Practical Assessment, Research, and Evaluation

Volume 12 *Volume 12, 2007* Article 7

2007

An Introduction To Multi-Battery Factor Analysis: Overcoming Method Artefacts

Gavin T. L. Brown

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 12, Number 7, May 2007 ISSN 1531-7714

An Introduction To Multi-Battery Factor Analysis: Overcoming Method Artefacts

Gavin T L Brown
University of Auckland, Auckland, New Zealand

Examination of participants' responses to factor or scale scores provides useful insights, but analysis of such scores from multiple measures or batteries is sometimes confounded by methodological artefacts. This paper provides a short primer into the use of multi-trait, multi-method (MTMM) correlational analysis and multi-battery factor analysis (MBFA). The principles of both procedures are outlined and a case study is provided from the author's research into 233 teachers' responses to 22 scale scores drawn from five batteries. The batteries were independently developed measures of teachers' thinking about the nature and purpose of assessment, teaching, learning, curriculum, and teacher efficacy. Detailed procedures for using Cudeck's (1982) MBFACT software are provided. Both MTMM and MBFA analyses identified an appropriate common trait across the five batteries, whereas joint factor analysis of the 22 scale scores confounded the common trait with a battery or method artefact. When researchers make use of multiple measures, they ought to take into account the impact of method artefacts when analyzing scale scores from multiple batteries. The multi-battery factor analysis procedure and MBFACT software provide a robust procedure for exploring how scales inter-relate.

Researchers in social science are encouraged to explore constructs with multiple instruments (Brewer & Hunter, 1989). In a study designed to examine how teachers' conceptions of assessment related to their conceptions of learning, teaching, curriculum, and assessment Brown (2002) administered 22 scales taken from five different inventories to 233 participants. The instruments were Brown's (2004b) *Conceptions of Assessment* inventory, and five abbreviated scales from the *Teaching Perspectives Inventory* (Pratt & Collins, 1998), two conceptions of learning scales from the *Approaches and Study Skills Inventory for Students* (ASSIST) (Tait, Entwistle, & McCune, 1998), four scales from the *Curriculum Orientations Inventory* (Cheung, 2000), and two scales from Guskey and Passaro's (1994) *Teacher Efficacy Scale*. The constructs measured by each inventory were largely distinct from each other. The assessment conceptions were (1) assessment is bad, (2) assessment is ignored, (3) assessment is inaccurate, (4) assessment is valid, (5) assessment describes performance, (6) assessment improves teaching, (7) assessment improves learning, (8) assessment makes schools accountable, and (9) assessment makes students accountable. The teaching perspectives were (1) teaching is nurturing children, (2) teaching involves students in apprenticeship learning, (3) teaching is transmission of important learning, (4) teaching aims to enact social reform, and (5) teaching is about cognitive development. The learning conceptions were (1) transformative (i.e., making meaning by changing how material is formed) and (2) reproductive (i.e., replicating material so that it is reproduced in the same form). The curriculum orientations were (1) content as a means of social reconstruction, (2) content as academic traditions, (3) content as technological systematic specification of objectives and processes, and (4) content as a humanistic concern for the individual learner. Finally, the efficacy conceptions were (1) a focus on external obstacles to successful teaching and (2) a focus on the internal resources of the teacher to accomplish successful teaching.

When the questionnaire was prepared, the items from each inventory were randomly ordered with only the items from the same inventory. In other words, the participants were asked to consider their conceptions of assessment, then their teaching perspectives, followed by their curriculum orientations, their learning conceptions, and their efficacy. This procedure was followed for two reasons. First, the items had their meaning in part from the other items related to the same domain, much as items in a test analyzed using classical test theory only have their meaning if they are kept with the other items of the test. Second, by presenting all the items of a domain together it was possible to reduce cognitive demand and fatigue due to changing field of reference. When considering items related to the one construct or domain, it is easier to respond accurately and honestly than when large leaps are needed to consider each subsequent item. By grouping according to domain (and thus also by method), the integrity of the item meanings is upheld and cognitive demand is reduced. Nevertheless, this does mean that the 22 scale traits are presented within the framework of five different methods.

Each of these five inventories had been independently developed to measure conceptually and theoretically well thought out domains. Additionally, each inventory had acceptable or robust psychometric properties for its constituent scales so there was good reason for thinking that the scale scores would provide valid about teacher thinking. Brown (2002) conducted confirmatory factor analysis with the items and scales for each inventory separately to confirm the existence of the scales. Each of the proposed 22 scales was found in the five inventories in the sample of New Zealand primary school teachers. It is worth noting that of the 22 scale scores derived from the five self-report inventories; only one common trait existed across two of the batteries (i.e., *Curriculum Orientation* social reconstruction and *Teaching Perspective* social reform). Thus, within five domains, measurements were found for 21 different traits by making use of five different inventories. Furthermore, with only 233 participants, there were too few cases with which to conduct exploratory factor analysis of all the items underlying the 22 scales; the ratio of cases to variables was much less than the recommended 20 to 1 (Osborne & Costello, 2005).

Other than the methodological problem of sample size, there exists a rather more substantive reason for using scale scores rather than creating new factors every time a new mix of instruments is used. Most inventories, as in the case of the inventories used here, have well established theoretical and empirical bases and their resulting factor or scale scores permit meaningful interpretations of the responses. Indeed, it is scale or factor score that matters in interpreting participant responses, not item level responses. For example, Marsh, Hau, Artelt, Baumert, and Peschar (2006) conducted a large-scale analysis of student approaches to learning by developing measurement models that made use of the contributing inventories' factor or scale scores rather than by reanalyzing all items simultaneously into integrated factors. Strauman and Wetzler (1992), in their analysis of two self-report measures of psychopathology, analyzed the scale level factors. Thus, it seems appropriate, when the goal is to see how scales relate to each other to analyze factor scale scores as if they were observed variables rather than conduct a reinterpretation of the underlying items.

There were good grounds for using the scale scores to report the strength of teachers' thinking within each domain represented by the five different inventories. That is the ratio of cases to variables was low and the factor scales had important meanings that had been previously and independently established. Furthermore, the objective of the research was to discover how the five domains inter-related to each other in the thinking of the teacher participants. Examination of the five sets of independent scale scores was insufficient. Also there was no pre-existing theory of how the various 22 scale scores should inter-relate, as similar research had not been conducted. Thus, the 22 scale scores had to be treated as if they were observed variables rather than as latent factors, so that the resulting scales were analyzed using standard exploratory maximum-likelihood estimation, oblique rotation factor analysis procedures. This approach to common factor analysis is usually called joint factor analysis, since it involves treating all variables jointly regardless of their origin.

Results of this analysis were largely meaningful, in that three factors could be easily identified and interpreted (Table 1). However, one factor (Factor IV) was difficult to interpret, in that it involved the two similar trait variables (i.e., the social reform perspective of teaching and the social reconstruction orientation to curriculum) and two other perspectives taken from the *Teaching Perspectives Inventory* (i.e., developmental perspective and transmission perspective). Either

DOI: https://doi.org/10.7275/ag25-0r66

Table 1. *Joint EFA Results for Conceptions of Teaching, Learning, Curriculum, Teacher Efficacy, and Assessment.*

Scales	Joint Factors I	II	III	IV
18. Assessment Student Accountability	**.66**	.35	-.04	-.08
14. Assessment Describe	**.63**	-.44	-.15	.04
13. Assessment Valid	**.56**	-.41	.17	-.14
17. Assessment School Accountability	**.56**	-.13	.09	-.26
20. Curriculum Academic	**.47**	.05	-.20	-.24
7. Learning Reproductive	**.45**	.09	-.12	-.10
21. Curriculum Technological	**.42**	-.15	-.31	-.01
9. Efficacy Internal	**.40**	.07	-.06	-.21
10. Assessment Bad	.13	**.79**	-.02	.01
11. Assessment Ignore	-.03	**.72**	-.02	-.09
16. Assessment Improve Learning	.39	**-.60**	-.13	-.09
15. Assessment Improve Teaching	.38	**-.53**	-.30	.08
12. Assessment Inaccurate	-.11	**.40**	-.31	-.09
8. Efficacy External	.20	**.36**	.13	.04
1. Teaching Nurturing	-.10	-.07	**-.67**	-.20
6. Learning Transformative	.02	-.05	**-.64**	-.10
22. Curriculum Humanistic	.24	.05	**-.51**	.16
2. Teaching Apprenticeship	.09	-.10	**-.39**	-.35
4. Teaching Social Reform	-.04	.03	-.02	**-.78**
5. Teaching Development	-.06	-.11	-.29	**-.67**
19. Curriculum Social Reconstruction	.20	.11	.09	**-.55**
3. Teaching Transmission	.36	.07	.09	**-.53**
Inter-factor Correlations				
I	1.00			
II	-.12	1.00		
III	-.20	.12	1.00	
IV	-.36	-.04	.28	1.00

Notes. The strongest loadings are shown in bold. Joint EFA conducted with maximum likelihood estimation and direct oblimin rotation.

this factor meant that teachers conceived of social transformation as involving transmission and developmental teaching approaches or else the shared origins of the three teaching perspectives factors was overwhelming and contaminating the shared trait of social reform. Unfortunately, joint factor analysis was unable to discriminate between the shared method (the Teaching Perspectives questionnaire) and shared traits (social reform or change) interpretations of the factor.

Fortunately, problems such as this have been studied before and procedures have been developed to disentangle method effects from common traits. Building on the logic of multi-trait, multi-method correlational analyses, multi-battery factor analysis was developed in the early 1980s. The point of this paper is to explain the logic underlying multi-battery factor analysis, give a fully worked example of multi-battery factor analysis, and show how the introductory problem was resolved with this procedure. This paper is a tutorial in the procedure and an explanation of how and when the procedure has value.

DISENTANGLING TRAITS AND METHODS

Multi-trait, multi-method analysis (MTMM) (Campbell & Fiske, 1959) was developed to address the issue of how to separate method effects from trait effects. The argument Campbell and Fiske made was that a trait should be congruent across methods, if the trait is being independently but commonly measured by multiple methods, inventories, or data sources. In other words, using Brown's dataset already introduced, if the trait of social reform existed independently of the type of measure, the common trait scores should be more like each other than they were like any other traits with which they shared a common method. If the traits did not exist, then they would be more like their method partners than their cross-method cousins. MTMM uses within method and between trait correlations to determine whether trait or method facets explain observed relationships. In MTMM correlational analysis, a common trait across method (i.e., monotrait, heteromethod) is accepted if the correlations between similar traits are greater than the correlations of different traits within the same method (heterotrait, monomethod) or between methods (heterotrait, heteromethod). However, it is up to the judgment of the researcher to determine whether the congruent correlations are greater than the within method or across trait correlations.

The logic of MTMM has been extended to factor analysis of multiple factors taken from multiple measurement instruments or batteries. In the case of two batteries, there are four sub-matrices of correlations—denoted R_{11}, R_{12}, R_{21}, and R_{22}. The sub-matrices R_{11} and R_{22} constitute the within-battery, multiple trait spaces, while the sub-matrices R_{12} and R_{21} are the between-battery, multiple-trait spaces. The number of such sub-matrices increases if there are more batteries. Figure 1 shows the nature of the within and between-battery spaces when two inventories are used.

Figure 1 *Supermatrix of Multiple Methods and Multiple Traits Correlations*

		Method 1 1 2 3	Method 2 1 2 3
R=	Method 1 Trait 1 Trait 2 Trait 3	$\mathbf{R}_{11}$ Within-Battery Information	$\mathbf{R}_{12}$ (Transpose of R_{21}) Between-Battery Information
	Method 2 Trait 1 Trait 2 Trait 3	$\mathbf{R}_{21}$ Between-Battery Information	$\mathbf{R}_{22}$ Within-Battery Information

When scale scores from two or more batteries are analyzed in standard, joint factor analysis, the between-battery space is ignored—the analysis only makes use of R_{11} and R_{22} sub-matrices (Cudeck, 1982). Researchers would accept that a common factor between batteries had been found if conceptually related scales from the two or more batteries loaded on one common factor, while conceptually opposing scales loaded on a different factor (Finch & West, 1997). With this approach, important information about how scales covary across batteries is ignored. Thus, it is possible that the method artefact will obscure some common trait that is being measured by two different inventories.

Alternatively, it is possible to use canonical correlations to identify common aspects of two or more source inventories. Canonical correlation reports the correlation between two orthogonally-related, reduced-rank component spaces (i.e., two linear composites designed to simplify only R_{11} and R_{22} sub-matrices respectively). Since each composite uses only within-battery information, the procedure maintains the boundaries between method artefacts, defeating the purpose of knowing how traits inter-relate across methods (Cudeck, 1982). Further, canonical correlation treats the composite variables (i.e., the traits within the battery) as error-less manifest variables, whereas factor analysis assumes that each trait contains true score and error information (Huba, Newcomb, & Bentler, 1981).

Multi-battery, or inter-battery when there are only two methods, factor analysis (MBFA) takes account of method factors (Cudeck, 1982; Tucker, 1958) when constructing interpretations of scale scores from multiple sources. Tucker's (1958) initial solution to this problem was to examine only the submatrix of the between-battery traits (i.e., R_{12} sub-matrix), and ignore the within-battery submatrices. Browne (1979), however, applied maximum-likelihood estimation to the problem of multiple scores from multiple sources to examine the variance-covariance matrix among battery specific factors plus residuals; that is making use of information in all within and between battery sub-matrices. The analysis seeks to identify the inter-battery factors that account for the between-battery covariances. "In practical terms, this means that the method is principally designed to explore similarities between the batteries, and de-emphasizes unique elements of either set" (Cudeck, 1982, p. 54). This procedure is most appropriate when measurement models for each battery "are well-defined… [and]…the purpose of the study was to explore common but unknown aspects of the behaviors assessed by the different methods" (Cudeck, 1982, p. 63).

The decision to move to multi-battery factor analysis should be driven by analytic intent. "If the analyst wishes to determine the common latent variable sources of variance for a set of variables that is grouped into two domains, the interbattery model is probably the more appropriate one. If, on the other hand, the investigator wishes to choose a small number of linear combinations of the original variables in each of the two sets in such a way as to maximize the correlations between the domains, then the canonical correlation model is probably more appropriate" (Huba, Newcomb, & Bentler, 1981, p. 295).

When the number of batteries or methods is three or more an iterative procedure is used until the maximum-likelihood estimate is achieved for multi-battery factors that are assumed to be uncorrelated. It is normally expected that the number of multi-battery factors will not exceed the lowest number of traits or factors supplied by one of the batteries. However, if the communalities do not exceed unity, the number of multi-battery factors can exceed this small value (Cudeck, 1982). In terms of Brown's (2006) data, since there were only two factors for conceptions of learning, there might be no more than two multi-battery factors. However, the goodness-of-fit for MBFA solutions can be evaluated with a number of fit indices; for example, the Tucker-Lewis index (TLI) with values >.95 indicating good fit (Cudeck, 1982); the model with the lowest Akaike Information Criterion (AIC) value being preferred; a model with a likelihood test ratio that exceeds the χ^2 critical value for the degrees of freedom being rejected (Browne, 1980); and a model with a high average off-diagonal residual also being rejected. So it is possible to find a well-fitting solution that exceeds the number of factors in the battery with the fewest factors.

Cudeck[1] (1982; 1991) developed a software application (MBFACT) which he is willing to make available to interested researchers. MBFACT has been successfully used in studies using both joint and inter-battery factor analysis (e.g., Brown, 2006; Finch, Panter, & Caskie, 1999; Meiring, Van De Vijver, Rothmann, & Sackett, 2006; Ransom, Fisher, & Terry, 1992). This paper will demonstrate with Brown's (2006) dataset introduced earlier how MBFACT can be used to address the method-trait problem. To demonstrate the congruence between MBFA and MTMM, first the MTMM results are reported and described before the detailed MBFA results and procedures.

THE MBFACT SOFTWARE

The MBFACT software operates within a Windows command space and requires the user to specify the number of batteries, the number of scales per battery, the number of factors to be tested, and the type of rotation method preferred, and to provision of the covariance matrix for the scales in the order listed. The output, in simple text format, reports factor structure, factor pattern, and factor correlation matrices. Additionally, the likelihood ratio test and the Tucker-Lewis Index are provided as goodness-of-fit measures.

MBFACT comes as a small (500Kb) executable file (mbf.exe), a MS Word document help file, and example input and output files. The help file explains how the input file is to be structured and what all the parameter values are. The input and output files contain the information used and reported in Cudeck (1982).

MBFACT Input File Structure.

I recommend creating this file in Microsoft Notepad or any other similar fixed format word processor as fixed field options are easy to apply. Optional title line(s) commence the MBFACT input file and each title line must begin with a NON-numeric character in column 1, otherwise the application will consider it to be an executable command. I find it useful to insert here comments such as the name of the study, the names of the batteries in the order they will be inserted, and other informative details so that I can recall what study the analysis applies to.

The first command section is ONE line that is a series of at least 16 numbers (for two batteries, increase by one for each additional battery) each of which is separated by a space. The nature of each number and its options are listed in order:

1st. The number of observations or cases;
2nd. The number of batteries;
3rd. The value for type of final transformation to be applied (0=Varimax—orthogonal; 1=Direct Quartimin—oblique; 2= both);
4th. The minimum number of factors to examine (usually 1);
5th. The maximum number of factors to examine (normally not more than the lowest

[1] Professor Robert Cudeck, Ohio State University can be contacted by email at cudeck.1@osu.edu

number of factors in any battery—although MBFACT provides a warning, it is possible to specify a maximum that may be theoretically valid and supported by the number of cases available);

6th. The value to indicate the format of variable or factor names being analyzed (0=no names; 1=names used and presented in Fortran style of 20 columns per name, 1 name per line; 2=names used and presented as 4 names per line and each name is exactly 20 columns wide);

7th. The value for variable selection option (0=all variables analyzed; 1=a subset of variables are to be analyzed—though normally the researcher would want to analyze all the variables from all the batteries and so this value should have a default of 0);

8th. The value to indicate the convergence criterion for communalities to control the number of iterations implemented (1=epsilon of .001; 2=epsilon of .0001; 3=stop after 10 iterations regardless of convergence—clearly the most robust procedure is 2 and modern computers should have no difficulty in calculating the required number of iterations in an acceptable period of time. If convergence cannot be obtained, then options 1 and 3 can be used);

9th. The value to indicate whether to include the original covariance matrix in the output file (0=no; 1=yes);

10th. The value to indicate whether to include the covariance matrix after any selection of variables in the output file (0=no; 1=yes);

11th. The value to indicate whether to include the orthogonalized covariance matrix in the output file (0=no; 1=yes);

12th. The value to indicate whether to include the Eigen values of the orthogonalized covariance matrix in the output file (0=no; 1=yes);

13th. The value to indicate whether to include the unrotated solution in the output file (0=no; 1=yes);

14th. The value to indicate whether to include the residual matrix in the output file (0=no; 1=yes);

15th. The number of variables or scales in the first battery;

16th. The number of variables or scales in the second battery; and

17th. Continue for as many batteries were specified in the second number. This row then concludes with a hard return (no semi-colon or full-stop).

The second command section consists of the variable names if option 6 was set as either 1 or 2. I normally use option 2 as this allows 20 characters by which to name each variable four scale or variable names are put in each row which ends in a hard return (not semi-colon or full-stop). Truncate variable names to end at 20 characters and pad variable names with blank spaces up to 20 characters if the variable name is shorter than 20 characters. I recommend switching on variable names as they are included in the output and assist in interpreting output.

The third command section is the covariance matrix for the factors in the order specified in the variable name list. Each row starts a new variable each of which is read row-wise. Only the lower triangular section of the variable covariance matrix is used. Thus, the first row consists only of the value 1. (**NB**: no trailing zero required, but decimal point is), since the covariance of the variable to itself is set at unity, even if the covariance matrix reports a different value. The covariance values are recorded at three decimal places in the format .136 and -.023 for positive and negative values respectively.

If the variable selection value of 1 was selected then the fourth command section specifies the structure of the selection process. In separate command lines, insert the number of batteries after selection is finished, the number of variables or factors in the first battery, the number of factors in the second battery, and so on. Then, insert the number of the first variable, the second one, and so on, using as many lines as needed to indicate all the selected variables.

The command sequence ends with a blank line, so do not leave blank lines between the various command sections, otherwise MBFACT will terminate and report as far as it got in the output file. If you intend to analyze more than one problem in one command file, do not insert a blank line; rather insert a new title section followed by all the appropriate command sections for the second analysis. The whole job ends with a blank line; otherwise MBFACT will fail to report output. Figure 2 shows the input file for Brown's (2006) data problem consisting of five batteries, 22 factors, and 233 cases.

MBFACT Command Sequence.

The first thing to note is that the MBFACT executable must be in the same folder as the input file, as there is no capacity to include folder or directory sequences. Once launched, MBFACT requests the name of the input file, which must be in 8.3 format (i.e., no more than eight character name and 3 character file type suffix). The default file name is MBF.DAT but any file name including those ending .txt can be read in. Once the file name is entered, press enter and MBFACT will request the name of the output file which must be in the same format as the input file. Note this file will be saved in the same folder as the input file. Once entered is pressed, MBFACT will calculate, displaying as scrolling text the output, which given contemporary computers scrolls past too quickly to be read. You will need to open the output file to see if the analysis has been run correctly. Should any of the instructions not be correctly formatted in the input file, there will be no error messages, but the output file will provide a running record of where it had got up to in the command sequence before it terminated. It is hoped that the verbose description of the input file given above will assist users in trouble-shooting any output faults.

Figure 2 *MBFACT Input Commands for Brown (2006) data set*

instructin.txt - Notepad

File Edit Format View Help

```
MBfact file teacher instructional conceptions
order = teaching, learning, efficacy, assessment, curriculum
batteries = 5
factors = 22
command begins
233 5 2 1 5 2 0 2 1 1 1 1 1 1 5 2 2 9 4
tpinurtur       tpiappren       tpitrans        tpireform
tpidevelop      learndeep       learnsurf       teffexternal
teffinternal    assessbad       assessignore    assessinaccurate
assessvalid     assessdescribe  assessimptchg   assessimplrng
assessschlacc   assessstudacc   currsocreform   curracademic
currtechnology  currhumanistic
1.

.143 1.

.074 .162 1.

.158 .237 .347 1.

.19 .242 .282 .429 1.

.179 .155 .067 .142 .181 1.

.078 .127 .294 .193 .104 .162 1.

-.09 -.012 .059 .026 -.057 -.044 .074 1.

.051 .133 .176 .171 .087 .077 .147 .024 1.

-.031 -.038 .06 .068 -.034 -.029 .013 .211 .036 1.

-.023 -.055 .04 .132 -.005 -.02 -.022 .186 .037 .389 1.
.067 .069 -.031 .147 .072 .073 .088 .074 .038 .19 .211 1.
.037 .071 .212 .165 .121 .05 .201 -0.053 .138 -.169 -.213 -.221 1.
.092 .14 .168 .108 .145 .13 .197 -.067 .141 -.187 -.205 -.152 .358 1.
.097 .15 .075 .073 .128 .149 .125 -.112 .091 -.221 -.273 -.071 .246 .358 1.
.100 .168 .117 .13 .156 .109 .124 -.117 .115 -.24 -.277 -.118 .341 .367 .359 1.
.062 .179 .266 .274 .177 .041 .274 .022 .201 -.055 -.064 -.029 .311 .313 .216 .278 1.
.035 .112 .271 .166 .141 .066 .273 .130 .135 .184 .114 .058 .180 .176 .067 .070 .268 1.
.065 .142 .271 .454 .197 .029 .162 .053 .167 .091 .052 .144 .149 .071 .044 .125 .249 .195 1.
.122 .198 .281 .272 .253 .140 .291 -.003 .227 .014 .004 .188 .250 .276 .216 .195 .341 .262 .35 1.
.115 .187 .158 .121 .187 .098 .187 -.014 .087 -.085 -.132 .031 .206 .276 .232 .217 .219 .187 .088 .272 1.
.194 .12 .038 .052 .075 .126 .107 -.031 .073 .017 -.012 .078 .087 .134 .158 .153 .075 .113 .089 .203 .239 1.
```

MBFACT Output File Structure.

The MBFACT output file repeats the initial title information from the input file and then reports in full verbose form the parameter entries in the first command line with a brief explanation as to what the value meant. This is useful for validating that the analysis has been run in the way requested and on the variable and battery structure intended. The input covariance matrix is repeated with variable names inserted as further validation that the variables have been analyzed in the order expected. Then the various requested matrices are reported for inspection. These are the values that are used in the subsequent MBFA.

The researcher specified the minimum and maximum number of factors in the input file; MBFACT reports the results for each analysis. If the researcher specified a maximum greater than the lowest number of variables in a battery, a warning will be printed stating that "KL is greater than the number of variables in the smallest battery". The researcher is reminded that although normally not expected, an inspection of the fit statistics and the theory underlying the analysis may mean that this warning can be safely ignored. In each analysis, the number of factors is specified and for every third iteration (i.e., 1st, 4th, 7th, etc.) the fit functions, largest change in communality, and largest communality are reported. Also various fit statistics for the solution are reported, including the test statistic, degrees of freedom, upper tail probability, the Tucker-Lewis Reliability Index, the Rescaled Akaike Information Criterion, and the Rescaled Akaike for the Saturated Model. Depending on print selections, MBFACT will also report the unrotated and rotated solutions, the residual covariance matrix, the average absolute off-diagonal residual, and the factor inter-correlations. Note, to ease interpretation, MBFACT inserts a blank line after each battery when reporting the factor solution and so if the factor names do not align with the battery breaks, the researcher knows that something was wrong with the input file. With these values, the researcher can determine which solution has the most valid structure and defensible fit statistics.

Determining the number of factors is a complex judgmental process, but the solutions reported by MBFACT can be evaluated using the same criteria used in normal factor analysis. For example, if a factor has fewer than three variables or if the variables have cross-loadings greater than .30 (Osborne & Costello, 2005), the researcher may wish to consider alternative solutions. A solution that has TLI values between .90 and .95 may be a valid solution, if on balance, the variables load appropriately on theoretically meaningful factors.

Alternative Applications for MBFA

Two alternative software applications or macros have been identified as alternatives to Cudeck's MBFACT. O'Connor (2002) reported a multi-battery factor analysis of the correlations between the factors of two personality inventories using Tucker's (1958) procedure. This is a defensible procedure when only the correlation matrices are available as in the case of a re-analysis of published studies. Dr O'Connor[2] has written MatLab and SPSS syntax files for this procedure and which he is willing to make available to interested readers. Because I have not used these tools, I do not report them in this paper, but interested readers will find that the syntax files are well commented.

Additionally, Huba, Palisoc, and Bentler (1982, p. 62) reported an inter-battery software, ORSIM2, that used "two matrices of canonical correlation loadings or weights and a vector of the canonical correlations" to identify inter-battery factors. It should be noted that the application makes use of canonical correlations and returns only orthogonally related inter-battery factors, the limitations of which have both been mentioned previously. Unfortunately, the application is no longer available, though it could be resurrected by consulting Bentler and Huba (1982) and Bentler (1977). Nevertheless, for the researcher who has access to the factor covariance matrix for a set of batteries the MBFACT software is a superior approach to identifying multi-battery factors.

DEMONSTRATING METHOD TRAIT DISENTANGLEMENT

Having described the how MBFACT operates, it is worthwhile to examine the multi-trait, multi-method correlation analysis and the multi-battery factor analysis results for the initial dataset problem. Five batteries were administered in one questionnaire, with each battery clearly demarcated from the other instruments so that participants could clearly identify the change of focus. Specific instructions were given with each inventory to assist in orienting the participant to the content of the inventory. To assist with questionnaire completion, each battery used the same self-report rating scale. Instead of using the traditional, five-point, balanced Likert scale, a six point positively-packed (Lam & Klockars, 1982) agreement scale was used. Positive-packing means there were more positive values than negative values; that is there were two negative responses (strongly disagree, mostly disagree) and four positive responses (slightly agree, moderately agree, mostly agree, and strongly agree). This type of rating scale has been found effective at generating variance in self-report data in contexts where participants are inclined to agree with all statements (Brown, 2004a).

MTMM

The scale reliabilities are reported in brackets on the diagonal; only one scale had an internal estimate of reliability (standardized alpha) of less .50 (Tables 2a and 2b). Thus, the scale reliabilities were acceptable to good especially considering that only two or three items made up six of the scales. The within-battery, multiple-trait sub-matrices (R_{11}, R_{22}, R_{33}, R_{44}, and R_{55}) are marked in bold, while the between-battery matrices are in plain font. Generally, the mono-method, hetero-trait correlations (i.e., the within-battery sub-matrices) between the various scale scores of each battery were higher (average r=.28) than the hetero-method, hetero-trait correlations (average r=.19). In contrast, the correlation between the two common trait scales was .53, suggesting strongly that there existed a mono-trait across inventories.

[2] Dr Brian O'Connor, Lakehead University, Canada, can be contacted by email at boconnor@lakeheadu.ca

Table 2a. *MTMM Analysis of Instructional Conceptions Scale Correlations and Reliabilities*

	Teaching					Learning		Efficacy	
Scales	1	2	3	4	5	6	7	8	9
Teaching									
1. Nurturing	*(.67)*								
2. Apprenticeship	**.39**	(.50)							
3. Transmission	**.17**	**.32**	*(.37)*						
4. Social Reform	**.28**	**.36**	**.44**	*(.78)*					
5. Development	**.46**	**.50**	**.49**	**.57**	*(.64)*				
Learning									
6. Transformative	.54	.40	.15	.23	.41	*(.61)*			
7. Reproductive	.16	.22	.43	.21	.16	**.31**	*(.58)*		
Efficacy									
8. External	-.18	-.02	.09	.03	-.09	-.08	.09	*(.65)*	
9. Internal	.15	.33	.37	.27	.19	.21	.27	**.04**	*(.65)*
Assessment									
10. Bad	-.07	-.08	.10	.09	-.06	-.06	.02	.31	.08
11. Ignore	-.05	-.10	.06	.16	-.01	-.04	-.03	.25	.07
12. Inaccurate	.13	.12	-.04	.16	.11	.14	.11	.09	.07
13. Valid	.08	.14	.34	.20	.20	.11	.28	-.07	.28
14. Describe	.22	.29	.39	.14	.26	.29	.30	-.10	.31
15. Improve Teaching	.24	.31	.13	.10	.23	.34	.19	-.17	.2
16. Improve Learning	.24	.35	.20	.17	.28	.25	.19	-.18	.25
17. School Accountability	.13	.32	.40	.31	.28	.08	.36	.03	.38
18. Student Accountability	.08	.23	.46	.22	.25	.15	.41	.19	.29
Curriculum									
19. Social Reconstruction	.14	.26	.42	.53	.32	.06	.22	.07	.32
20. Academic	.24	.34	.41	.30	.38	.26	.37	-.00	.41
21. Technological	.27	.37	.27	.16	.33	.21	.27	-.02	.18
22. Humanistic	.42	.23	.06	.06	.12	.26	.15	-.04	.15

Note. Scale alpha reliabilities in brackets. Within-battery correlations in bold.

Table 2b. *MTMM Analysis of Instructional Conceptions Scale Correlations and Reliabilities*

	Assessment									Curriculum			
Scales	10	11	12	13	14	15	16	17	18	19	20	21	22
Assessment													
10. Bad	*(.68)*												
11. Ignore	**.61**	*(.78)*											
12. Inaccurate	**.28**	**.28**	*(.63)*										
13. Valid	**-.28**	**-.32**	**-.31**	*(.73)*									
14. Describe	**-.33**	**-.33**	**-.22**	**.60**	*(.78)*								
15. Improve Teaching	**-.39**	**-.45**	**-.11**	**.42**	**.65**	*(.68)*							
16. Improve Learning	**-.42**	**-.45**	**-.18**	**.57**	**.66**	**.66**	*(.79)*						
17. School Accountability	**-.08**	**-.09**	**-.04**	**.45**	**.49**	**.34**	**.44**	*(.81)*					
18. Student Accountability	**.32**	**.18**	**.09**	**.30**	**.31**	**.12**	**.13**	**.41**	*(.75)*				
Curriculum													
19. Social Reconstruction	.14	.08	.19	.22	.11	.07	.20	.35	.31	*(.85)*			
20. Academic	.02	.01	.23	.35	.41	.33	.29	.44	.39	**.47**	*(.65)*		
21. Technological	.14	-.21	.05	.34	.48	.41	.38	.33	.32	**.14**	**.40**	*(.67)*	
22. Humanistic	.03	-.02	.11	.13	.22	.26	.25	.17	.18	**.13**	**.28**	**.38**	*(.66)*

Note. Scale alpha reliabilities in brackets. Within-battery correlations in bold.

Clearly, the common trait of social change is identified through MTMM and the effect of common methods is also seen, so it should not surprise us that common factor analysis would confound the trait and method artefacts as seen in Table 1.

MBFA

Like the earlier joint factor analysis, maximum likelihood estimation and oblique (i.e., direct quartimin) rotation were used to test models containing one to five factors. Given the degrees of freedom, none of the models would be rejected by the likelihood test statistic (Table 3). Only the four and five factor solutions exceeded .90 for the TLI. The model with the lowest AIC was the two factor solution, while the four factor solution had a smaller AIC than the five factor solution. Three models had average off-diagonal residuals less than .05, and the four and five factor solutions both fell between .02 and .03. Based on fit statistics the four and five factor solutions had the best overall profiles. Inspection of the variable loadings on the factors indicated that the four factor solution was more interpretable since only one variable loaded on the fifth factor in the five factor solution. Thus, like the joint factor solution, four factors were found.

Table 3 *MBFA Model Fit Statistics for Five Instructional Conceptions Batteries*

			Model Fit Statistics		
# of Factors	*df*	TLI	Likelihood Test Statistic	Rescaled Akaike Information Criterion	Average Absolute Off-Diagonal Residual
1	155	.64	129.63	1.404	.065
2	134	.79	73.89	1.344	.051
3	114	.88	43.63	1.386	.044
4	95	.92	27.98	1.483	.027
5	77	.95	19.47	1.601	.024

Note. Rescaled Akaike for Saturated Model = 2.181

In the MBFA four factor solution, three of the factors had the same scales or variables loading on them as the common factor analysis (Table 4). After adjusting for sign, the absolute difference in factor loading for the 20 scales which did not change factor was an average of .13 (*SD*=.08). As expected the two procedures had very similar results when factor structures did not confound trait and method.

However, the fourth factor which contained the common trait of social change showed noticeably different results. In the joint analysis condition, the two common traits attracted two other scales from the *Teaching Perspectives Inventory*, thus confounding trait and method. In contrast, the multi-battery factor analysis isolated the two common traits across the two batteries into a single, uncontaminated factor related to social change. The two teaching perspectives scales were, in effect, freed from the tyranny of method to load on two different factors. The resulting factor pattern is interpreted much more readily in the multi-battery condition. The common trait of social change is identified across the two batteries, without cross-contamination from scales which share method, as would be predicted by the MTMM analysis.

Table 4 *Joint & Multi-battery EFA Results for Conceptions of Teaching, Learning, Curriculum, Teacher Efficacy, and Assessment.*

	Joint Factor Analysis				Multi-battery Factor Analysis			
Scales	I	II	III	IV	I	II	III	IV
18. Assessment Student Accountability	**.66**	.35	-.04	-.08	.19	**.50**	.01	.02
14. Assessment Describe	**.63**	-.44	-.15	.04	-.32	**.35**	-.04	.17
13. Assessment Valid	**.56**	-.41	.17	-.14	-.31	**.35**	.10	-.03
17. Assessment School Accountability	**.56**	-.13	.09	-.26	-.13	**.43**	.20	.00
20. Curriculum Academic	**.47**	.05	-.20	-.24	.04	**.48**	.08	.28
7. Learning Reproductive	**.45**	.09	-.12	-.10	-.01	**.50**	.04	-.00
21. Curriculum Technological	**.42**	-.15	-.31	-.01	-.11	**.35**	-.07	.29
9. Efficacy Internal	**.40**	.07	-.06	-.21	.02	**.24**	.13	.05
10. Assessment Bad	.13	**.79**	-.02	.01	**.77**	.11	.00	.01
11. Assessment Ignore	-.03	**.72**	-.02	-.09	**.83**	.04	.08	.28
16. Assessment Improve Learning	.39	**-.60**	-.13	-.09	**-.43**	.14	.11	.17
15. Assessment Improve Teaching	.38	**-.53**	-.30	.08	**-.34**	.18	-.03	.17
12. Assessment Inaccurate	-.11	**.40**	-.31	-.09	**.49**	.04	.08	.28
8. Efficacy External	.20	**.36**	.13	.04	**.23**	.17	-.00	-.17
1. Teaching Nurturing	-.10	-.07	**-.67**	-.20	.00	-.07	.07	**.39**
6. Learning Transformative	.02	-.05	**-.64**	-.10	.00	-.02	.05	**.37**
22. Curriculum Humanistic	.24	.05	**-.51**	.16	.04	.15	-.12	**.38**
2. Teaching Apprenticeship	.09	-.10	**-.39**	-.35	-.05	.09	.16	**.27**
4. Teaching Social Reform	-.04	.03	-.02	**-.78**	.06	.00	**.72**	.09
5. Teaching Development	-.06	-.11	-.29	**-.67**	-.02	.04	.27	**.29**
19. Curriculum Social Reconstruction	.20	.11	.09	**-.55**	.03	.12	**.59**	-.07
3. Teaching Transmission	.36	.07	.09	**-.53**	-.01	**.46**	.20	-.15
Inter-factor Correlations								
I	1.00				1.00	-.08	.02	-.19
II	-.12	1.00				1.00	.39	.26
III	-.20	.12	1.00				1.00	.23
IV	-.36	-.04	.28	1.00				1.00

Notes. The strongest loadings are shown in bold. Joint EFA conducted with direct oblimin rotation, while multi-battery EFA used direct quartimin oblique rotation.

The inter-correlations between the four factors were low regardless of procedure. Even the social change factor had considerably similar correlations with the other factors in both joint and multi-battery conditions. The mean difference of correlations is .05, after adjusting for sign.

CONCLUSION

The major complication in using factor analysis to explore the structure of scale scores taken from two or more batteries is the power of inventory or battery method to obscure meaningful relations or common traits that may be present. MTMM analysis has been successfully extended to factor analysis with multi-battery factor analysis. This paper has explained the logic and procedures of MBFA and given a detailed tutorial in the use of Cudeck's (1982) MBFACT software. A case study (consisting of 233 cases, five batteries, and 22 scale scores) contrasting joint and multi-battery factor analysis has demonstrated the ability multi-battery factor analysis to overcome the confound of method. One factor in joint factor analysis was made up to two common trait variables and two variables from one of the related batteries; whereas, in MBFA, the two common trait variables were isolated and the two battery-related variables were freed to move to other multi-battery factors.

Results from MBFA are usually very similar to common or joint factor analysis, but have the added virtue of having taken into account method covariance. Once an exploratory MBFA solution is found, however, the researcher is still obliged to test the fit of the solution to the data through application of confirmatory factor analysis (see Payne, Finch, & Tremble Jr., 2003 for an example). However, the researcher can have confidence that a confirmatory model would have good grounds for separating or combining traits across methods because of the application of MBFA. The final decision to accept the MBFA solution ought to depend on the results of a confirmatory analysis.

A further alternative to MBFA would be to increase sample size so that modeling could be done with factors as true latent variables rather than as observed variables. With large samples relative to the number of variables it might be possible to complete a fresh factor analysis using all the items from all batteries and combine the items into new integrated factors and achieve the goal of overcoming method artefacts. This may change the factors and reduce comparability with other studies that had used the original factors. However, it may actually uncover better measurement precision for common traits by using the best of items from the contributing batteries.

Nevertheless, as long as comparison of factor scores is required, and as long as researchers use multiple sources for their measurement instruments, MBFA should be used to disentangle method and trait artefacts. The use of multi-battery factor analysis is warranted whenever researchers are exploring new constructs, have small sample sizes, want to maximize the possibility of detecting common traits, and want to ensure that the resulting factors overcome artificial similarities created by shared method.

References

Bentler, P. M. (1977). Factor simplicity index and transformations. *Psychometrika, 42*, 277-295.

Bentler, P. M., & Huba, G. J. (1982). Symmetric and asymmetric rotations in canonical correlation analysis: New methods with drug variable examples. In N. Hirschberg & L. G. Humphreys (Eds.), *Multivariate Applications in the Social Sciences* (pp. 21-46). Hillsdale, NJ: Erlbaum.

Brewer, J., & Hunter, A. (1989). *Multimethod Research: A Synthesis of Styles*. Newbury Park, CA: SAGE.

Brown, G. T. L. (2004a). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports, 94*, 1015-1024.

Brown, G. T. L. (2004b). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Policy, Principles and Practice, 11*(3), 305-322.

Brown, G. T. L. (2006). Teachers' instructional conceptions: How learning, teaching, curriculum, assessment, and self-efficacy interconnect. In A. P. Presscott (Ed.), *The Concept of Self in Education, Family and Sport* (pp. 1-48). New York: Nova Science Publishers.

Browne, M. W. (1979). The maximum-likelihood solution in inter-battery factor analysis. *British Journal of Mathematical and Statistical Psychology, 32*, 75-86.

Browne, M. W. (1980). Factor analysis of multiple batteries by maximum likelihood. *British Journal of Mathematical and Statistical Psychology, 33*, 184-199.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin 56,* 81-105.

Cheung, D. (2000). Measuring teachers' meta-orientations to curriculum: Application of hierarchical confirmatory analysis. *Journal of Experimental Education, 68*(2), 149-165.

Cudeck, R. (1982). Methods for estimating between-battery factors. *Multivariate Behavioral Research, 17*, 47-68.

Cudeck, R. (1991). Program MBFACT: Maximum Likelihood Solution for Multiple Battery Factor Analysis. (Version 3) [Software]. Columbus, OH: Author.

Finch, J. F., & West, S. G. (1997). The investigation of personality structure: Statistical models. *Journal of Research in Personality, 31*, 439-485.

Finch, J. F., Panter, A. T., & Caskie, G. I. L. (1999). Two approaches for identifying shared personality dimensions across methods. *Journal of Personality, 67*(3), 407-438.

Guskey, T. R., & Passaro, P. D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal, 31*, 627-643.

Huba, G. J., Newcomb, M. D., & Bentler, P. M. (1981). Comparison of canonical correlation and interbattery factor analysis on sensation seeking and drug use domains. *Applied Psychological Measurement, 5*(3), 291-306.

Huba, G. J., Palisoc, A. L., & Bentler, P. M. (1982). ORSIM2: A FORTRAN program for symmetric and asymmetric orthogonal rotation of canonical variates and interbattery factors. *The American Statistician, 36*(1), 62.

Lam, T. C. M., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement, 19*(4), 317-322.

Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing, 6*(4), 311-360.

Meiring, D., Van De Vijver, F., R., R. S., & Sackett, P. R. (2006, July). *Internal and External Bias of Cognitive and Personality Measures in South Africa.* Paper presented at the 18th International Congress of the International Association for Cross-Cultural Psychology (IACCP), Spetses, Greece.

Osborne, J. W., & Costello, A. B. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7).

Payne, S. C., Finch, J. F., & Tremble Jr., T. R. (2003). Validating surrogate measures of psychological constructs: The application of construct equivalence to archival data. *Organizational Research Methods, 6*(3), 363-382.

Ransom, D. C., Fisher, L., & Terry, H. E. (1992). The California Family Health Project: II. Family world view and adult health. *Family Process, 31*, 251-267.

Strauman, T. J., & Wetzler, S. (1992). The factor structure of SCL-90 and MCMI scale scores: Within-measure and interbattery analyses. *Multivariate Behavioral Research, 27*(1), 1-20.

Tait, H., Entwistle, N. J., & McCune, V. (1998). ASSIST: A reconceptualisation of the Approaches to Studying Inventory. . In C. Rust (Ed.), *Improving Student Learning: Improving Students as Learners* (pp. 262-271). Oxford: Oxford Centre for Staff and Learning Development.

Tucker, L. R. (1958). An inter-battery method of factor analysis. *Psychometrika, 23*(2), 111-136.

Citation

Brown, Gavin T L (2007). An Introduction To Multi-Battery Factor Analysis: Overcoming Method Artefacts. *Practical Assessment Research & Evaluation*, 12(7). Available online: http://pareonline.net/getvn.asp?v=12&n=7

Author

Correspondence concerning this article should be sent to Dr Gavin T. L. Brown, School of Teaching, Learning & Development, Faculty of Education, University of Auckland, Private Bag 92019, Auckland, New Zealand or by email to gt.brown@auckland.ac.nz .