Re-Visiting Exploratory Methods for Construct Comparability: Is There Something to be Gained From the Ways of Old?

Bruno D. Zumbo
University of British Columbia

Stephen G. Sireci and Ronald K. Hambleton
University of Massachusetts – Amherst

Running Head: Exploratory Methods

Presented in the symposium *Construct Comparability Research: Methodological Issues and Results,* organized by Kadriye Ercikan, for the National Council on Measurement in Education (NCME) meetings, April 2003, Chicago, Illinois.

Send Correspondence to:

Professor Bruno D. Zumbo
University of British Columbia
Dept. of ECPS, Scarfe Building, 2125 Main Mall
Vancouver, B.C.
CANADA V6T 1Z4

Fax: (604) 822-3302
Phone: (604) 822-1931
e-mail: bruno.zumbo@ubc.ca

Version: April 19, 2003

Abstract

This paper reviews methodological issues that arise in the investigation of construct comparability across key comparison groups such as ethnic and gender groups, or adapted/translated versions of tests in the same or different cultures. The authors advocate a multi-method approach to investigating construct comparability. In particular, multi-group exploratory factor analysis is described, in the context of an example, as a complement to the standard multi-group confirmatory factor analysis. We also describe a graphical method of investigating if the congruence coefficients in multi-group exploratory factor analysis may be spuriously inflated, hence strengthening the exploratory methodology. The example, from the Canadian School Achievement Indicators Program 1996 Science Assessment, shows how the confirmatory approach may not support construct comparability when the exploratory does. Reasons for why this may happen, and why the exploratory approach is a good complement, are discussed.

Re-Visiting Exploratory Methods for Construct Comparability: Is There Something to be Gained From the Ways of Old?

As Kadriye Ercikan, this symposium's convener, reminds us, questions of construct comparability frequently arise when considering key comparison groups. The key comparison groups that are most commonly considered are different ethnic and gender groups. In addition, the potential for differences in constructs for examinees who take different language versions of tests, which are from different cultural groups and those who take tests in a language different than their first language, is well documented.

The matter of construct comparability addresses, at its basic level, whether individuals from different groups respond in the same or similar manner to items, tasks, or other such measurement opportunities. If one can demonstrate that individuals from different groups are responding similarly then one can infer that these individuals attach the same meaning to the construct as a whole. In short, then, like the classical reliability of tests, construct comparability is a property of data (actually, the intersection of respondents and tasks) and hence establishing construct comparability becomes an empirical question.

The empirical test of construct comparability is typically examined by a pair wise comparison of factors (i.e., latent variables) or dimensions across groups. Although there are several methods such as exploratory factor analysis, multidimensional scaling (MDS) and cluster analysis available, in educational measurement like many other social sciences, multi-group confirmatory factor analysis (MG-CFA) has become the standard and commonly recommended approach to investigating construct comparability. In this paper, we (a) review some methodological issues that suggest that MG-CFA is not always necessarily the only choice, and (b) review some exploratory factor analysis methods that may be used along with MG-CFA to investigate construct comparability. It should be noted that the use of multiple methods (i.e., MDS and CFA) has been advocated and demonstrated by Sireci and Allalouf (2003), and Robin, Sireci, and Hambleton (in press) for investigating construct comparability. In this paper we expand that repertoire to include multi-group exploratory factor analysis (MG-EFA) and highlight how MG-EFA is different than MG-CFA. This paper begins with an overview of MG-CFA and then turns to some of the exploratory methods.

Factor Analysis Methods for Construct Comparability

Before turning to a detailed description of the confirmatory and exploratory factor analytic methods, let us make a few remarks about the factor analysis methods in general. The factor analysis methods for construct comparability are guided by two important fundamental principles (see Meredith, 1993 for a detailed overview and analytical description):

1. Construct comparability is established if the factor loadings of the items (i.e., the regressions of the items on to the latent variables) are invariant across groups. Therefore, if one is investigating the construct comparability across males and females of, for example, a large-scale test of science knowledge one needs to focus on establishing the similarity of the factor loadings across the groups of males and females.
2. Staying with our example of large-scale testing of science knowledge, it is important to note that the results of factor analyses based on the two groups combined may be quite different than if one were to analyze the same two groups separately. That is, when conducting factor analyses if a set of factors are orthogonal in the combined population of men and women then the same factors are typically correlated when the

> groups are analyzed separately. However, if one finds that the factors are correlated in the separate groups this does not imply that the same set of factors are orthogonal in the combined population of men and women. A practical implication of this principle is that one need not be surprised if the factor analysis results found in studies of combined groups does not replicate when the groups are examined separately with multi-group methods. Likewise, one must proceed cautiously in translating results from combined group factor analyses to separate group factor analyses.

The matter of construct comparability hence becomes one of comparing the factor solutions that have been conducted separately for the groups of interest. The comparison should be of a statistical nature involving some sort of index or test of similarity, rather than purely impressionistic, because factor loadings, like all statistics, are effected by sampling and measurement error.

An early statistical approach to comparing the factor solutions across groups is provided by exploratory factor analysis. The strategy is to quantify or measure the similarity of the factor loadings across groups by rotating the two factor solutions to be maximally similar and then computing some sort of similarity index. The emphasis in the previous sentence is on *quantifying* or *measuring* the similarity rather than computing a formal statistical hypothesis test. The formal statistical hypothesis test involving simultaneous factor analysis in several populations awaited the development of a statistical estimation theory (in particular maximum likelihood methods; rather than the numerical methods such as principal components) and the corresponding sampling theory for the factor analysis statistics (e.g., factor loadings) provided by Jöreskog (1971). In the MG-CFA approach there are a number of well-known testable hypotheses for full or partial measurement invariance between groups (Byrne, 1994; Byrne, Shavelson, & Muthén, 1989; Jöreskog, 1971).

At this juncture, two points are worth noting. First, the exploratory factor analysis methods were superseded by the MG-CFA because MG-CFA's reliance on a formal hypothesis test made the decision of factorial comparability, as well as the whole factor analysis enterprise objective, at the very least it appears so on a surface analysis. In addition, the reliance on exploratory factor analysis' similarity indices may capitalize on sample-specific subtleties (sometimes referred to as capitalizing on chance) and these similarity indices can be spuriously large and misleading in some situations (see, Barrett, 1986; Horn, 1967; Horn & Knapp, 1973; Korth & Tucker, 1975; Paunonen, 1997; ten Berge, 1986).

Second, there is an understanding now that the sole reliance on MG-CFA may also be limiting for the following three reasons.

(a) CFA generally appears to perform optimally when the factor pattern solution exhibits simple structure (see, for example, Church & Burke, 1994), typically, with each variable measuring one and only one factor in the domain. Of course, one should be reminded that simple structure was introduced by Thurstone as a way of side-stepping the rotational problem in exploratory factor analysis: the problem is that when one has more than two factors there is no way of choosing among the many possible rotations that perform equally well in describing the data. Thurstone's solution relied on a form of parsimony and the extent to which it is a reasonable solution is still debatable. Furthermore, it has been recently highlighted, however, in the assessment of dimensionality for educational measures that many tests, items or tasks may measure multiple factors albeit some of them minor factors. In short, clean simple structure is an ideal that may not be seen with many tests. Clearly, then, the application of CFA hypothesis tests to measures may be potentially misleading because CFA methods have a strict requirement that "off-loadings" on minor factors are zero, when in reality they may be non-zero but still relatively small in magnitude compared to the primary target factor of interest. This has come to be referred to in some of the psychometric literature as the situation of trailing minor factors.

(b) The statistical estimation and sampling theory of the commonly used estimators in CFA assume continuous observed variables. Educational measures, however, often have skewed and/or binary (or typically at most 5-point rating scale) item response data. In addition, the item scores do not only have to be continuous variables but the chi-squared test is valid under multivariate normality of the observed item scores.

(c) Many of the fit and modification indices need large sample sizes before they begin to evince their asymptotic properties.

The above remarks regarding the MG-CFA and MG-EFA focus on the reliance of one method, in fact either method, to the exclusion of the other. We are proposing that together MG-CFA and MG-EFA complement each other well. We are, in essence, advocating an extensive multi-method approach to construct comparability. In short, we have much to learn from the ways of old – i.e., exploratory factor analysis. For example, MG-EFA does not assume a clean simple structure nor does it assume strict dimensionality; therefore, if one is investigating construct comparability one can start with MG-CFA and if strong invariance (as described below) is not supported by the data, then one can move to MG-EFA because it allows for possible secondary minor factors and does not necessitate a clean simple structure. Likewise, MG-EFA is less stringent in terms of sample size and the distributional assumptions of the observed item responses.

Demonstration of the Methodology with an Example Data Set

We will demonstrate the use of the multi-method strategy in the context of a real data set. The School Achievement Indicators Program (SAIP) 1996 Science Assessment was developed by the Council of Ministers of Education, Canada (CMEC) to assess the performance of 13- and 16-year old students in science. There are three forms (Forms A, B, and C) in the Science Assessment. The SAIP 1996 Science Assessment was administered to 13-year-old students in two languages, namely, English and French. They were 5261 English-speaking and 2027 French-speaking students. This study used the 66 items from the assessment data from Form B (13-year olds).

Multi-Group Confirmatory Factor Analysis

The most common application of MG-CFA involves testing two hypotheses (i) whether the factor loadings are the same across the two groups, and (ii) whether the factor loadings and uniquenesses (i.e., error variances) are the same across the groups. The former is typically referred to as strong invariance whereas the latter is full invariance. The former helps one ascertain whether the same latent variable is being measured whereas the latter helps one ascertain whether one is measuring the same latent variable with the same level of accuracy (i.e., error variance). Both of these hypotheses take advantage of the ability to test nested models in CFA by differences in the chi-squared statistics – typically, with maximum likelihood estimation. In this case, the full and strong invariance hypotheses are tested against the baseline model of no constraints across the two groups. As a technical note to consider when implementing MG-CFA, as Cudeck (1989) reminds us, CFA and particularly MG-CFA should always be conducted on covariance matrices (and not correlation matrices).

The one strength that has brought MG-CFA to prominence is that a researcher is provided a formal statistical test, a chi-squared test computed as a difference from the baseline model, to investigate full and strong invariance. This formal test was meant to liberate the researcher from relying on what was often subjective criterion for construct comparability of the exploratory methods.

Given that the 66 items of the SAIP Assessment are summed to a single composite score, a one-factor MG-CFA was fit using regular Pearson covariance matrices and maximum likelihood estimation. Although the items are binary a Pearson covariance matrix was fit because the other four options were not viable: (i) maximum likelihood estimation with a tetrachoric covariance matrix is not recommended, (ii) diagonally weighted least squares with the corresponding asymptotic covariance matrix and the tetrachoric covariance matrix is limited due to the fact that no more than 25 items can be used due to the excessive[1] computer memory demands with the so-called weight matrix, i.e., asymptotic covariance matrix of the vectorized elements of the observed covariance matrix, and (iii) the Satorra-Bentler corrected chi-square in LISREL and Muthen's estimation method for binary data in the software Mplus are also limited by the large number of items that are found in large-scale educational measurement.

From Table 1, the difference in chi-square values between the baseline model and the full invariance model is statistically significant, $\Delta\chi^2 = 2554.55$, $\Delta df = 132$, $p < 0.0001$, indicating that the hypothesis of full invariance is not tenable. The difference in chi-square values between the baseline model and the strong invariance model is statistically significant, $\Delta\chi^2 = 612.12$, $\Delta df =$

---

[1] With p variables there are l elements in the same covariance matrix, and the weight matrix is of order lxl, where l=(p(p+1))/2. Therefore, as an example, for a model that has 20 items, the weight matrix would contain 22,155 distinct elements and for 25 items the weight matrix would contain 52,975 distinct elements.

66, $p < 0.0001$, indicating that the hypothesis of strong invariance is also not tenable. Therefore, neither the stricter test of full invariance nor the weaker test of strong invariance are tenable in the SAIP data suggesting, on their own, that we do not have construct comparability.

Insert Table 1 about here

Multi-Group Exploratory Factor Analysis

Given that the MG-CFA analyses resulted in the conclusion that construct equivalence was not tenable, a MG-EFA was conducted. The MG-EFA is conducted in three steps:

1. An exploratory factor analysis was conducted separately for the English and French versions of the test. In both cases, the same numbers of factors are extracted. Earlier analyses of these data (Koh, Zumbo, Ercikan, 2002) found that the first eigenvalue is nearly five times the size of the second and the remaining eigenvalues are nearly equal. This suggests that there may be a secondary minor factor. Therefore, two factors were extracted using unweighted least squares estimation[2] and varimax rotation was performed. Varimax rotation was chosen because the Procrustes rotation to maximal similarity is less subject to capitalizing on chance and spurious similarity if the Procrustes rotation is to orthogonal axes.
2. Orthogonal rotation to congruence is not available in widely used statistical packages. The Procrustean approach we are adopting, based on Cliff (1966) and Schönemann

[2] Unweighted least squares estimation is useful with binary data because, unlike maximum likelihood estimation, this estimation method is more robust to lack of normality.

(1966), allows us to rotate two factor solutions to maximal congruence[3]. As in any type of factor rotation, transformation matrices T_e and T_f, for the English and French version, respectively, are required to rotate the original factor axes to new positions. The original factors are those specified by the varimax factor matrices B_e and B_f, and the new positions are referenced by the matrices $S_e = B_e T_e$ and $S_f = B_f T_f$, respectively. Now, let $W_e = (B'_e B_f)(B'_f B_e)$ and $W_f = (B'_f B_e)(B'_e B_f)$ and the transformation matrices, T_e and T_f, are created from the eigenvectors of W_e and W_f as columns, respectively. Appendix A lists the SPSS syntax for the above computations.

3. The agreement between rotated factor loadings can be investigated by a congruency coefficient (e.g., Barrett, 1986; Kaiser, Hunka, and Bianchini, 1971; Paunonen, 1997; Wrigley & Neuhauss, 1955). There are two widely used measures of agreement between factor loadings. The first and simplest is the degree of correlation between corresponding factors. The second widely used approach to quantify agreement between factor loadings, for each factor, from two different groups is Tucker's congruence index, phi. The phi index for the *k*th-factor of English (denoted *e*) and French (denoted *f),*

$$\phi_k(e,f) = \frac{\sum_{i=1}^{p} b_{ie} b_{if}}{\sqrt{\sum_{i=1}^{p} b_{ie}^2 \sum_{i=1}^{p} b_{if}^2}},$$

[3] It is assumed in this presentation that each group has the same number of factors and that there are two or more factors. If each group is unidimensional then this type of rotation is unnecessary.

where b_{ie} and b_{if} denote the *i*th factor loading, for *i=1,2, ..., p,* for the *p* items of the English and French versions, respectively. This index measures the identity of two factors up to a positive multiplicative constant hence allowing for differences in the eigenvalues for the separate factor solutions across groups. These two similarity measures can be used to decide on the number of reliable components across the key groups.

The sampling distribution of the phi index is unknown, however, Hurley and Cattell (1962) show that if the value of the coefficient is greater than 0.90 that is evidence that the factors are similar. Although it is not used in this study due to software constraints, Chan, Ho, Leung, Chan, and Yung (1999) introduce a bootstrap procedure that allows, within sampling limits, to consider sample-to-sample variability when interpreting the congruency index.

Note, however, that both the simple correlation and the phi congruence index share the same limitation that all correlation coefficients would possess; that is, because they measure linear association it is possible for these indices to be large yet the loadings are very different in magnitude. This would mean that the magnitude of the indices would suggest construct comparability when it is in fact not true.

A test of whether spurious results are being shown can be constructed based on recent findings by Rupp and Zumbo (2002) who show, in the context of invariance of item response theory parameters, that a Pearson correlation coefficient, or in our case a large Tucker's phi, is a necessary but not sufficient condition for invariance to hold. In essence, the problem is that the correlation based methods fail to capture additive shifts in factor loadings between groups or non-linearity. To check to see if the phi or correlation is working appropriately, one can plot the linear relationship between the two groups.

The ideal case would have a line starting at the zero-zero point and increasing at a 45-degree angle. Therefore, a good indicator of whether the phi or correlation coefficients are working appropriately is if the confidence band (for individual estimates) of the regression includes the zero-zero intercept. As we can see in Figures 1 and 2, the regression confidence band on the regression intercept includes zero. The scatter plots include the sunflower symbols for multiple occurrences of data points. It should be noted that, as expected from the results above, Figure 1 has the regression line on nearly a 45-degree line from the zero-zero point (a correlation of large absolute magnitude), whereas Figure 2 does not[4]. In sum, the phi and correlation coefficient can be interpreted without concern of spurious inflation.

Insert Figures 1 and 2 about here

Turning back to the example, for demonstration purposes, Tables 2 and 3 display the original and Procrustes maximal similarity rotated factor loadings for the first 10 items. The Pearson correlation of Factor 1 for the English and French versions is .85 (Spearman r =.86) and .67 (Spearman r=.66) for Factor 2. Likewise, the Tucker's phi coefficient of congruence is .97 and .67 for Factors 1 and 2, respectively. Applying the criterion listed above for the phi coefficient, we conclude that Factor 1 is equivalent across the groups but that the second (minor) factor is not.

As a test of the robustness of our results, because of the concern that the items are binary, the above exploratory factor analysis was repeated using PRELIS's option of

[4] As a further check, because the plots depict a regression they were also constructed with the X and Y variables reversed and the conclusions are the same.

exploratory factor analysis with a tetrachoric correlation matrix. We followed the same steps as before except that the factor analysis was a maximum likelihood estimation method with the tetrachoric correlation matrix. Like before, the factor solutions from the exploratory factor analysis were rotated using Procrustes orthogonal rotation to maximal similarity. Again, the congruence coefficient, phi, was computed. The resulting phi coefficients are .98 and .23 for factors 1 and 2, respectively. These results support the conclusions found with the unweighted least squares estimation and suggest that our conclusions are not due to the fact that we have binary data and unweighted least squares estimation. Note that unlike confirmatory factor analysis, as described above, the number of variables does not limit the exploratory factor analysis of the tetrachoric correlation matrix.

Insert Tables 2 and 3 about here

As a final step in our multi-method approach, Sireci's multi-group MDS strategy was used. This is an important step in our methodology because the MDS methodology used does not make the same assumption (nor is it precisely the same statistical model) as the factor analysis models and hence one is able to investigate the robustness of the findings across modeling strategies. Turning back to our example data set, two random samples without replacement were drawn from each group (half the group in each sample) and a weighted MDS was applied. Figure 3 is the weight space for the MDS. From Figure 3 one can see that the conclusions are the same as the MG-EFA, that the language groups each have one large weight on one dimension with a minor secondary dimension. The secondary dimension does not appear to be equivalent across groups, whereas the primary dimension does appear to be equivalent.

Insert Figure 3 about here

Conclusion

In this paper we reviewed some exploratory methods for construct comparability and measurement invariance. We focused on the use of an orthogonal Procrustes rotation method and the computation of a phi congruence coefficient to complement multi-group confirmatory factor analysis. We are recommending that use of exploratory methods as a complementary method to multi-group confirmatory factor analysis in contexts where the more stringent tests of measurement invariance and construct comparability via CFA fail. In our example we found that the strict tests of construct comparability failed (i.e., the MG-CFA) suggesting that the two language versions are not comparable. However, the exploratory factor analyses (i.e., MG-EFA) found a minor secondary factor. The primary factor was shown to be equivalent across language groups but the minor secondary factor was not. Given that the items of the example data set are summed to a total composite score and that the first factor dominated over the others in terms of the variance accounted for, inferences from the total scores are comparable in the English and French versions.

A limitation of the strategy we are advocating is that it does not, yet, fully exploit the idea of exploratory factor analysis as quest for discovering underlying structure in data. That is, the

MG-EFA approach we describe above is a kind of hypothesis testing (as opposed to the more inductive approaches in some types of exploratory factor analysis) or quasi-confirmatory approach, although not in the sense of formal statistical hypothesis testing provided in MG-CFA. In essence, we are suggesting complementing the strict hypothesis testing MG-CFA with a less-strict MG-EFA when the MG-CFA shows lack of construct equivalence. In future work we will investigate the use of more inductive/exploratory approaches to discovering common underlying structure across the key comparison groups.

References

Barrett, P. (1986). Factor comparison: An examination of three methods. Personality and Individual Differences, 7, 327-340.

Byrne, B. M. (1994). Testing for the factorial validity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslach Burnout Inventory. Multivariate Behavioral Research, 29, 289-311.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin, 105, 456-466.

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. Journal of Personality and Social Psychology, 66, 93-114.

Cliff, N. (1966). Orthogonal rotation to congruence. Psychometrika, 31, 33-42.

Horn, J. L. (1967). On subjectivity in factor analysis. Educational and Psychological Measurement, 27, 811-820.

Horn, J. L., & Knapp, J. R. (1973). On the subjective character of the empirical base of Guilford's structure-of-intellect model. Psychological Bulletin, 80, 33-43.

Hurley, J., & Cattell, R. B. (1962). The Procrustes program: Producing direct rotation to test a hypothesized factor structure. Behavioral Science, 7, 258-262.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

Kaiser, H. F., Hunka, S., & Bianchini, J. C. (1971). Relating factors between studies based upon different individuals. Multivariate Behavioral Research, 6, 409-422.

Koh, K., Zumbo, B. D., & Ercikan. K. (2002). Evaluating Construct Comparability across the English and French versions of the School Achievement Indicators Program (SAIP): A Look at Translation Effects from CFA and IRT perspectives. Annual Meeting of the National Council for Measurement in Education (NCME), New Orleans, Louisiana.

Korth, B., & Tucker, L. R. (1975). The distribution of chance congruence coefficients from simulated data. Psychometrika, 40, 361-372.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika, 58, 525-543.

Paunonen, S. V. (1997). On chance and factor congruence following orthogonal Procrustes rotation. Educational and Psychological Measurement, 57, 33-59.

Robin, F., Sireci, S.G., & Hambleton, R.K. (in press). Evaluating the equivalence of different language versions of a credentialing exam. *International Journal of Testing.*

Rupp, A. A., & Zumbo, B. D. (2002). How to Quantify and Report Whether Invariance Holds for IRT Models: When Pearson Correlations Are Not Enough. Manuscript under review.

Schönemann, P. H. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31, 1-10.

Sireci, S. G., & Allaluof, A. (2003). Appraising item equivalence across multiple languages and cultures. Language Testing, 20, 148-166.

ten Berge, J. M. F. (1986). Rotation to perfect congruence and the cross validation of component weights across populations. Multivariate Behavioral Research, 21, 41-64.

Wrigley, C. S., & Neuhauss, J. O. (1955). The matching of two sets of factors. American Psychologist, 10, 418-419.

Table 1. Multi-group Confirmatory Factor Analysis Results for the Measurement Invariance

Model	**Chi-square**	**Degrees of freedom**
Baseline[1] (No between-group constraints, except number of factors and the factor pattern)	8592.14	4158
Strong Invariance[2] (Number of factors and Factor loadings invariant)	9204.26	4224
Full Invariance[3] (Number of factors, Factor loadings, and Error variances invariant)	11146.69	4290

Notes:

[1] Configural invariance

[2] Configural invariance and Metric invariance

[3] Configural invariance, Metric invariance, and item uniqueness invariance

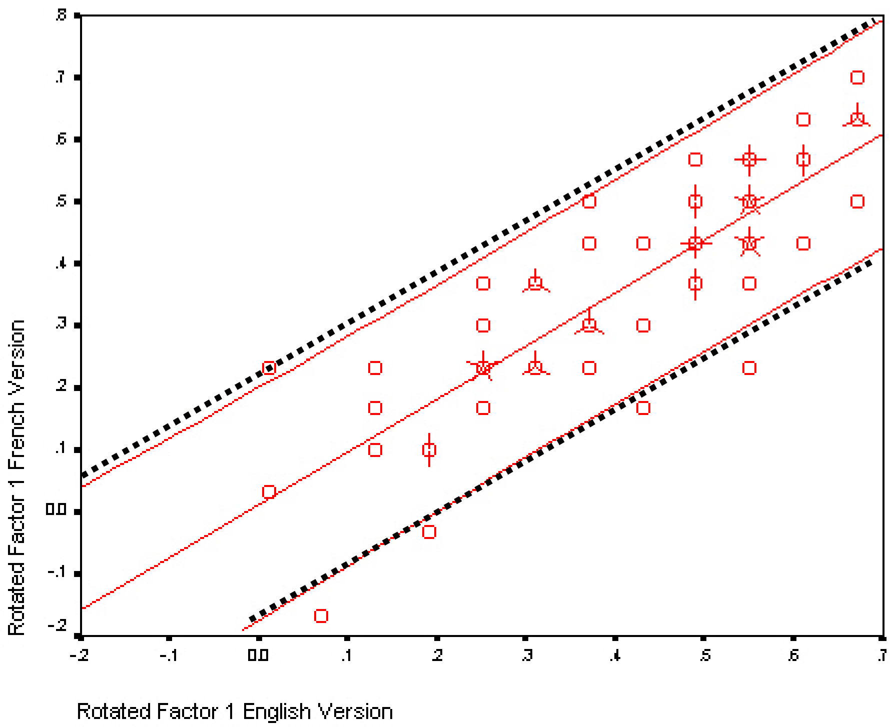
Table 2. The Varimax Factor Loadings for the First Ten Items, From the Separate Factor Analyses.

	English		**French**	
	Factor 1	Factor 2	Factor 1	Factor 2
Item1	.23	.03	.16	-.01
Item 2	.26	.13	.21	.15
Item 3	.18	.11	.10	.12
Item 4	.35	.02	.39	.08
Item 5	.12	.12	.09	.05
Item 6	.40	.11	.27	.18
Item 7	.41	.11	.33	.08
Item 8	.26	.35	.21	.34
Item 9	.31	.29	.26	.17
Item 10	.26	.28	.20	.32

Table 3. The Varimax Factor Loadings for the First Ten Items, After Procrustean Orthogonal Rotation to Maximal Similarity.

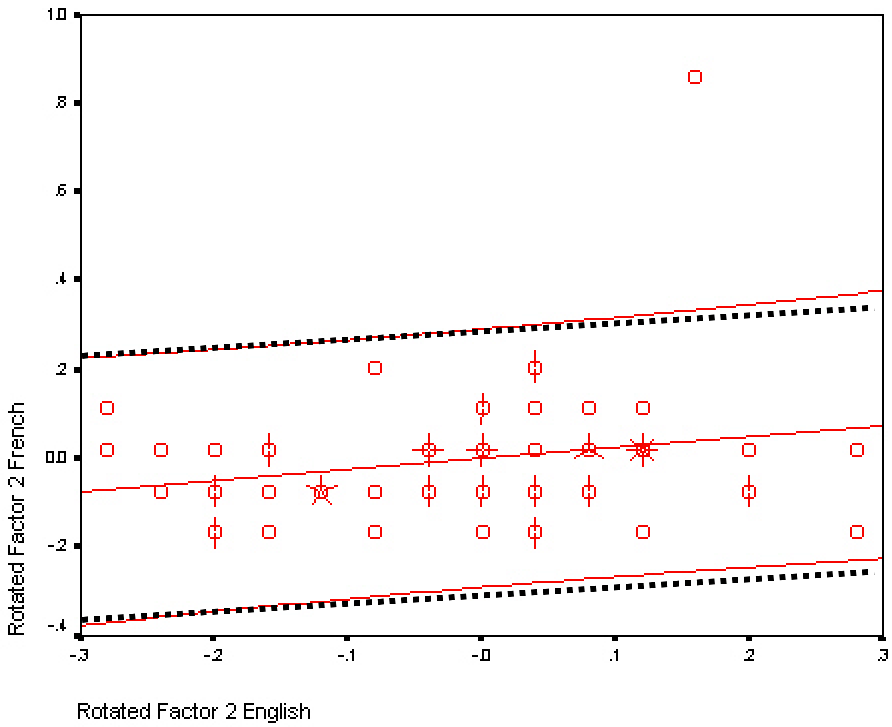
	English		French	
	Factor 1	Factor 2	Factor 1	Factor 2
Item1	.20	-.12	.12	-.10
Item 2	.28	-.05	.25	-.01
Item 3	.21	-.02	.15	.03
Item 4	.29	-.19	.36	-.17
Item 5	.17	.02	.10	-.01
Item 6	.38	-.16	.32	-.02
Item 7	.39	-.17	.31	-.13
Item 8	.42	.12	.37	.15
Item 9	.42	.04	.31	-.02
Item 10	.38	.06	.35	.13

Figure 1. The Plot of the Relationship Between the Factor Loadings of the Two Groups, a Test of Whether the Correlation or Phi are Appropriate. (Factor 1)



Note: The dotted lines represent the 95% confidence band for individual estimates.

Figure 2. The Plot of the Relationship Between the Factor Loadings of the Two Groups, a Test of Whether the Correlation or Phi are Appropriate. (Factor 2)

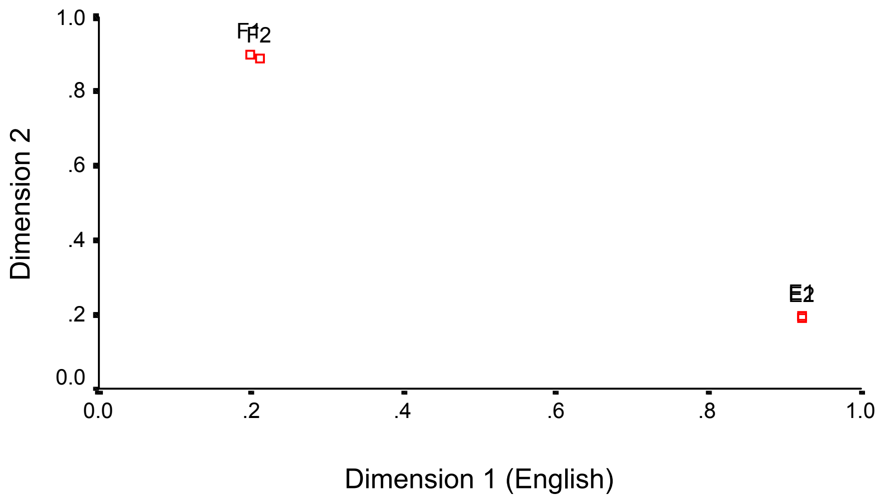


Note: The dotted lines represent the 95% confidence band for individual estimates.

Figure 3. The weight space and MDS weights from the example data set.

Figure 3

WMDS Weight Space



E1=English Sample 1, E2=English Sample 2

F1=French Sample 1, F2=French Sample 2

Subject Weights

Subject Number	Weird-ness	Dimension 1	Dimension 2
1	.7260	.9219	.1945
2	.7286	.9222	.1926
3	.7344	.1972	.8995
4	.7125	.2112	.8877
Overall importance of each dimension:		.4460	.4180

Appendix A

SPSS Syntax to Perform the Procrustes Orthogonal Rotation to Maximal Similarity

```
*.
* Orthogonal rotation to congruence.
*.
set workspace=100000.
MATRIX.
*.
*  The Varimax solution for the English.
*   Be is the factor loading vector.
*.
*  In future applications change the variable number and names for the Get commands.
*  Have the loadings SPSS file open.

GET Be /VARIABLES = f1eng f2eng.
GET Bf /VARIABLES = f1fren f2fren.
Print Be.
Print Bf.

*.
* Compute the transposes.
*.
COMPUTE BeT=T(Be).
COMPUTE BfT=T(Bf).

COMPUTE We=(BeT*Bf)*(BfT*Be).
Print We.
COMPUTE Wf=(BfT*Be)*(BeT*Bf).
Print Wf.

* CALL EIGEN(M, A, B) will assign to A a matrix having the same.
* dimensions as M, containing the eigenvectors of M as its columns, and.
* will assign to B a column vector having as many rows as M, containing.
* the eigenvalues of M in descending numerical order.
* The eigenvectors  in A are ordered to correspond with the eigenvalues in B; thus, the first.
* column corresponds to the largest eigenvalue, the second to the second.
* largest, and so on.

CALL EIGEN(We,Te,Lamdae).
CALL EIGEN(Wf,Tf,Lamdaf).
Print Te.
Print Tf.

COMPUTE Se=Be*Te.
COMPUTE Sf=Bf*Tf.

Print Se.
Print Sf.
*  In future applications change the variable number and names for the Save command.
SAVE {Se,Sf} /OUTFILE=rotated_loadings.sav
/VARIABLES = rf1eng, rf2eng, rf1fren, rf2fren.

END MATRIX.
```