Factor Analysis and Correspondence Analysis Composite and Indicator Scores of Likert Scale Survey Data

Matthew J. Powers
The Joint Staff, Future Joint Force Development
116B Lake View Parkway, Suffolk VA 23435
UNITED STATES OF AMERICA

matthew.j.powers.mil@mail.mil

ABSTRACT

Surveys are valuable tools in complex irregular warfare arenas for gaining insight into population attitudes, beliefs, and perceptions. Factor analysis and correspondence analysis unlock statistical power underlying survey response data by measuring latent traits that describe human terrain as continuous random variables or aggregated values appropriate for further analysis. These variables are survey composite scores that are calculated through factor analysis of linearly related response sets and correspondence analysis of those responses with no obvious relationship, linear or otherwise. Factor analysis composite scoring respects the variability of the raw survey data while calculating latent traits and translating the regression scores to reflect the semantics of the original survey Likert scales. Similarly, perpendicularly projecting correspondence analysis biplot column coordinate points onto row coordinate vectors (or vice-versa) and calculating point-intercept distances generates composite scores that are easily translatable to the same original survey Likert scales. This innovative analytic approach anticipates linear and non-linear response sets and ensures easily interpretable results that adhere to the common demand to treat survey responses as ordinal variables.

1.0 BACKGROUND

This paper presents innovative correspondence analysis indicator score calculation to complement Starkweather's (2012) approach to composite or indicator scoring through factor analysis of survey data. Surveys are valuable common assessment tools used to gain insight into complex military operational environments such as irregular warfare (IW) and Security Cooperation (SC). They are an intuitively suitable data collection method for measuring perceptions and attitudes in these population-centric arenas. A common problem amongst survey assessments exists in the univariate analysis approach that results from low appreciation for analytic potential underlying a well-crafted survey instrument. In analyses of surveys whose respondent and question bank sizes produce volumes of charts and pictures, assessments are expensive, repetitive, confusing, and difficult to use. Conversely, long-winded narrative often fails to create a clear picture while too-brief verbal descriptions suggest biased selection and partiality. Problems compound when "analysis" is misrepresented as simply cherry-picking from a set of responses, an approach that is also highly vulnerable to analyst bias (Wilson and Stern 2001).

In addition to the problems of analyst bias, only indirect and incomplete measurement of latent factors represented by attitudes, perceptions, and beliefs that underlie human terrain is possible via single survey questions. That is, "individual survey questions are often imperfect measures of the population traits of interest," thereby requiring extraction of relevant information about the population or populations. Factor analysis composite scoring and correspondence analysis indicator scoring are two proposed methods for measuring these relevant latent traits (Fricker et al 2012).

The next section of this paper discusses why factor and correspondence Analyses are useful measurement methods amidst the common analytic limitations inherent to surveys. Next, the paper introduces and describes calculation of factor analysis composite and correspondence analysis indicator scores, including an experimental example and limitations of both methods. This paper concludes with the advantages and disadvantages of these approaches prior to answering the question of how these analyses can change how the Alliance thinks.

1.1 Why are Factor Analysis and Correspondence Analysis Useful?

Survey data are often of the Likert form. The questions are of an ordinal type that ask the respondent to provide input on levels of adequacy, agreement, etc. Military assessment practices such as stoplight charts are typically of the ordinal variety as well. Unfortunately, this presentation technique has limited data analysis to falsely identify trends within complex systems while ignoring the underlying nuance. Factor analysis combines Likert response survey data into a composite score of the unmeasured latent trait. Similarly, correspondence analysis calculates indicator scores for response-type variables that do not satisfy the linear requirements necessary for factor analysis loading. These scores, be they composite or indicator, are themselves variables of interest appropriate for traditional analysis such as linear regression or simple averaging. These methods and their resultant scores overcome the analytic limitations imposed by misunderstanding or mishandling of survey data (Starkweather 2012).

1.2 Analytic Limitations of Survey Data Analysis: Misunderstanding and Mishandling

Synthesizing survey data into useful information requires understanding that item-by-item analysis can be overwhelming and that it is infeasible to capture every characteristic of interest in a complex operational environment with single survey questions (Fricker et al 2012). Disorganized quantitative or qualitative single-question “laundry list and fuzzy jumble” analysis that attempts question-by-question summaries without analysing relationships or significant meaning is a poor reporting practice that confuses the audience and simultaneously fails to offer insight into the human terrain. Frequently, quantitative analysis stops at response summary statistics while open responses and interviews attempt to explain the results. This approach is anecdotal at best and there is no framework to describe variable relationships or population characteristics (Carifio and Perla 2007).

The nuanced qualitative nature of complex military operational environments is frequent justification for analytic shortcomings; preaching that the collection, analysis, and interpretation of data “cannot be reduced to numbers.” On the contrary, self-imposed limitations (such as the phrase “reduced to numbers”) represent a fundamental lack of analytic understanding, be it quantitative or qualitative. Survey-based assessment are not “reducing” data to numbers, but rather “recoding” them to numbers, thereby opening the door to a myriad of possibilities (Payne and Osburg 2013).

When recoding Likert responses to numbers, it is important to be familiar with the controversy surrounding the practice of treating these numbers as “numbers.” Is it ever appropriate to treat ordinal variables as continuous (Kulzy and Fricker 2015)? It is common to simply average each response value across several questions to calculate a “composite score for the domain which the questions are believed to be assessing” or individually for single question indicator scores. This is controversial because it assumes equal “distance” between Likert anchor points and, with respect to composite scoring, treats each question as contributing equally. Factor analysis reveals that equal contribution to latent variable structure is often not the case while correspondence analysis displays the varying distances between response options (Starkweather 2012).

The other side of the controversy surrounding numeric analysis of ordinal scale variables manifests in Carifio and Perla (2007), "Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes" and by extension Glass et al (1972). Carifio and Perla (2007) confront the concept that non-parametric procedures are the only proper approach to analyzing Likert and/or ordinal-type data. They refer to the Glass et al (1972) Monte Carlo study of analysis of variance (ANOVA) which demonstrates that "the F-test was incredibly robust to violations of the interval data assumption…and could be used to do statistical tests at the scale and subscale level of data that was collected using a 5 to 7 point Likert response format with no resulting bias." Glass et al (1972) concludes that a priori testing of Likert data, akin to that of factor analysis which models observed variables as linear combinations of a known number of factors, with a sufficient number of scale points is extremely robust. Assessors need not sacrifice "statistical power and sensitivity by using non-parametric statistical tests" when analyzing Likert scale data. Regardless of the varying opinions towards analytic propriety, the existence of controversy is evident. Factor analysis composite scoring and/or correspondence analysis indicator scoring attempt to unite the opposing arguments by generating variables that are amenable to analyses that do not sacrifice statistical power.

2.0 COMPOSITE AND INDICATOR SCORES

Likert scaled data analysis of linearly related response subsets representing an unmeasured underlying "continuously scaled latent factor" requires numeric recoding via one factor models and/or correspondence analysis indicator scores of unrelated response sets, both of which are scalable to reflect the semantics of the original scale. In addition to the benefits of interpretability, the composite and indicator scores respect the varying response intervals and closely approximate the latent construct of the original data (Starkweather 2012).

What follows is an explanation of Starkweather's (2012) approach to factor analysis composite score and correspondence analysis indicator score calculations.

2.1 Factor Analysis Composite Scores

Factor Analysis composite scores are appropriate when there is strong belief behind underlying data structure. Hypothesised structures are common in carefully constructed surveys aimed specifically at measuring these latent traits. These beliefs are confirmed through linearly related response vectors or common loading within a multi-factor model. Starkweather (2012) is highly recommended reading for a detailed description and practical example of this process. However, the general procedure for generating factor analysis composite scores includes the following four steps:

- Recode ordinal (such as Likert) responses to numeric responses
- Run a single factor analysis model onto the variables related to the latent factor of interest
- Save the factor scores and factor loadings
- Rescale the factor scores to calculate composite scores that reflect the original response range
  - This requires the factor loadings, the weighted mean, and the weighted standard deviation of the original data. The factor loadings are the weights applied to the original means and standard deviations.

This process appeals to analysts who appreciate analytic possibilities of continuously scaled variables regardless of their levels of agreement with Carifio and Perla (2007) or Glass et al (1972). Audience appeal exists because of the interpretability inherent in the semantic rescaling of the latent factor scores. Despite the intuitive comprehension of rescaled factor Analysis composite scores, Starkweather (2012) points out that values may be "slightly below" or "slightly above" the original scale endpoints because this process models the latent "true scores." This possibility is in contrast with the following correspondence analysis approach of measuring perpendicularly projected biplot column points onto row point line segments that assigns highest scaled values to points with no perpendicular intercept (Starkweather 2012).

2.2 Correspondence Analysis Indicator Scores

In cases where there are response variables that do not display necessary linear relationships, Starkweather (2012) recommends exploring correspondence analysis without explicitly explaining a process. This paper describes in detail a 4-step indicator scale that obeys the Borg and Groenen (2005) non-Euclidean projected assessment:

- Transform responses into a contingency table
- Apply correspondence analysis and generate biplot row/column coordinates
- Perpendicularly project column points onto line segments between response-option row points and measure point-intercept distances to/from segment endpoints
- Rescale this distance to reflect original data scale by utilizing intermediate identically calculated response row point projections onto the same line segment

The next section describes an example of both factor analysis and correspondence analysis scoring methods applied to controlled data. This demonstration clarifies the steps of both approaches and enables comparison of the results.

3.0 COMPSOITE AND INDICATOR SCORES OF A CONTROLLED EXPERIMENT

In this experiment we have survey data compiled from 40 responses to 7-point Likert scaled questions regarding the adequacy of nine "requirements." These responses translate from the range of "1-Extremely Inadequate" to "7-Extremely Adequate", while 4 represents a middle or neutral opinion. The recoding from adequacy context to numbers satisfies step 1 of the 4-step factor analysis composite score process. These fictional data are designed so that three separate factors load onto the nine requirements by threes in sequential order (see figure 1). For example, factor 1 loads heaviest onto requirements 1-3. Other design features of these data are as follows:

- Responses to requirements 1, 4, and 7 distribute uniformly between 1-7
- Requirement 2 is the least adequate requirement
- Requirement 9 is the most adequate requirement

	requirement1	requirement2	requirement3	requirement4	requirement5	requirement6	requirement7	requirement8	requirement9
1	6	2	4	5	4	4	7	7	7
2	5	2	4	4	4	3	1	4	5
3	6	2	4	7	5	4	7	7	7
4	5	2	4	1	3	2	4	5	7
5	1	1	3	5	4	4	3	5	6
6	1	1	3	6	5	4	7	6	7
7	4	1	4	7	5	4	7	7	7
8	1	1	3	4	4	3	2	4	5
9	3	1	3	3	4	3	6	6	7
10	6	2	4	5	4	4	7	7	7
11	3	1	3	2	3	2	4	5	7
12	4	1	4	7	5	5	3	5	6
13	1	1	3	6	4	4	4	5	7
14	7	2	4	1	3	2	7	7	7
15	1	1	3	2	3	2	1	4	5
16	6	2	4	4	4	3	2	4	6
17	7	2	4	3	4	3	4	5	7
18	3	1	3	2	3	2	2	4	6
19	7	2	4	5	4	4	7	6	7
20	6	2	4	6	5	4	3	5	6
21	4	1	4	3	4	3	7	6	7
22	5	2	4	1	3	2	6	6	7
23	3	1	3	4	4	3	1	4	5
24	6	2	4	6	4	4	4	5	7
25	4	1	3	5	4	4	2	4	5
26	5	2	4	3	4	3	6	6	7
27	7	2	4	5	4	4	5	6	7
28	3	1	3	4	4	3	3	5	6
29	3	1	3	4	4	3	4	5	7
30	3	1	3	4	4	3	6	6	7
31	6	2	4	1	3	2	7	6	7
32	2	1	3	5	4	4	7	7	7
33	1	1	3	4	4	3	3	5	6
34	2	1	3	5	4	4	4	5	7
35	7	2	4	3	3	3	7	6	7
36	4	2	3	1	3	2	5	6	7
37	7	2	4	4	4	3	7	7	7
38	7	2	5	7	5	5	2	4	6
39	4	2	3	2	3	2	3	5	6
40	6	2	4	3	3	3	2	4	6

Figure 1. Fictional survey data of 7-point Likert responses regarding adequacy of nine requirements. By design, requirements 1-3, 4-6, and 7-9 are heavily loaded on by three separate factors identified separately by color in the figure.

The second of Starkweather's (2012) 4-step process is to run a single-factor analysis model on the variables that cluster through heaviest loads or simple linear relationships. With knowledge of the data structure we know that three separate "requirement factors" generate their own unique sets of regression scores and factor loadings which are necessary to calculate the rescaled composite scores, thereby satisfying Starkweather's (2012) steps 3 and 4, respectively (see figure 2).

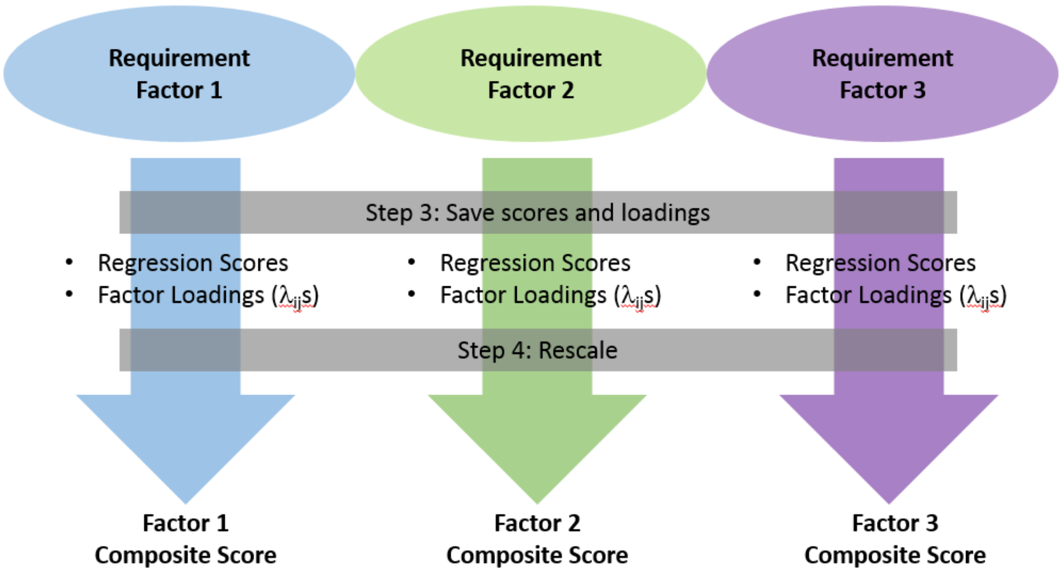

Figure 2. Composite score calculation steps 3 and 4. Regression scores and factor loadings combine with original response data means and standard deviations to rescale the composite scores to reflect the survey scales.

Final factor analysis composite scores for these data are discussed after the correspondence analysis indicator score process example.

3.1 Controlled Example: Correspondence Analysis Indicator Score

Step 1 of the 4-step indicator score process requires translating response data (figure 1) into a contingency table (table 1) amenable to correspondence analysis.

Table 1. Contingency table of requirement adequacy response distributions. Column header "X1" represents a Likert response of "1-Extremely Inadequate," etc.

	X1	X2	X3	X4	X5	X6	X7
requirement1	6	2	7	6	4	8	7
requirement2	19	21	0	0	0	0	0
requirement3	0	0	18	21	1	0	0
requirement4	5	4	6	9	8	4	4
requirement5	0	0	11	23	6	0	0
requirement6	0	9	15	14	2	0	0
requirement7	3	6	6	7	2	4	12
requirement8	0	0	0	9	13	11	7
requirement9	0	0	0	0	5	10	25

Correspondence analysis biplots reveal underlying data structure but the distances between row points (requirements) and column points (response options, “X1,” “X2,” etc.) cannot be directly interpreted. However, the relationship between these coordinate sets can be assessed by perpendicular projection, thereby enabling a measurement between point-line intercepts and response coordinates to calculate an interval score (see figure 3) and respectively satisfying steps 2 and 3 of the 4-step process.

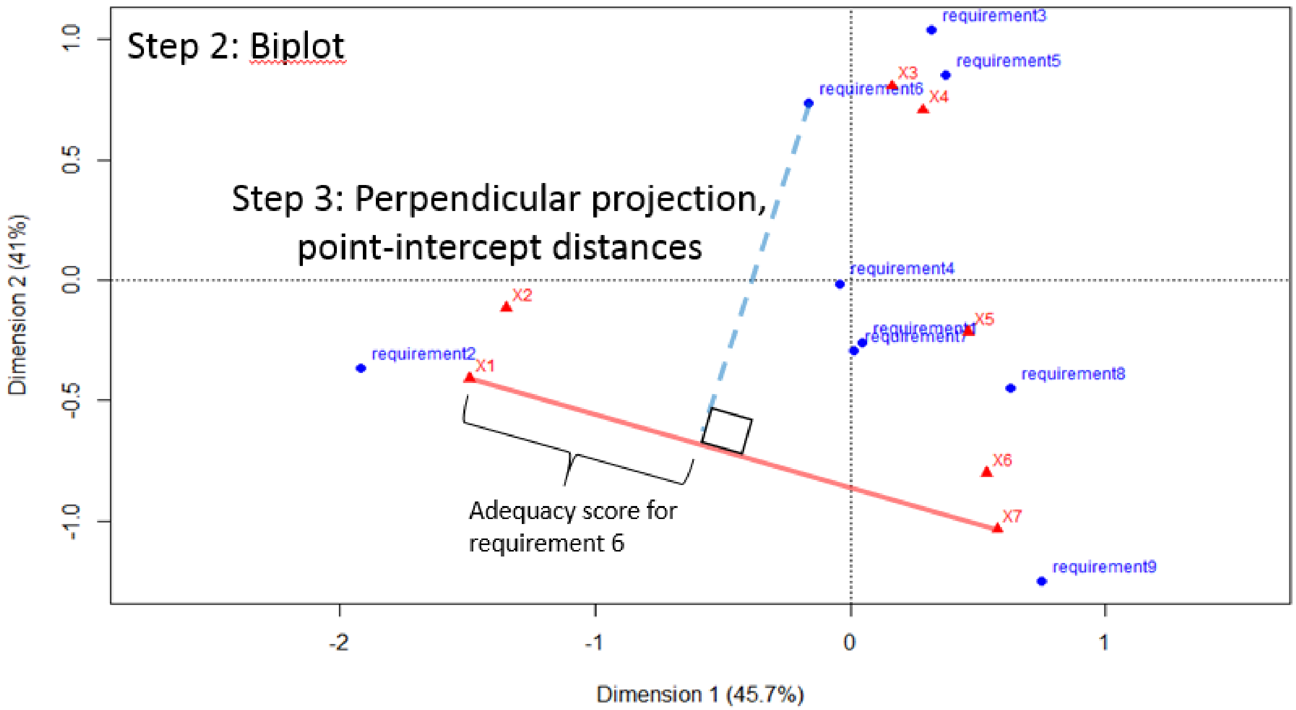

Figure 3. Correspondence analysis biplot of experiment data contingency table with sample perpendicular projection of requirement 6 coordinates onto line vector between extreme response points of *X1-Extremely Inadequate* and *X7-Extremely Adequate*.

Scaling simply requires a similar projection-measurement process of intermediate column points onto the same line segment in figure 3. The resultant non-identical intermediate values visualize the theoretical uneven distances between response options underlying the aforementioned controversy regarding survey data treatment (see figure 4).

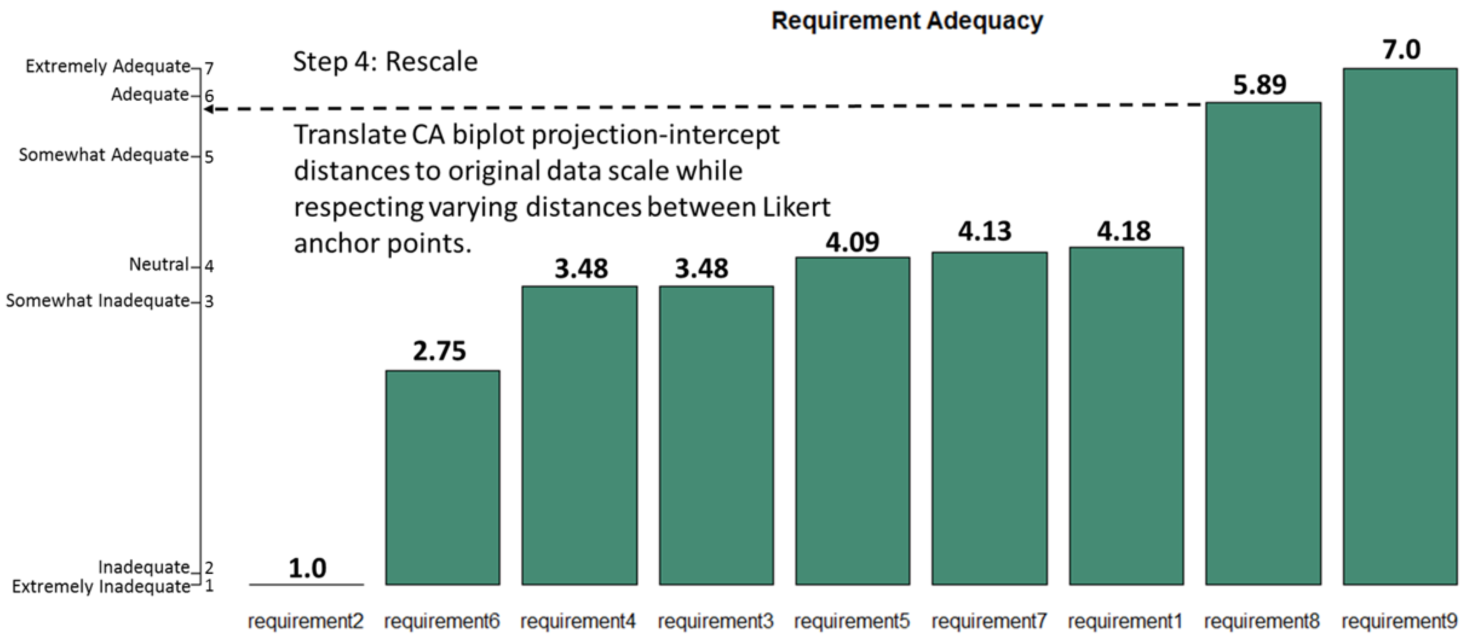

Figure 4. Scaled point-intercept values from Correspondence Analysis biplot. This scaling process demonstrates the unequal distance between the seven response options as seen on the seven tick marks on the Y-axis.

3.2 Analyzing the Results

Since factor analysis composite scores originate from multiple variables we compare them to appropriate mean correspondence analysis indicator scores and grouped response averages. The results are plotted in figure 5.

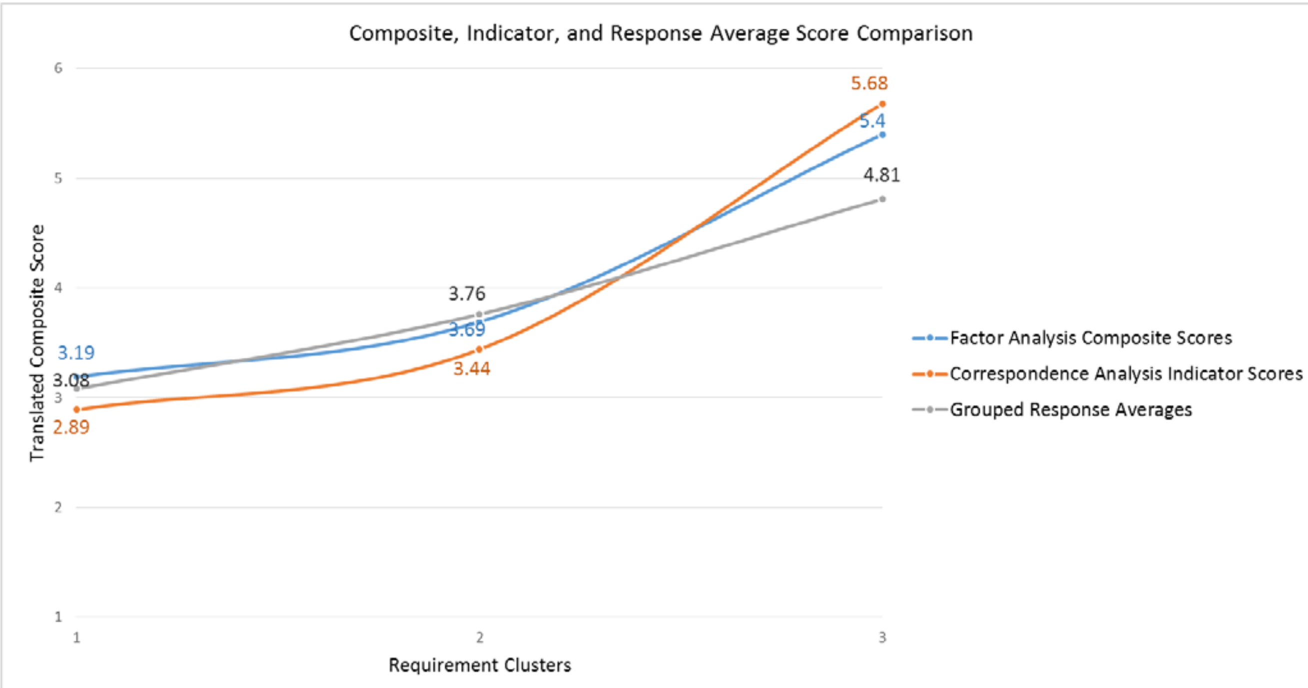

Figure 5. Comparison of composite scores to mean indicator scores and response averages. The most substantial difference is between response averages and composite/indicator scores at requirement cluster 3 with differences of 0.59 and 0.87, respectively (far right plot points). Response averages do not display the same curve the other two trend lines.

Figure 6 compares requirement-by-requirement correspondence analysis indicator scores against response averages.

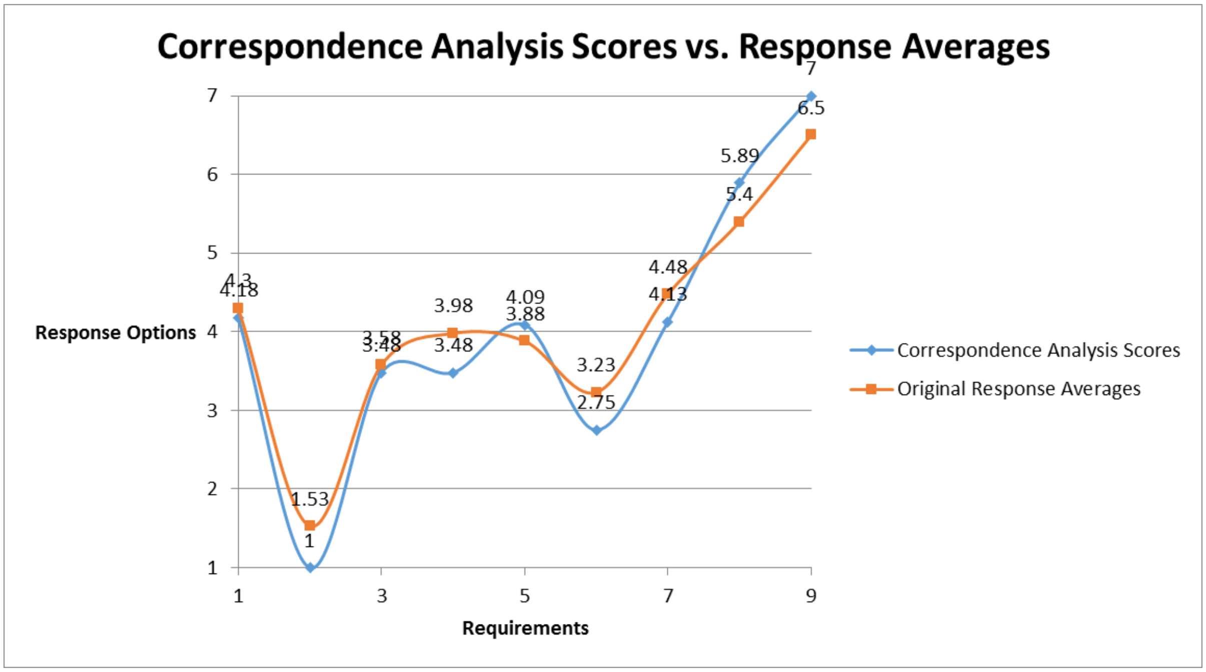

Figure 6. Comparison of correspondence analysis indicator scores and response averages. The mean score difference is 0.37 with a max of 0.53. These differences are relatively small and they carry similar semantics while producing nearly homogeneous curves, hence the controversy regarding the propriety of averaging.

Response averages aside, figure 5 shows that while factor analysis and correspondence analysis produce similar composite/indicator scores, a question remains; which is the preferred method? One answer to that question is that both methods can be combined. First, apply factor analysis to variables that demonstrate linear relationships, and for remaining variables apply the correspondence analysis method. Of course, the real answer is that it depends. It depends on the discovery goals of the overall analysis, such as latent traits or data structure. It depends on if the analysis stops with the score or continues with traditional statistical analysis along the lines of regression. Regardless of which method is preferred or required it is helpful to understand the limitations.

3.2 Limitations

3.2.1 Factor Analysis Composite Score Limitations

Starkweather (2012) highlights a common factor analysis criticism of substantial (and perhaps unexpected) loadings on each latent factor in single multi-variable models. Should this occur during the exploratory phase then a conflict of interest arises in terms of the multi-model analysis plan vs. the actual single model loadings, calling into question the data structures underlying multiple models. This conflict may require "subjective choices" and "data manipulation," both of which are common practices in mathematical and statistical modelling and should be reported in the analysis. Correspondence analysis is not as vulnerable to this criticism (Starkweather 2012).

The theoretical existence of latent factors loading onto multiple variables coerces multiple variables represented by a single composite score. While this simplification may be advantageous, it can be a limitation if analysis requires item-by-item scoring for variables that display heavy loading that cannot be ignored.

3.2.2 Correspondence Analysis Indicator Score Limitations

The biggest analytic limitation to the correspondence analysis method is that the result is a single number vice a continuous random variable. This in turn limits the analytic possibilities. Also, since the score is a distance measurement between column point-line intercepts and row response point line vectors, it is possible for points to possess no perpendicular intercept (see Figure 3, requirement points 2 and 9, for example). These cases assume extreme response values, which are 1.0 and 7.0 in our experiment.

There are rare cases where contingency table data structure can break the model. For instance, if an outlying single response exists in an extreme response column, the corresponding row point coordinates may plot in such a way that an extreme-to-extreme line segment enables no feasible perpendicular projections. One fix for this that has shown to be robust against substantial indicator score deltas is analogous to sensitivity analysis where the offending column is removed from calculation. Such a fix requires consideration as to the outlying degree of the offending data point(s). Another possibility is to project each column point onto appropriate line segments inclusive of intermediate row points. This method enjoys the advantages of measuring distances between responses since row-to-row distances are directly interpretable, but requires assumptions regarding line segment projection choice and it is a difficult process to automate.

Despite these limitations, both factor analysis composite scores and correspondence analysis indicator scores offer many advantages.

4.0 CONCLUSION

Factor analysis composite scores and correspondence analysis indicator scores unlock traditional analysis potential, particularly the composite scores that manifest as continuous random variables of interest. Moreover, these methods are olive branches between analysts who subscribe to the non-parametric-only approach to survey analysis and those who adhere to the Carifio and Perla (2007) assessment. These approaches avoid scattershot, item-by-item analyses by revealing theoretical latent traits and data structure.

4.1 How does this analysis change how Alliance thinks?

Resistant-to-change command climates ensure that radical adjustments to existing assessment processes are not likely to be quickly realized. These existing processes include surveys and analogous methods such as stoplight charts or other flavors of ordinal data collection. Factor analysis and correspondence analysis increase the utility of commonly assessed ordinal data by breaking the shackles of fake arithmetic, shotgun analysis, and volumes of charts.

This paper proposes survey analysis as a means of understanding the operational environment. Latent factors and otherwise indistinguishable data structures reveal how the operational environment reacts to missions across human terrain. Surveys are difficult to craft, administer, and collect, but the Alliance acknowledges their value as an assessment tool. Survey analysis oftentimes falls short of investigating the underlying data and focuses too heavily on statistical summaries. Misleading analysis invites a decision that is heavily dependent on commander's bias with little or no rigorous backing (Downes-Martin 2011). These rigorous, theory-driven analysis approaches reveal operational environment nuances so commanders can better understand *why* a condition exists (Williams and Morris 2009).

REFERENCES

[1] Borg, Ingwer, and Patrick JF Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer Science & Business Media. (2005).

[2] Carifio, James, and Rocco J. Perla. "Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes." *Journal of Social Sciences* 3, no. 3 (2007): 106-116.

[3] Downes-Martin, S. "Operations Assessment in Afghanistan is Broken". *Naval War College Review*. 64(4). (2011).

[4] Fricker Jr, Ronald D., Walter W. Kulzy, and Jeffrey A. Appleget. "From data to information: Using factor analysis with survey data." (2012).

[5] Glass, Gene V., Percy D. Peckham, and James R. Sanders. "Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance." *Review of educational research* (1972): 237-288.

[6] Kulzy, Walter W. and Ronald D. Fricker Jr., "The Survey Process: With an Emphasis on Survey Data Analysis." PHALANX, 45(4), 30-34 (2015).

[7] Payne, L. A. and Osburg, J. "Leveraging Observations of Security Force Assistance in Afghanistan for Global Operations." RAND Arroyo Center Santa Monica, CA. (2013)

[8] Starkweather, Jon. "How to Calculate Empirically Derived Composite or Indicator Scores." *Benchmarks Online*, 9 June 2016, http://it.unt.edu/sites/default/files/benchmarks-02-2012.pdf (February 2012).

[9] Wilson, I. M. and Stern, R. D. "Approaches to the Analysis of Survey Data." Statistical Services Centre, The University of Reading, UK. (2001).