–554). Hoboken, NJ: Wiley.
- Gower, J. C., & Hand, D. J. (1996). Biplots (Monographs on Statistics and Applied Probability, No. 54). London, UK: Chapman & Hall.
- Gower, J., Lubbe, S., & le Roux, N. (2011). Understanding biplots. Hoboken, NJ: Wiley.
- Green, P. E., & Rao, V. (1970). Rating scales and information recovery—How many scales and response categories to use? *Journal of Marketing*, 34, 33–39.
- Greenacre, M. J. (1993). Biplots in correspondence analysis. *Journal of Applied Statistics*, 20, 251–269.
- Greenacre, M. J. (2010). Biplots in practice. Madrid, Spain: Fundación BBVA.
- Greenacre, M. J. (2016). Correspondence analysis in practice (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Greenacre, M. J., & Primmer, R. (2013). Multivariate analysis for ecological data. Bilbao, Spain: Fundación BBVA.
- Harrington, D. (2009). Confirmatory factor analysis. New York, NY: Oxford University Press.
- Hartmann, P., Kruuse, N., & Nyborg, H. (2006). Testing the cross-racial generality of Spearman's hypothesis in two samples. *Intelligence*, 35, 47–57.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41–45. doi:<https://doi.org/10.1126/science.7604277>
- Herrnstein, R. J., & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. New York, NY: Free Press.
- Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using ω^2 and d . *American Psychologist*, 36, 892–901. doi:<https://doi.org/10.1037/0003-066X.36.8.892>
- Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta analysis. *Psychological Bulletin*, 104, 53–69. doi:<https://doi.org/10.1037/0033-2909.104.1.53>
- Jacoby, J., & Matell, M. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research*, 8, 495–500.
- Jensen, A. R. (1973). Level I and Level II abilities in three ethnic groups. *American Educational Research Journal*, 10, 263–276.
- Jensen, A. R. (1980). Bias in mental testing. New York, NY: Free Press.
- Jensen, A. R. (1985). The nature of the Black-White difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, 8, 193–263.
- Jensen, A. R. (1992). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research*, 27, 225–233.
- Jensen, A. R. (1998). The g factor: The science of mental ability. Westport, CT: Praeger/Greenwood.
- Jensen, A. R., & Johnson, F. W. (1994). Race and sex differences in head size and IQ. *Intelligence*, 18, 309–333. doi:[https://doi.org/10.1016/0160-2896\(94\)90032-9](https://doi.org/10.1016/0160-2896(94)90032-9)
- Johari, S., & Sclove, S. L. (1976). Partitioning a distribution. *Communications in Statistics—Theory and Methods*, A5, 133–147.
- Kim, S.-K., & Grochowalski, J. H. (in press). Exploratory visual inspection of category associations and correlation estimation in multidimensional subspaces. *Journal of Classification*.
- Kim, S.-K., McKay, D., Taylor, S., Tolin, D. F., Olatunji, B. O., Timpano, K. R., & Abramowitz, J. S. (2016). The structure of obsessive compulsive symptoms and beliefs: A correspondence and biplot analysis. *Journal of Anxiety Disorders*, 38, 79–87.
- Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271–280. doi:<https://doi.org/10.1007/s004420100716>
- Lehmann, D., & Hurlbert, J. (1972). Are three-point scales always good enough? *Journal of Marketing Research*, 9, 444–446.
- Le Roux, B., & Rouanet, H. (2004). Geometric data analysis: From correspondence analysis to structured data. Dordrecht, The Netherlands: Kluwer.
- Lynn, R. (2006). Race differences in intelligence: An evolutionary analysis. Augusta, GA: Washington Summit.
- Maccoby, E. E., & Jacklin, C. N. (1974). The psychology of sex differences. Stanford, CA: Stanford University Press.
- Mackintosh, N. J. (1998). IQ and human intelligence. New York, NY: Oxford University Press.
- Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77–101. doi:<https://doi.org/10.1037/0003-066X.51.2.77>
- Nishisato, S. (1980). Analysis of categorical data: Dual scaling and its applications. Toronto, ON: University of Toronto Press.
- Nishisato, S. (1994). Elements of dual scaling: An introduction to practical data analysis. Hillsdale, NJ: Erlbaum.
- Nishisato, S. (1999). Data types and information: beyond the current practice of data analysis. In R. Decker & W. Gaul (Eds.), *Classification and information processing at the turn of the millennium* (pp. 40–51). Heidelberg, Germany: Springer.
- Nishisato, S. (2007). Multidimensional nonlinear descriptive analysis. Boca Raton, FL: Chapman & Hall/CRC.
- Pearson, K. (1904). Mathematical contribution to the theory of evolution: XIII. On the theory of contingency and its relation to association and normal correlation. *Drapers' Company Research Memories, Biometric Series, I*, 1–35.

- Reynolds, C. R., & Kaiser, S. M. (1990). Test bias in psychological assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (pp. 487–525). Chichester, UK: Wiley.
- Reynolds, C. R., & Suzuki, L. A. (2013). Bias in psychological assessment: An empirical review and recommendations. In J. R. Graham, J. A. Naglieri, & I. B. Weiner (Eds.), *Handbook of psychology: Assessment psychology* (pp. 82–113). Hoboken, NJ: Wiley.
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54, 297–330.
- Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, 11, 235–294. doi:<https://doi.org/10.1037/1076-8971.11.2.235>
- Sackett, P. R., & Wilk, S. L. (1994). Within group norming and other forms of score adjustments in pre-employment testing. *American Psychologist*, 49, 929–954.
- Shaw, D. G., Huffman, M. D., & Haviland, M. G. (1987). Grouping continuous data in discrete intervals: Information loss and recovery. *Journal of Educational Measurement*, 24, 167–173.
- Sourial, N., Wolfson, C., Zhu, B., Quail, J., Fletcher, J., Karunanathan, S., . . . Bergman, H. (2010). Correspondence analysis is a useful tool to uncover the relationships among categorical variables. *Journal of Clinical Epidemiology*, 63, 638–646.
- Spearman, C. E. (1927). *The abilities of man: Their nature and measurement*. New York, NY: Blackburn Press.
- Stefansky, W., & Kaiser, H. F. (1973). Note on discrete approximations. *Journal of the American Statistical Association*, 68, 232–234.
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York, NY: Wiley.
- Ter Braak, C. J. F. (1983). Principal components biplots and alpha and beta diversity. *Ecology*, 64, 454–462.
- Ter Braak, C. J. F., & Verdonschot, P. E. M. (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences*, 57, 255–289.
- Vernon, P. A. (1981). Level I and Level II: A review. *Educational Psychologist*, 16, 45–64.
- Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117, 250–270. doi:<https://doi.org/10.1037/0033-2909.117.2.250>
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth edition: Technical and interpretive manual*. San Antonio, TX: Pearson.
- Wechsler, D. (2009). *Wechsler Memory Scale—Fourth edition: Technical and interpretive manual*. San Antonio, TX: Pearson.
- Weiss, L. G. (2003). The WISC-III in the United States. In J. Georgas, L. G. Weiss, F. van de Vijver, & D. H. Saklofske (Eds.), *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (pp. 41–59). Amsterdam, The Netherlands: Academic Press.
- Weiss, L. G., Chen, H., Harris, J. G., Holdnack, J. A., & Saklofske, D. H. (2010). WAIS-IV use in societal context. In L. G. Weiss, D. H. Saklofske, D. Coalson, & S. E. Raiford (Eds.), *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 97–139). Amsterdam, The Netherlands: Academic Press.
- Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (2015). *WISC-V assessment and interpretation: Scientist-practitioner perspectives*. Amsterdam, The Netherlands: Elsevier/Academic Press.