FRAME ANALYSIS: A MODERN APPROACH TO FACTOR ANALYSIS

Master's Thesis
Presented to
The Academic Faculty

by

Ryan C. Sanchez

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
December 2021

FRAME ANALYSIS: A MODERN APPROACH TO FACTOR ANALYSIS

Approved by:

Dr. Michael D. Hunter, Advisor
School of Psychology
Georgia Institute of Technology

Dr. James S. Roberts
School of Psychology
Georgia Institute of Technology

Dr. Keaton Fletcher
School of Psychology
Georgia Institute of Technology

Date Approved: August 30, 2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
CFrA	Constrained Frame Analysis
CFrA-R	Constrained Frame Analysis – Replicating
CFrA-NR	Constrained Frame Analysis – Non-Replicating
FA	Factor Analysis
FV	Frame Vector
IPIP	International Personality Item Pool
RMSE	Root-mean-squared error

SUMMARY

We introduce and develop a new statistical method for exploring latent structures: Frame Analysis. Frame Analysis drops the one-to-one correspondence between factor dimensionality and vector space representation found in Factor Analysis. This minor change obviates factor rotations, simple structure, and provides equal status for cross-loaded items. We show that in Frame Analysis, manifest items are defined by only one frame loading and are uniquely characterized as a linear combination of latent variables: a frame vector. Through a series of simulations, we characterize Frame Analysis performance in three scenarios: Exploratory, Constrained, and Partially-Constrained. Finally, we apply Frame Analysis to archival five-factor personality data and provide evidence that hierarchical personality models are disguised frame vectors.

INTRODUCTION

Identifying hidden traits and relations is an essential method in the modern research psychologist's tool bag. This practice's origin is traceable back to Spearman (1904, 1927) and his inquiry into human mental ability. Spearman (1904) observed that schoolchildren who scored high in one domain also scored high in other disciplines. This observation led him to hypothesize that a hidden general ability 'g,' measurable only indirectly, explained the correlations. He stated, "Whenever branches of intellectual activity are at all dissimilar, then their correlations with one another appear wholly due to their being all variously saturated with some common fundamental Function" (Spearman, 1904, p. 273). Thus, by compressing many overt assessments into a single covert ability, he established the first method for scientifically quantifying the unmeasurable: a factor analysis.

Thurstone (1935) mathematically grounded and extended the nascent factor analysis method into a robust exploratory technique to systematically uncover hidden traits. He introduced several important concepts in factor analysis that remain impactful a century later. First, factors represent primary traits that map the range of observed measurements (Thurstone, 1935, p. 150). Whether experimentally measured, self-reported through an inventory, or by other means, primary traits reflect the belief that an "unlimited number of phenomena can be comprehended in terms of a limited number of concepts or ideal constructs" (Thurstone, 1935, p. 44). Second, Thurstone (1935, p. 69, 120) used linear algebra to conceptualize factors mathematically as vectors oriented in a multidimensional space; the expansion coefficients (i.e., pattern coefficients) defined the relationship between the unobserved factors and the measured variables.

Third, Thurstone identified that rotational indeterminacy presented an interpretive barrier for factors. To circumvent this indeterminacy, he proposed sparse representations of expansion coefficients, simple structure (Thurstone, 1935, p. 151) to constrain identification and facilitate factor interpretability. Factors should be rotated or transformed such that each item is described by only one factor. Browne (2001) provided a "rotation bible" for the assortment of possible constraints leading to simple structure. This systematic search for simple structure, or something approximating it, led researchers to hypothesize hierarchical or bi-factor models (Holzinger & Swineford, 1937) to aid factor interpretation.

In addition to rotational indeterminacy interpretive problems, factor proliferation (Shaffer et al., 2015) via rotational duplication is also possible. Rotational duplicates are traits formed from linear combinations of other, established traits. For example, the direction north-east is a rotational duplicate of north and east. Ludeke et al. (2019) demonstrated that the agreeableness and emotionality traits of the HEXACO personality model are possible rotational duplicates of the five-factor personality model. Likewise, the construct grit may be a rotational duplicate of the five-factor personality constructs of conscientiousness, agreeableness, and neuroticism (Tables 4 and 9, Duckworth & Quinn, 2009; Table 4, Crede et al., 2017).

Rotational duplication can also affect inventories that reportedly measure the same constructs. When assessing personality with a five-factor inventory, should the researcher use the NEO-PI-R (Costa & McCrae, 2008), the Personality Inventory for *DSM-V* (Krueger et al., 2012), the Five-Dimensional Personality test (van Kampen, 2012), the Big Five Aspects Scale (DeYoung et al, 2007) or the Big Five Inventory 2 (Soto & John, 2017)?

John (2008, p.125) succinctly captured the issue with such an abundance of five-factor personality inventories "... it's possible to ask questions such as '*which* Big Five?' or '*whose* Big Five?'".

Factor analysis's modern standard operating procedure tacitly captures Thurstone's legacy. Yong and Pearce (2013) outline this current procedure which includes 1.) determining the number of factors (dimensionality), 2.) factor extraction, 3.) rotation for interpretation (simple structure), and possibly 4.) dropping cross-loaded items. The present work reimagines steps 3 and 4 by dropping the implicit equivalence between dimensionality and the number of vectors. Thurstone (1935, p. 69, 120) *defined* factors as vectors which implies the number of factors and vectors are equal. Our alternative approach decouples the number of vectors from the number of factors extracted. The number of factors still defines the dimensionality of the vector space, but we propose allowing linearly dependent vectors to partition this space. These new linearly dependent vectors are compositions of the dimensionality[1] and form an overdetermined system. Figure 1 shows a vector space that is exactly determined (left) and overdetermined (right). By hypothesizing an overdetermined vector space, we can no longer use traditional factor analytic methods for analysis. We need to establish new analytical strategies and standard operating procedures that properly account for the vector space redundancy while maintaining an interpretive connection with classical factor analysis.

In the present work, we propose and develop a new statistical technique that we call frame analysis. Frame analysis builds upon the foundation of frame vectors, just as factor analysis builds upon the foundation of basis vectors. Frame analysis reconceptualizes

[1] Imagine north and east on a map. The directions north-east and north-north-east are compositions of these two dimensions.

factor extraction and obviates both factor rotation to simple structure and dropping cross-loaded items. In the upcoming sections, we outline properties, define terms, and explore three use cases for frame analysis. Next, we provide several examples as a point-of-departure from both traditional exploratory and confirmatory factor analysis. We then perform a simulation study to characterize the performance of frame analysis under conditions with known solutions. Finally, we apply frame analysis to archival five-factor personality data to demonstrate an alternative understanding of the item-factor relationship.

1.1 Frame Analysis

The multiple factor analytic approach outlined by Thurstone (1935, p. 120) extracts a set of basis vectors (primary axes) equal to the number of factors hypothesized. This one-to-one correspondence affords a traditional linear algebraic approach for further interpretation. However, we wish to extend the factor analytic approach to include redundancy or linear dependence. Figure 1 demonstrates the difference between an exact and overdetermined vector space; mathematicians define an overdetermined vector space as a frame (Kovacevic & Chebira, 2007a; Kovacevic & Chebira, 2007b; Morgenshtern & Bolcskei, 2011). Note that a frame maintains the same dimensionality, or rank, as the extracted basis vectors. Therefore, we denote the analysis performed in this overdetermined vector space as a frame analysis and explicate the properties below.

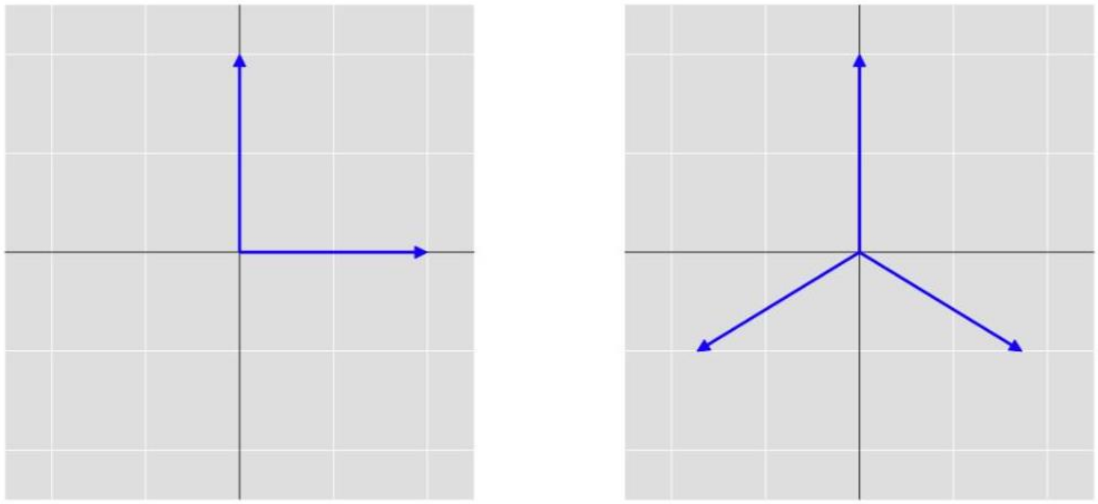

Figure 1 – Standard and Frame Representation of 2D Space. *Left*. Standard representation of two-dimensional space. Each point in the grid is represented by a unique combination of the blue axes. *Right*. Frame Representation. Each point is represented by an infinite number of combinations of the blue axes.

1.1.1 Basis and Frame Vectors

An n-dimensional vector space is spanned by a set of n linearly independent vectors forming a basis set. Any point in this vector space spanned by the basis vectors is uniquely determined and identified by the pattern coefficients in factor analysis. Frames include the basis set and linear combinations of these basis vectors to include p vectors ($p > n$). Figure 1 (left) illustrates a two-dimensional space defined by two orthogonal basis vectors and an overrepresented (right) two-dimensional space, a frame. Similar to factor analysis, pattern coefficients in frame analysis are contingent upon the vectors used to define the frame.

1.1.2 Taxonomic Frame Vectors

Frame vectors can either be data-driven or selected a priori for desirable properties. Methods for finding data-driven frame vectors are conceptually similar to oblique rotations discussed by Browne (2001). Several drawbacks to data-driven strategies are 1.) they can only estimate frame vectors for collected data, 2.) researchers must specify the number of frame vectors, and 3.) may contain more than one frame coefficient. In Figure 2 (left), collected data are used to define three frame vectors. Data in proximity to each frame vector would result in a simple structure pattern coefficient representation. However, the lone data point in quadrant IV would exhibit non-zero contributions on at least two frame vectors.

An alternative to data-driven techniques is to cover the *n*-dimensional space with equally spaced vectors, a vector taxonomy. The angular spacing defines a quantization error that represents a lower bound on similarity. Returning to Figure 2 (right), frame vectors are evenly spaced in polar angle, 22.5 degrees, covering the entire two-dimensional space. Each data point has a sparse frame loading represented by the bisecting vector of the containing wedge. By definition, data contained within a wedge are equivalent, and this compression results in quantization error. Quantization error is controlled by increasing or decreasing the angle between frame vectors. We note that the cosine of the wedge angle is equivalent to correlation and in the above example is $\cos(22.5) \approx .92$. Practically this means items correlated above .92 are, by definition, equal.

Moving forward, we discard the data-driven approach to computing frame vectors and adopt the taxonomic system. When discussing frame vectors, we imply taxonomic frame vectors and label the collection of taxonomic frame vectors for a given dimensionality: the frame dictionary. Appendix A describes the creation of the frame dictionary.

1.1.3 Frame Loadings

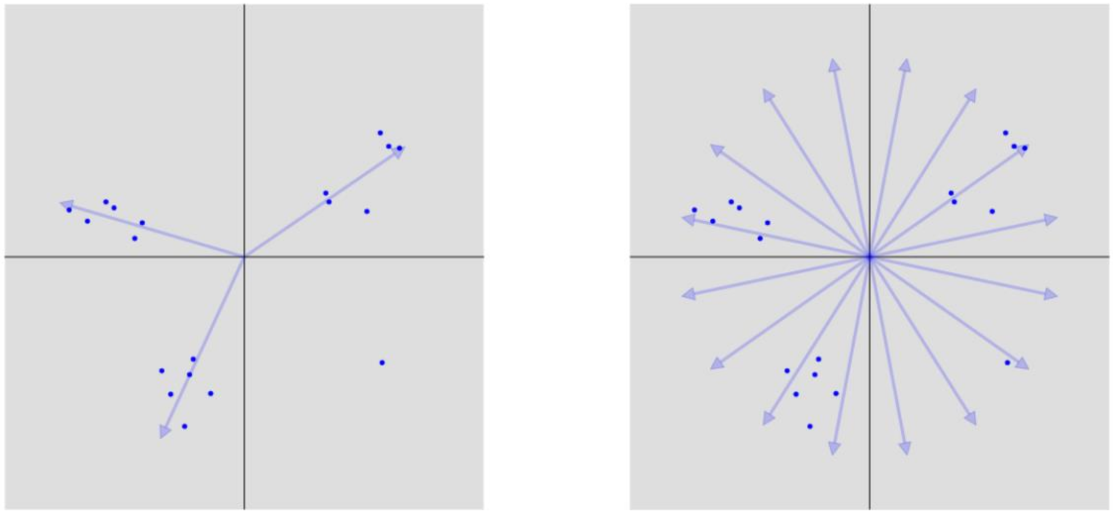

Figure 2 - Two approaches for computing frame vectors. *Left*. Data-driven frame vectors. The lone measurement in quadrant IV will not have a sparse representation. *Right*. Taxonomic frame vectors. Measurements lying in a partition are definitionally similar.

Frame loadings represent an item's correlation with the underlying latent variables. Computing the analog of structure coefficients in frame analysis is more complex than in factor analysis. Because the vector space is exactly determined, there is only one possible inverse to compute. Conversely, the redundancy found in frames results in an infinite number of possible inverses, a family of matrix inverses defined by the left inverse (Morgenshtern and Bolcskei, 2011). Using the frame vector taxonomy outlined above, a solution where each item is associated with only one frame vector is possible. Here we see an advantage in frame analysis that is not present in factor analysis: Rotation to simple structure is obviated by representing each item with only one frame vector.

Determining the frame loadings and associated frame vectors requires solving an overdetermined system. The most common method for solving overdetermined systems is the pseudo-inverse, representing a least-squares solution. Unfortunately, least-squares methods penalize sparse solutions that would result in the desired one-to-one relationship between frame vector and item. Therefore, we seek an alternative inverse that minimizes the number of non-zero frame loadings to one. Because we adopted a frame dictionary that nominally covers the entire vector space, this inverse is equivalent to identifying the closest frame vector. We provide an algorithm outline for calculating frame loadings in Appendix B.

Interpreting the frame loading's meaning is straightforward and we note that the single frame loading determined in a frame analysis is related to the communality found in factor analysis. This understanding has several benefits. First, the frame loading maintains a connection with factor analysis. Second, frame analysis decouples the frame loading from the coordinate transformation. The frame loading is the square root of the common variance explained and is unchanged, up to quantization error, under rotations. Finally, the frame vector correlation matrix, the inner product of frame vectors, captures inter-item correlation information similar to the factor correlation matrix (Harman, 1967, p.240).

1.1.4 Frame Rotations

Another frame analysis benefit is that it obviates the need for rotation to simple structure in an exploratory analysis. Since the multidimensional vector space is uniformly partitioned, which guarantees only one frame loading, orthogonal rotations or oblique transformations provide no extra information or interpretative benefit. However, an orthogonal rotation less than the frame separation angle that minimizes quantization error may still be applied. The result of such a rotation is a potential reduction in the number of expressed[2] frame vectors.

1.1.5 Exploratory Frame Analysis

Exploratory frame analysis determines which frame vectors are expressed in the dataset and the strength of that expression. Due to the one-to-one relationship between items and frame vectors, items expressed on the same frame vector are categorically related. When the number of expressed frame vectors equals the dimensionality, then frame analysis is conceptually equivalent to factor analysis, up to quantization error. When the number of expressed frame vectors is greater than the dimensionality, this is analogous to having at least one cross-loaded item in factor analysis. Again, each item in a frame analysis has only one non-zero frame loading, and cross-loaded items manifest in the frame correlation matrix.

Interpretation in an exploratory frame analysis is simplified compared to factor analysis. The single frame loading *is* the square root of the communality and a standardized (partial-) regression coefficient. The frame correlation matrix defines the relatedness among all items, and the element-wise square of this matrix quantifies the shared variance.

1.1.6 Constrained Frame Analysis

Constrained frame analysis operates like confirmatory factor analysis (Jöreskog, 1969). However, a significant difference between the two analyses is the method used to estimate model parameters. Confirmatory factor analysis is couched in structural equation

[2] We denote a frame vector with a non-zero frame loading as expressed.

modeling and requires setting reference items for identification purposes. In constrained frame analysis, frame vectors are supplied from previous studies and used to determine frame loadings that minimize the same cost function used during the exploratory processes. The result is that there are no methodological differences between exploratory and constrained frame analysis, and the researcher can use their favorite factor analytic extraction method and use either the correlation or covariance matrix.

It is necessary to orthogonally rotate either the extracted loadings or the supplied frame vectors when performing a constrained frame analysis. This rotation aligns the current extracted coordinate system with the previously supplied one and offers no interpretative information beyond this alignment. The quantization scheme used in a taxonomic frame aids constrained frame analysis since it only penalizes models that change quantization bins. A comparison between the frame vectors from an exploratory frame analysis and the supplied frame vectors may help identify and diagnose changing items when a misfit does occur.

1.1.7 Partially-Constrained Frame Analysis

Exploratory and constrained frame analysis represent two ends of the informational landscape. When researchers have no information about the latent space arrangement, an exploratory frame analysis can reveal the relationship between items. When researchers have complete information about the latent space relationships, constrained frame analysis can validate this structure. A third condition exists when researchers have only partial information about the latent space arrangement.

This third condition occurs when the manifest variables consist of known and unknown items. Supplied frame vectors from a previous analysis constrain the known items, and the orienting process inherent in constrained frame analysis results in identifying the frame vectors for the unknown items. Performing a partially constrained frame analysis allows researchers to determine how new items relate to established items, determine a new construct's discriminant and convergent validity with a legacy construct, or link equivalent forms that share a small subset of items.

1.2 Frame Analysis Examples

In this section, we show simple examples for the methods described above. Each illustration is performed on synthetic data after applying principal axis factoring to the correlation matrix. We set the lower bound on correlation to 0.925, which results in a frame dictionary with eight frame vectors in two dimensions. We limit our analysis to a two-dimensional factor space to facilitate illustrations and comprehension, but frame analysis can operate in any finite multidimensional vector space.

1.2.1 Exploratory Frame Analysis

We start with a 10-item simulated inventory that has seven simple structure items and three cross-loaded items. Table 1 shows the generating loadings and the resulting correlation matrix from simulated data. Table 2 shows the extracted loadings after principal axis factoring, frame analysis loadings, and frame correlations. Factor analysis's rotation to simple structure is also shown for comparison. The frame loadings associated with frame vectors 1 and 4 have zero correlation, suggesting that items related to these two frame vectors share no variance. Similarly, frame vectors 3 and 5 are also orthogonal; however, there is a 45-degree rotation between these two orthogonal pairs.

Table 1 – True Loadings and Simulated Correlation Matrix

Item	Loadings		Correlation Matrix									
	Factor 1	Factor 2	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
1	1.0	0.0	—									
2	1.0	0.0	.559	—								
3	1.0	0.0	.505	.551	—							
4	0.0	1.0	.037	.058	.003	—						
5	0.0	1.0	-.038	.028	-.067	.559	—					
6	0.0	1.0	.006	.028	-.027	.596	.590	—				
7	.707	.707	.490	.478	.407	.404	.371	.397	—			
8	-1.0	0.0	-.582	-.582	-.565	-.043	.056	.006	-.450	—		
9	-.309	-.951	-.181	-.173	-.131	-.596	-.566	-.587	-.511	.157	—	
10	-.588	.809	-.303	-.310	-.336	.467	.477	.506	.053	.349	-.396	—

Table 2 also demonstrates the advantages frame analysis has over factor analysis. First, the number of frame vectors immediately informs the researcher about item heterogeneity. In our two-dimensional example, if the number of expressed frame vectors is greater than two, we can infer that at least one item is categorically composed of the two factors. In factor analysis terminology, a linear transformation to a Thurstonian simple structure solution is not possible. Similarly, if only one frame vector was expressed, then we may have overestimated the dimensionality.

Second, cross-loaded items (7, 9, and 10) are afforded equal status under frame analysis. The frame loading's magnitude is related to the amount of common variance explained by the item, and the sign defines the direction in multidimensional space.

Finally, the frame loadings are independent of global rotations, and the frame correlation matrix identifies the relationships between items. These metrics succinctly capture the amount of common variance explained by the item (frame loadings) and the amount of variance common with other items (frame correlation matrix).

Similar to factor analysis, the researcher has discretion in naming the frame vectors if desired[3]. External requirements often influence the naming approach taken. One option is to assign a unique construct to each frame vector without considering the redundancy inherent in frames. Another option is to identify and name a subset from frame vectors. In the above example, selecting any two linear independent frame vectors is necessary and sufficient. A majority voting criterion would select frame vectors 3 and 5 from Table 2; other scenarios where a minority or single item with high frame loadings or interpretability are also valid criteria.

Table 2 – Extracted and Rotated Factor Loadings Compared to Frame Loadings

	Extracted Loadings		Rotated Loadings		Frame Loadings				
Items	Factor 1	Factor 2	Factor 1	Factor 2	FV 1	FV 2	FV 3	FV 4	FV 5
1	-.437	.601	-.743	-.020	–	–	–	–	.743
2	-.461	.600	-.755	-.040	–	–	–	–	.755
3	-.373	.608	-.712	.037	–	–	–	–	.713
4	-.649	-.395	-.042	-.759	–	–	-.760	–	–
5	-.594	-.457	.040	-.748	–	–	-.747	–	–
6	-.643	-.449	.005	-.784	–	–	-.784	–	–
7	-.764	.213	-.608	.510	-.791	–	–	–	–
8	.431	-.647	.777	-.011	–	–	–	–	-.777
9	.747	.260	.209	.763	–	.784	–	–	–
10	-.274	-.719	.437	-.633	–	–	–	-.759	–
							Frame Correlations		
					FV 1	.924	.707	0	-.707
						FV 2	.924	.383	-.383
							FV 3	.707	0
								FV 4	.707
									FV 5

Note. FV: Frame Vector

Only 5 of the possible 8 frame vectors had non-zero frame loadings. The 3 unexpressed frame vectors are omitted.

[3] Although naming is possible, we do not recommend naming every frame vector since the number of frame vectors increases considerably with dimensionality.

A third option is to use the multidimensional canonical bases and align a set of expressed frame vectors as close as possible. In this option, the multidimensional space is defined by a set of *n* orthogonal vectors that require naming. An attractive feature of this option is that expressed frame vectors can still be individually named if necessary but are recognized as a linear combination of orthogonal factors. Researchers can now use these canonical bases as the standardized frame vectors for (partially-) constrained frame analysis.

1.2.2 Constrained Frame Analysis

We simulated a second dataset with the same parameters as the exploratory frame analysis example above but with slightly more unique variance per item. A principal axis factoring extracts two factors, and we note that both constrained and exploratory frame analysis share the same setup. We then expand the extracted loadings onto the supplied frame vectors found during the exploratory analysis, maintaining the item-frame vector

Table 3 – Constrained Frame Loadings

Items	Extracted Loadings		Constrained Frame Loadings				
	Factor 1	Factor 2	FV 1	FV 2	FV 3	FV 4	FV 5
1	-.459	.377	–	–	–	–	.577
2	-.401	.536	–	–	–	–	.669
3	-.374	.365	–	–	–	–	.516
4	-.419	-.387	–	–	-.567	–	–
5	-.494	-.375	–	–	-.620	–	–
6	-.558	-.410	–	–	-.693	–	–
7	-.539	.202	-.562	–	–	–	–
8	.369	-.471	–	–	–	–	-.598
9	.589	.158	–	.610	–	–	–
10	-.223	-.617	–	–	–	-.642	–

Note. FV: Frame Vector

Extracted loadings were from a principal axis factor solution. Constrained frame analysis uses previous frame vectors and forces an item-vector relationship.

relationship above. The effects of increased unique variance on the frame loadings are immediately evident in Table 3; the absolute frame loadings are smaller.

By constraining items to correspond with specific frame vectors, we impart the relative relationship between items that was discovered from a previous analysis. Conceptually, this is like confirmatory factor analysis. However, our method leverages a frame analytic approach instead of a structural approach outlined by Jöreskog (1969).

1.2.3 Partially-Constrained Frame Analysis

To demonstrate a partially constrained frame analysis, we simulated a third dataset that consisted of 10 items: 5 items from the previous studies (1, 2, 6, 7, and 9) and five previously unused items. The frame vectors associated with the known items constrain and orient the factor space, whereas the five unknown items are assigned frame vectors after the factor space orientation. Table 4 shows the found relationship between the unknown

Table 4 – Partially-Constrained Frame Loadings

Items	Extracted Loadings		Frame Loadings					
	Factor 1	Factor 2	FV1	FV2	FV3	FV4	FV5	FV6
1	-.585	.444	–	–	–	–	.734 [a]	–
2	-.591	.440	–	–	–	–	.737 [a]	–
6	-.441	-.566	–	–	-.717 [a]	–	–	–
7	-.739	-.069	-.741 [a]	–	–	–	–	–
9	.618	.468	–	.772 [a]	–	–	–	–
11	-.398	-.632	–	–	-.745 [b]	–	–	–
12	.638	-.420	–	–	–	–	-.762 [b]	–
13	-.580	-.394	–	-.701 [b]	–	–	–	–
14	.487	-.589	–	–	–	–	–	-.756 [b]
15	-.724	-.004	-.717 [b]	–	–	–	–	–
					Frame Correlations			
			FV 1	.924	.707	0	-.707	-.383
				FV 2	.924	.383	-.383	0
					FV 3	.707	0	.383
						FV 4	.707	.924
							FV 5	.924
								FV 6

Note. FV: Frame Vector

Extracted loadings were from a principal axis factor solution.

[a] Constrained frame loadings

[b] Exploratory frame loadings

items and the available five items. We note that four new items (11, 12, 13, and 15) provided redundant measurements given by frame loadings on the existing frame vectors. Item 14 is unique enough to warrant an additional expressed frame vector, frame vector six. We should conclude that the new items provide additional measurements in our theoretical analysis but do not specify new, unique constructs. Additionally, we know how item 14 (frame vector 6) relates to item 10 (frame vector 4) by examining the frame correlation matrix, even though we did not include item 10 in the analysis. Frame analysis's ability to relate items not included in the same analysis is a novel contribution and provides research opportunities not currently possible.

1.3 Experiment 1: Simulation Study

We ran a simulation study to characterize the performance of exploratory, constrained, and partially constrained frame analysis. We simulated cross-loaded items randomly from a uniform distribution centered at zero and terminating at +/- one and normalized the squared sum of the loadings for each item to one. Hypothetical respondents with latent traits were selected from a standardized multivariate Gaussian distribution, $N(0, \mathbf{I})$[4], and used to simulate the response data. We manufactured unique variance by adding Gaussian noise, $N(0, \sigma)$, $\sigma \sim N(.25, .05)$, to each item of the simulated response data. Finally, we randomly selected four quantile thresholds between zero and one, exclusive, from a uniform distribution to quantize the simulated data into five Likert-like response categories.

Next, we computed the polychoric correlation (Olsson, 1979) matrix from the simulated digitized data and used principal axis factoring to determine the extracted loadings. All data were simulated with a pseudo-random number generator for repeatability, and Python scripts are available upon request for explicit computations and replicability. Specific considerations for each algorithm used in the simulation study are outlined below.

1.3.1 Metric Definitions

The two metrics we used to measure loading similarity are root-mean-squared error and congruence. We define the calculation of these metrics below.

[4] N(x, y) represents a normal distribution with mean, x, and variance y or covariance **y**.

1.3.1.1 Root-Mean-Squared Error

Root-mean-squared error (RMSE) is similar to a standard deviation but with the mean replaced by a known value. We use this metric to compare the extracted loadings from factor analysis with reconstructed loadings from frame analysis. Let L represent the extracted loadings matrix, R represent the reconstructed loadings matrix from frame analysis, and indices *i*, *j*, identify the matrix components. Then the RMSE is computed as:

$$\boldsymbol{RMSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{IJ}}\sum_{\boldsymbol{ij}}\left(\boldsymbol{L_{ij}} - \boldsymbol{R_{ij}}\right)^{\mathbf{2}}}$$

1.3.1.2 Congruence

Congruence is the root mean squared dot product between the frame vectors from a previous and current analysis. We use this metric in partially constrained frame analysis to compare the frame vectors expressed for the unconstrained items with the known frame vectors on the same items from a previous investigation. Let the P matrix be the set of frame vectors associated with items from an earlier analysis, and the C matrix represents a set of frame vectors expressed by the same items in the current analysis, then average congruence is computed as:

$$d = diag(PC^T)$$
$$c = \sqrt{\frac{1}{K}\sum_{k} d_k^2}$$

The *diag* operator selects elements from the main diagonal of the matrix. The average congruence is insensitive to ∓ sign changes, and we report the minimum or maximum of

Equation 2 to determine if a reflection has occurred. Congruence is valid on the closed-domain from [0, 1], where numbers closer to 1 suggest a higher similarity.

1.3.2 Exploratory Frame Analysis

Our simulation was a 4 factor (2, 3, 4, and 5) x 4 test length (10, 25, 50, and 100) x 4 sample size (100, 250, 500, and 1000) fully crossed factorial design for a total of 64 conditions. We set the taxonomic frame vector quantization scheme to 22.5 degrees which corresponds to an upper bound on the correlation between two different items of approximately .925. The number of frame vectors in the frame dictionary is a function of the number of factors, *n*, and in our quantization scheme is equal to 8, 39, 184, and 860 for 2, 3, 4, and 5 factors, respectively. We measured reconstruction performance by computing the root-mean-squared error between the frame loadings and the extracted factor loadings for 250 realizations in each condition.

1.3.3 Constrained Frame Analysis

Like the constrained frame example above, we leveraged the exploratory frame analysis simulation to gauge constrained frame analysis performance. We simulated a second dataset in each exploratory frame analysis condition with the same parameters and factor loadings to characterize the root-mean-squared error between the frame loadings and extracted loadings in a replication situation. The third dataset in each exploratory frame analysis condition was simulated with the same parameters as the second dataset but with unrelated factor loadings. This dataset characterized the root-mean-squared error between the frame loadings and extracted loadings in a non-replication situation.

1.3.4 Partially-Constrained Frame Analysis

We simulated a scenario where two hypothetical inventories share a subset of items. Both inventories consisted of 3 factors, 30 items, and 500 people. We varied the common items between 5, 10, and 15. Items in the first inventory were randomly created from a uniform distribution centered at zero and terminating at +/- one. The squared sum of the loadings for each item was normalized to one. A second inventory was created with the same parameters as the first inventory. Factor loadings were extracted using principal axis factoring for both simulated inventories. We used the overlapping items to align the two inventories and computed the frame loadings and vectors for the non-overlapping items. We measured the root-mean-squared error between the overlapping items' loadings and extracted loadings. The average similarity computed for non-overlapping items between the recovered frame vectors from the second inventory and the frame vectors found in the first inventory assessed their congruence.

1.3.5 Results

The simulated exploratory and constrained frame analysis reconstructed loadings results are shown in Figures 3 - 6. We note the standard deviation for the 10-item test length in all conditions is larger than the other test lengths, which suggests fortuitous constrained alignments were driving the variability. We performed a Type 3 two-way ANOVA on each factor condition to determine the dependence of sample size and test length and average RMSE.

For two factors, in Figure 3, the average RMSE for exploratory frame analysis remained independent across sample size, $F(3, 3984) = 1.94$, $p > 0.05$, but was dependent on test length, $F(3, 3984) = 432.35$, $p < .001$, $\eta^2 = .245$. The average RMSE slightly increased as the test length increased. In the replicating constrained frame analysis condition, the RMSE was dependent on both test length and sample size, $F(3, 3984) = 41.38$, $p < .001$, $\eta^2 = .016$ and $F(3, 3984) = 1231.87$, $p < .001$, $\eta^2 = .473$, respectively. The average RMSE decreased as sample size increased but increased as test length increased. There is a discernible decrease in RMSE variability as sample size and test length increase. In the non-replicating constrained frame analysis condition, the average RMSE was only dependent on test length, $F(3, 3984) = 240.30$, $p < .001$, $\eta^2 = .153$. The average RMSE increased as test length increased. In this condition, the RMSE variability appeared to only depend on test length.

For the three factor condition, in Figure 4, the average RMSE for exploratory frame analysis was only dependent on test length, $F(3, 3981) = 1947.54$, $p < .001$, $\eta^2 = .594$. The average RMSE was larger than the two factor condition and also increased as test length increased. The average RMSE for constrained frame analysis was dependent on test length, $F(3, 3981) = 147.38$, $p < .001$, $\eta^2 = .038$, sample size, $F(3, 3981) = 2409.75$, $p < .001$, $\eta^2 = .620$, and the interaction, $F(9, 3981) = 2.52$, $p < .01$, $\eta^2 = .002$. The interaction effect size is minute, and, therefore, we interpret the marginal means as average RMSE decreasing with increasing sample size and increasing with longer test lengths. The RMSE variable appears to be dependent on both sample size and test length. We note that the three factor condition had larger *average* RMSE for both exploratory and constrained frame analysis than the two factor condition, whereas constrained frame analysis showed larger *variability* in RMSE for the three factor condition. In the non-replicating constrained frame analysis condition, the average RMSE was only dependent on test length, $F(3, 3984) = 787.12$, $p <$

.001, η^2 =.371. The average RMSE increased as test length increased and is smaller than the two factor condition. The RMSE variability appeared independent of sample size.

For four factors, in Figure 5, the average RMSE for exploratory frame analysis was dependent on test length, $F(3, 3966) = 4965.18$, $p < .001$, η^2 =.787, and sample size, $F(3, 3966) = 8.16$, $p < .001$, η^2 =.001. Again, the average RMSE increased as test length increased but was not larger than the three factor condition. We did observe a dependence on sample size, however, the effect size is negligible at .1%. The average RMSE for constrained frame analysis was dependent on test length, $F(3, 3966) = 116.27$, $p < .001$, η^2 =.030, sample size, $F(3, 3966) = 2445.10$, $p < .001$, η^2 =.626, and the interaction, $F(9, 3966) = 2.52$, $p < .01$, η^2 =.004. Similar to the three factor condition, the interaction is negligible and the marginal means had the same trends as the three factor condition. In the non-replicating constrained frame analysis condition, the average RMSE was dependent on test length, $F(3, 3966) = 2050.97$, $p < .001$, η^2 =.604, and sample size, $F(3, 3966) = 15.76$, $p < .001$, η^2 =.005. The average RMSE increased as test length increased and is smaller than the three factor condition. The sample size dependence's effect size is negligible at .5%. As in the other conditions, the RMSE variability appeared independent of sample size.

For five factors, in Figure 6, the average RMSE for exploratory frame analysis was dependent on test length, $F(3, 3812) = 6832.62$, $p < .001$, η^2 =.829, sample size, $F(3, 3812) = 16.01$, $p < .001$, η^2 =.002, and the interaction, $F(9, 3812) = 5.78$, $p < .001$, η^2 =.002. We note the same dependence and direction of average RMSE on test length as the other factor conditions. The sample size and interaction effect sizes, although significant, are small enough to forgo interpretation. The average RMSE for constrained frame analysis was dependent on test length, $F(3, 3812) = 40.49$, $p < .001$, η^2 =.013, sample size, $F(3, 3812) =$

1723.74, $p < .001$, $\eta^2 = .565$, and the interaction, $F(9, 3812) = 23.76$, $p < .01$, $\eta^2 = .023$. Unlike the other factor conditions, the interaction term effect size is large enough to warrant consideration. Examining Figure 6, we notice large RMSE variability in the 10-item condition. Excluding this condition, leads to a negligible interaction effect size seen in the other factor conditions, $F(6, 2988) = 7.22$, $p < .001$, $\eta^2 = .002$. We interpret this as the average RMSE with few items per factor as being significantly higher than the other conditions. The RMSE variability trends similarly to the other factor conditions. In the non-replicating constrained frame analysis condition, the average RMSE was dependent on test length, $F(3, 3812) = 1268.89$, $p < .001$, $\eta^2 = .491$, sample size, $F(3, 3812) = 11.77$, $p < .001$, $\eta^2 = .005$, and the interaction, $F(9, 3812) = 2.82$, $p < .01$, $\eta^2 = .003$. The average RMSE increased as test length increased and is smaller than the three factor condition. The sample size and interaction effect size is negligible at .5%. The RMSE variability appeared mostly independent of sample size; the 10 item condition showed greater variability in the 100 and 250 sample size conditions. The five factor condition had the lowest average RMSE in all factor conditions simulated.

Figures 7 - 10 show the RMSE between the observed polychoric correlation matrices and those reconstructed by frame analysis. The reconstructed correlation is the outer product of the reconstructed loadings with the unique variance added. Due to this fact, we note that the trends observed in the previous section carry over here and errors compound with the factor analysis reconstruction errors. We plot the average reconstruction RMSE for traditional factor analysis in all figures, which is the average RMSE lower bound for frame analysis reconstruction.

In the two factor simulation condition, in Figure 7, we observed similar trends as the reconstruction loadings with respect to test length dependence. Two notable differences are the stronger sample size dependence and larger average RMSE errors overall. The sample size dependence mimics that of factor analysis but is attenuated due to the quantization error at the loading level. The increased larger average RMSE is due to the compounding errors with the factor analysis reconstruction.

Constrained frame analysis in the replicating conditions showed a similar pattern: an average RMSE decreasing as test length increased. This sample size dependence was stronger than in the exploratory condition but asymptotically approached the exploratory RMSE as the sample size increased. Constrained frame analysis's RMSE variability did show a test length dependence and decreased as test length increased. In addition, the RMSE standard deviation also had a slight sample size dependence suggesting a sample size by test length interaction. The 10-item condition had the largest variability, which suggests fortuitous, constrained alignments were driving the variability. Constrained frame analysis's non-replicating condition showed both a sample size and test length dependence. This trend contrasts with the test length only dependence in loading reconstruction.

We observed similar trends in the three factor, in Figure 8, four factor, in Figure 9, and the five factor conditions, in Figure 10, as the two factor condition. We highlight constrained frame analysis's much larger variability in RMSE for the 10-item condition in the four and five factor figures. This variability suggests the decreased number of items per factor is driving fortuitous and unlucky results.

The partially constrained frame analysis results are shown in Table 5. The average RMSE over the realizations remained stable, whereas the RMSE standard deviation

Table 5 – Partially-Constrained Frame Analysis RMSE and Congruence Performance

	RMSE		Congruence		
Item Overlap	CFrA	EFrA	Minimum	Average	Maximum
5	0.069 (0.019)	0.079 (0.005)	.840 (0.217)	.961 (0.033)	1.0 (0.0)
10	0.082 (0.013)	0.078 (0.006)	.886 (0.024)	.973 (0.011)	1.0 (0.0)
15	0.086 (0.010)	0.077 (0.007)	.895 (0.024)	.976 (0.010)	1.0 (0.0)

Note. Standard errors are in parentheses

decreased as the item overlap increased for the constrained component. The exploratory analysis remained stable and did not vary as the overlap varied. The average congruence between the frame vectors expressed after alignment and the frame vectors from the initial exploratory frame analysis remained high, >.9, but demonstrated greater instability with only five overlapping items.

1.3.6 Discussion

We expected the reconstructed loading's RMSE in exploratory frame analysis to remain stable across test length and sample size conditions as this tracks the quantization error in the frame dictionary. The discovered dependence on test length suggested that as the number of items increased, the uniformity of error within a quantization bin also increased. The average RMSE equivalence in the three, four, and five factor conditions at 100 items suggests this increase occurs asymptotically and is defined by the quantization level. We hypothesize the smaller rate observed in the two-factor condition is due to the uniform partitioning of the frame dictionary in two dimensions versus the quasi-uniform partitioning in three or more dimensions.

The replicating constrained frame analysis demonstrated primarily a sample size dependence on loading reconstruction. The decrease in average RMSE appears asymptotic and approaches the exploratory frame analysis average RMSE. We explain this as a statistical phenomenon whereby more samples increase the probability of selecting the correct frame vector. In a constrained frame analysis, this phenomenon can occur either in the constraining frame vectors or the constrained solution. These results suggest in most practical cases, 10 items on two or more factors may be insufficient. The slight dependence on test length is explained by the same phenomenon as in exploratory frame analysis, the quantization bin utilization.

The non-replicating constrained frame analysis showed a test length dependence on average RMSE loading reconstruction which also decreased across factor conditions. The test length dependence is likely due to the increased probability of finding a fortuitous least-squares rotation in the coordinate system alignment. We hypothesize the factor dependence is due to a restriction of extreme values driven by an increase in factors. As the number of items per factor decreases, the probability of observing large differences from the generating value shrink, resulting in a smaller overall RMSE.

Examining the reconstruction correlation RMSE shows frame analysis does not reach the lower bound provided by factor analysis. This result is expected since frame analysis's current implementation begins with the extracted loadings from factor analysis. The offset between exploratory frame analysis and factor analysis captures the quantization error imparted by the frame dictionary. A frame analysis-specific extraction algorithm might achieve better performance than the current simulation; however, performance on par with factor analysis is likely unachievable due to quantization. Constrained frame analysis's replicating condition had similar RMSE patterns as the frame loading's RMSE, which suggests either observing differences in loadings or correlation would result in similar conclusions. Constrained frame analysis's non-replicating condition showed a test length and sample size dependence for conditions with more than two factors. We have no working hypothesis for the root cause of this trend but note that the RMSE errors are still considerably larger than both exploratory and replicating constrained frame analysis. The correlation reconstruction results suggest analysis at the loading level is appropriate moving forward.

We also demonstrated that constrained frame analysis is a viable alternative approach to confirmatory factor analysis. An advantage that constrained frame analysis maintains over confirmatory factor analysis is the algorithmic consistency; both exploratory and constrained frame analysis use the same extraction methods and provide the same outputs. Constrained frame analysis also requires fewer estimated variables since it uses factor analysis extraction algorithms. However, a critical missing feature in constrained frame analysis is the lack of a goodness-of-fit metric.

One possible metric for assessing goodness-of-fit is inspired by RMSEA 0.05 cutoff criterion from structural equation modeling (Browne & Cudeck, 1992; Hu & Bentler, 1999). Subtracting the exploratory frame analysis RMSE (quantization error) from the constrained frame analysis RMSE and comparing it to a 0.05 threshold might indicate a good versus bad fit. This criterion would suggest a poor fit for replicating constrained frame analysis sample sizes of 100, as seen in Figures 1-4. Another possibility for assessing constrained frame analysis goodness-of-fit is to use a data-driven parallel analysis-like algorithm on the extracted loadings (Horn, 1965). The results from the non-replicating constrained frame analysis approximate this algorithm. Devising a goodness-of-fit metric remains an active area of research.

Partially constrained frame analysis is unique to this new paradigm. We demonstrated that constraining a common set of items between two inventories resulted in high average congruence on the unconstrained items. This outcome suggests our method correctly aligns the coordinate systems between the two inventories, and inferences between the non-overlapping items are valid. Applications for this process include assessing discriminant and convergent validity on new items and determining rotational duplication between two different inventories. We plan to conduct a study that evaluates the required number of questions per factor necessary to stabilize the congruency.

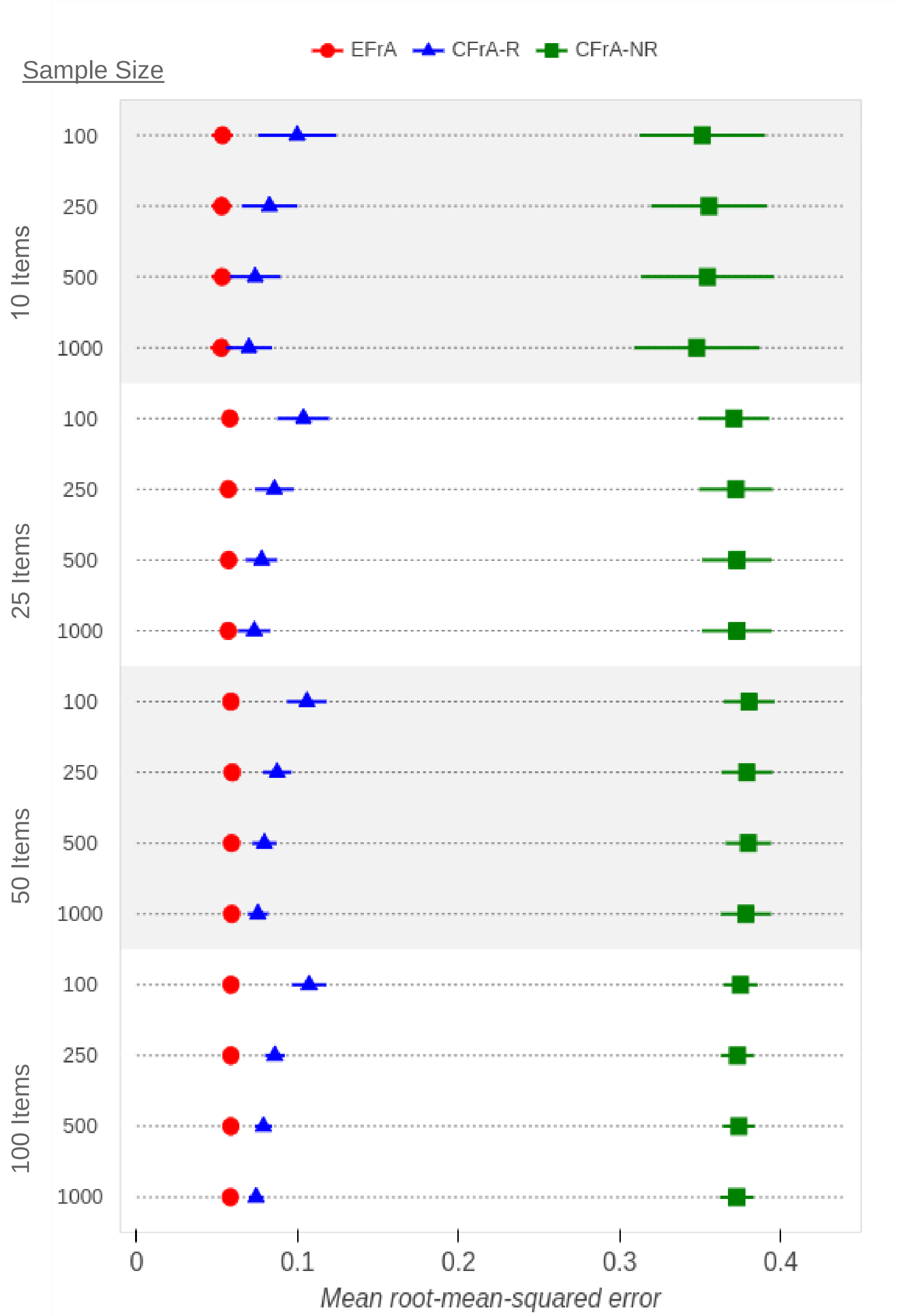

Figure 3 - Root-Mean-Squared Error between Extracted and Reconstructed Loadings for 2 Factor Exploratory and Constrained Frame Analysis averaged over Realizations. *Note*. Solid horizontal lines are standard deviations.

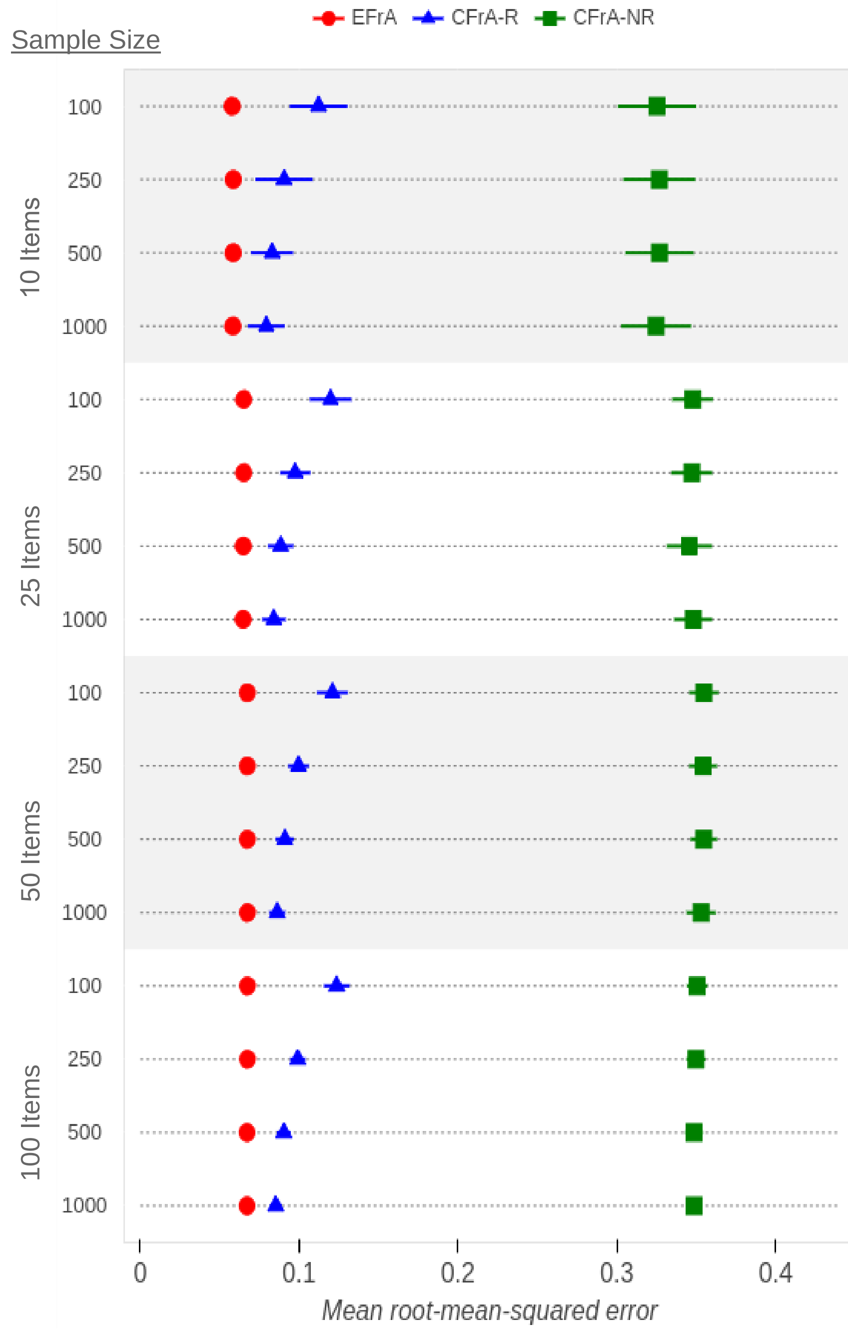

Figure 4 - Root-Mean-Squared Error between Extracted and Reconstructed Loadings for 3 Factor Exploratory and Constrained Frame Analysis averaged over Realizations. *Note*. Solid horizontal lines are standard deviations.

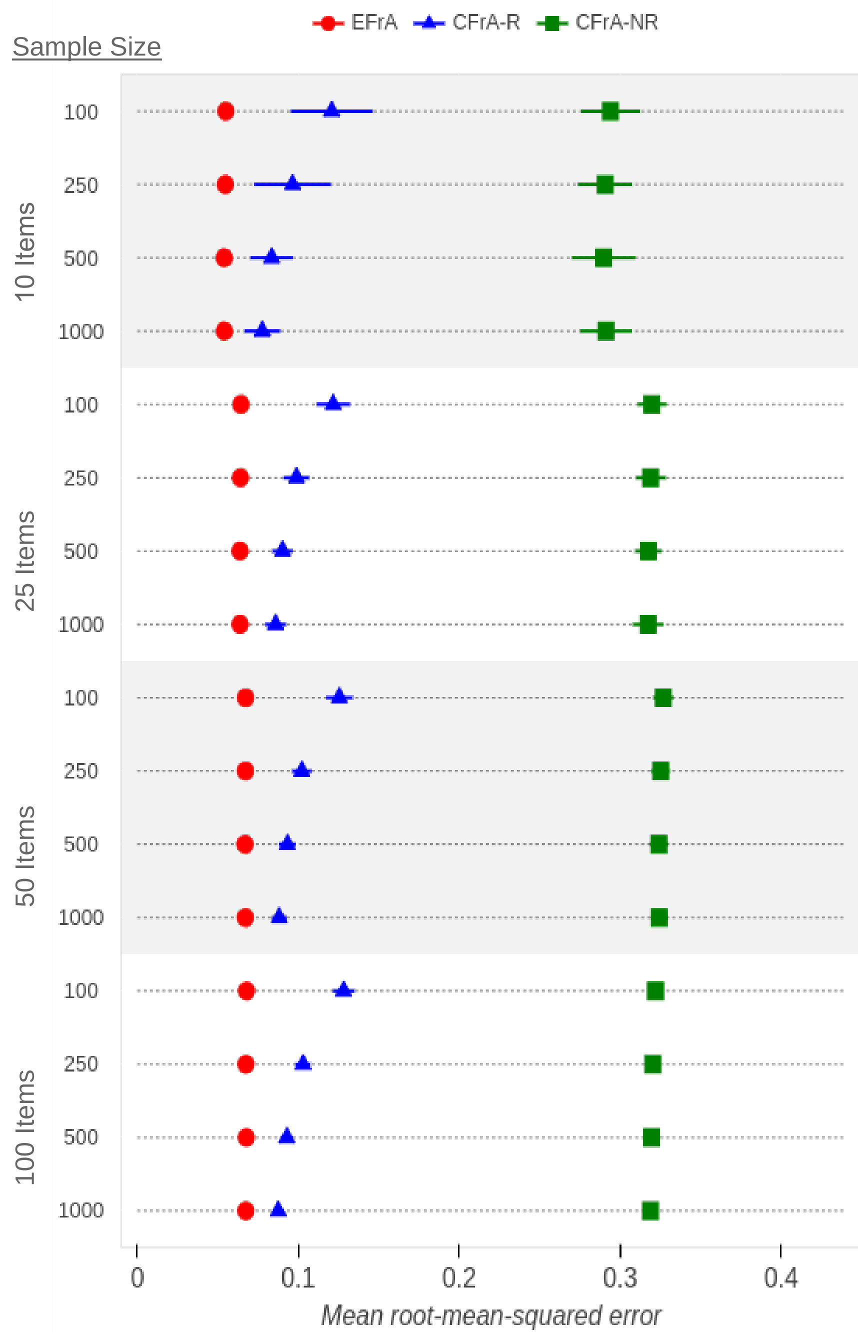

Figure 5 - Root-Mean-Squared Error between Extracted and Reconstructed Loadings for 4 Factor Exploratory and Constrained Frame Analysis averaged over Realizations. *Note*. Solid horizontal lines are standard deviations.

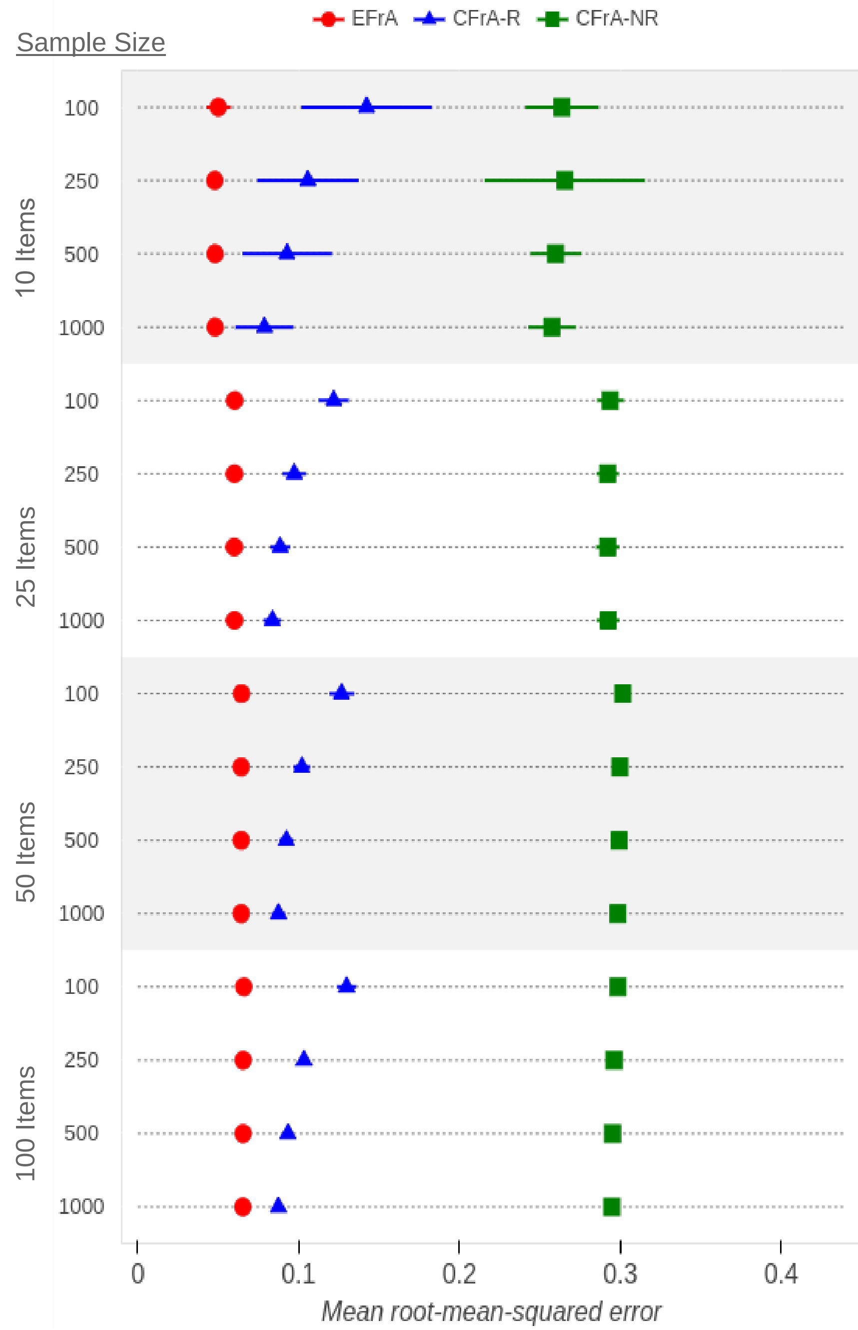

Figure 6 - Root-Mean-Squared Error between Extracted and Reconstructed Loadings for 5 Factor Exploratory and Constrained Frame Analysis averaged over Realizations. *Note*. Solid horizontal lines are standard deviations.

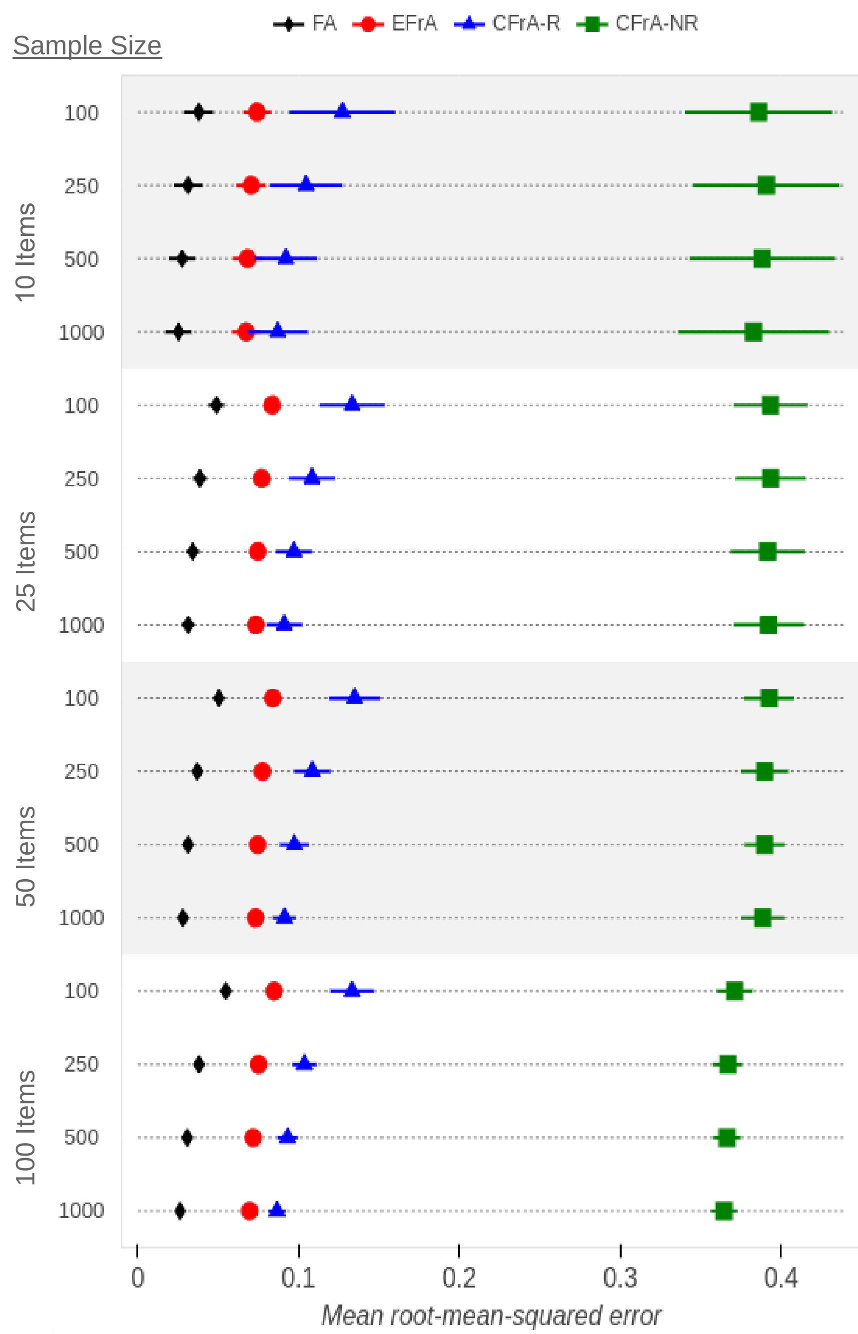

Figure 7 - Average Root-Mean-Squared Error between Polychoric Correlation and Reconstructed Correlation for 2 Factor Exploratory and Constrained Frame Analysis. *Note.* Solid horizontal lines are standard deviations.

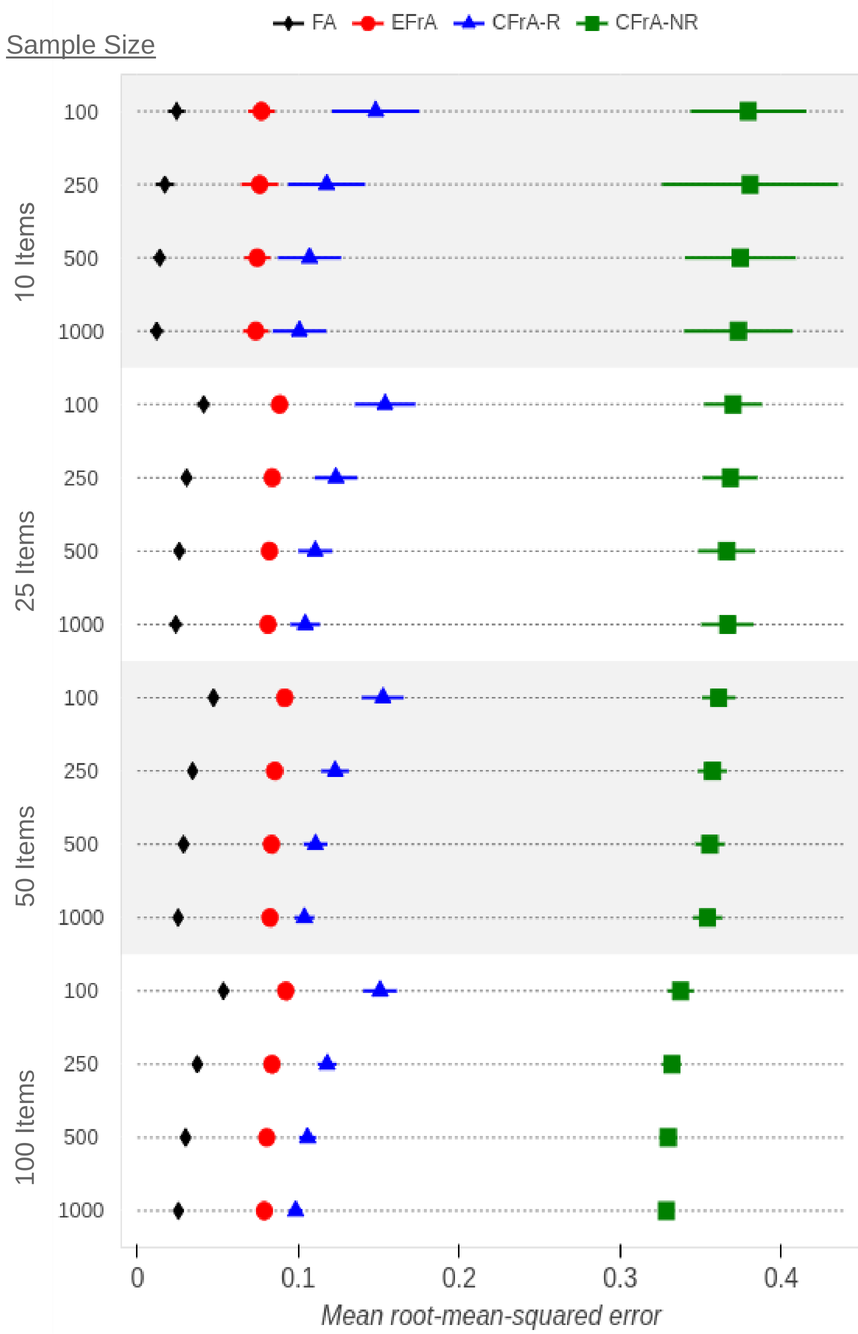

Figure 8 - Average Root-Mean-Squared Error between Polychoric Correlation and Reconstructed Correlation for 3 Factor Exploratory and Constrained Frame Analysis. *Note.* Solid horizontal lines are standard deviations.

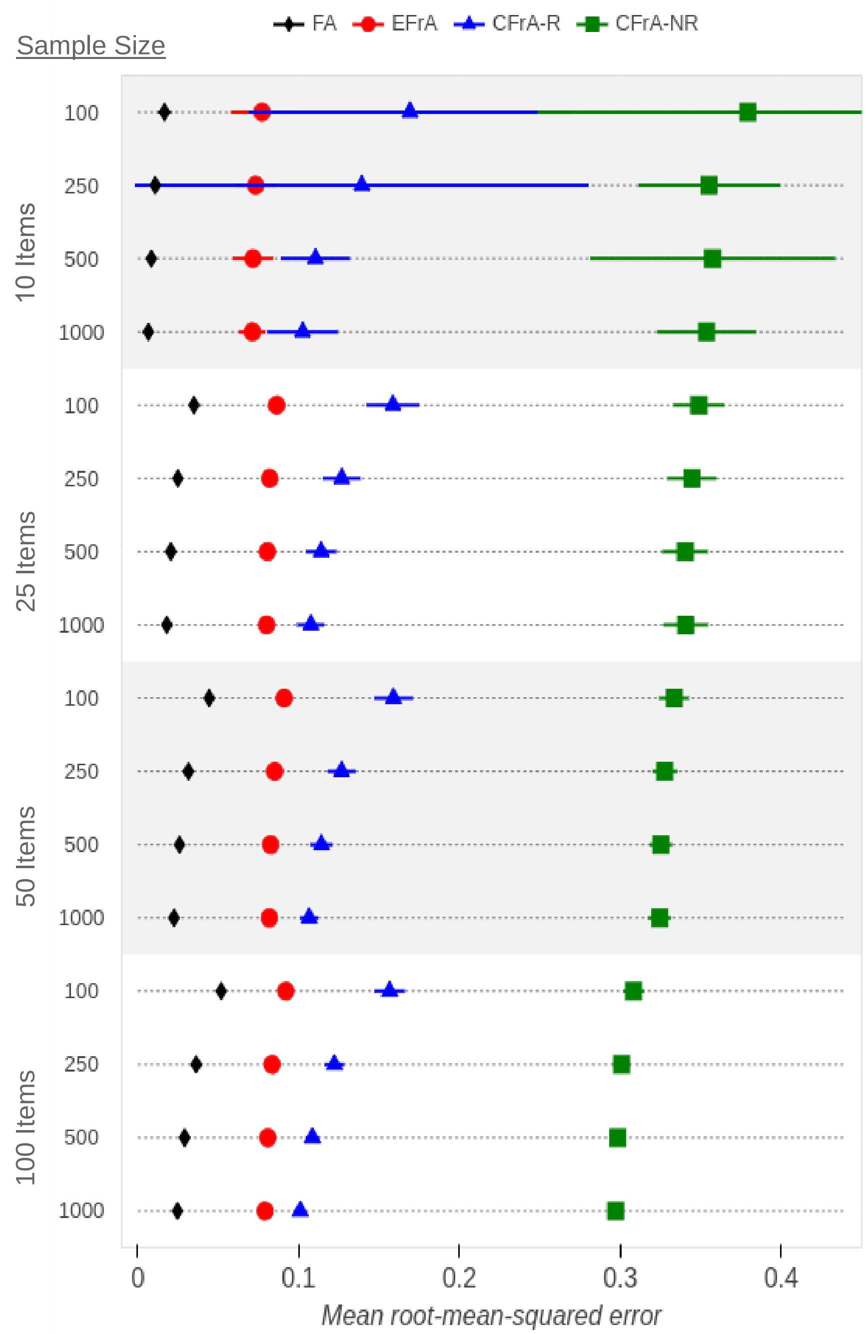

Figure 9 - Average Root-Mean-Squared Error between Polychoric Correlation and Reconstructed Correlation for 4 Factor Exploratory and Constrained Frame Analysis. *Note.* Solid horizontal lines are standard deviations.

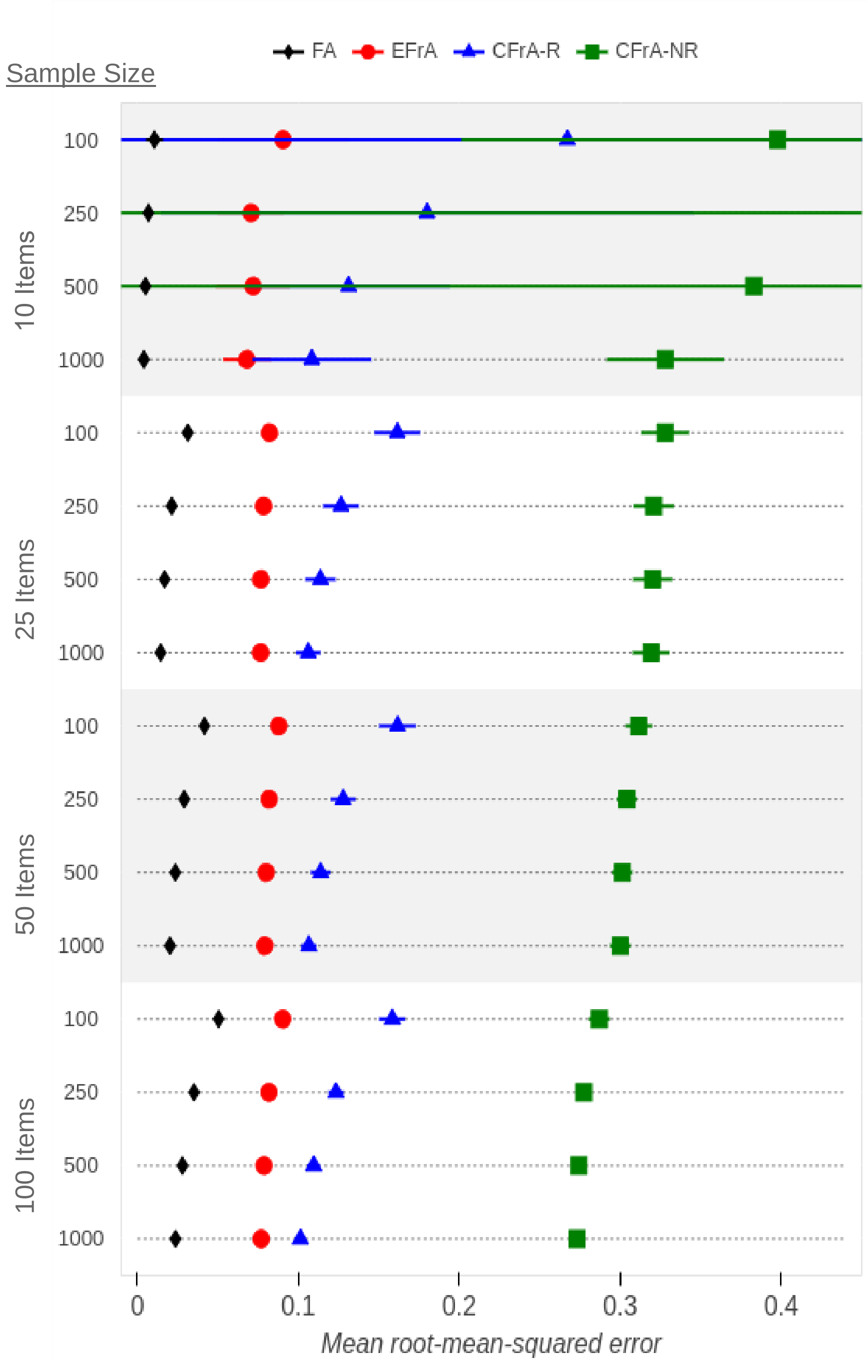

Figure 10 - Average Root-Mean-Squared Error between Polychoric Correlation and Reconstructed Correlation for 5 Factor Exploratory and Constrained Frame Analysis. *Note*. Solid horizontal lines are standard deviations.

1.4 Experiment 2: Five-Factor Personality Application

We performed the above simulation study to quantify frame analysis performance in controlled situations. To demonstrate the benefits of frame analysis in a research application, we applied exploratory, constrained, and partially constrained frame analysis to archival personality data. We chose personality data since it was the original impetus for developing frame analysis and allows frame analysis to demonstrate an alternative interpretation of the hierarchical and facet personality models (Costa & McCrae, 1995a; DeYoung et al., 2007).

1.4.1 Circumplex Personality Models

Proto-frame analytic concepts are found in Leary's (1957, p.62-67) two-factor and Hofstee et al.'s (1992) five-factor circumplex models of personality. Hofstee et al (1992) envisioned adjectives as mixtures of two primary traits, i.e., agreeableness and conscientiousness, resulting in 10 two-dimensional circumplexes. In their analysis, the two most significant non-zero loadings are kept after extraction and varimax rotation. These loadings locate each item in a two-dimensional plane partitioned into 12 sections of 30 degrees. Markey and Markey (2006) extend the circumplex model into three dimensions using latitude and longitude analogs to locate items.

Costa and McCrae (1995a) argued against the circumplex model on two grounds. First, there is poor representation of items in all circumplex segments (Costa & McCrae, 1995a). Second, it co-locates certain traits on the circumplex that represent different behaviors (Costa & McCrae, 1995a). We address the first complaint by noting the misplaced earnestness for a simple structure. Thurstone (1935, p. 156) suggested that researchers discard items that corrupt simple structure. This selection process has downstream effects since items selected for the inventory preference homogeneity over breadth (Little et al., 1999).

Furthermore, a vector space's lack of representation is not an argument against the existence of that vector space; mankind's representation in the universe is infinitesimally tiny, yet our existence is not in doubt. The co-location of items in Hofstee et al.'s circumplex personality model (1992) might be explained by the projecting of cross-loaded items into two dimensions. It is possible the appearance of co-location vanishes once the full dimensionality is considered. Our proposed frame analysis maintains this dimensionality and can be loosely interpreted as a higher dimensional analog to circumplex models.

Costa and McCrae (1995) favored an alternative to five-factor circumplex personality models. They fractionated the Big Five factors (consciousness, openness, agreeableness, extraversion, and neuroticism) into smaller facets "organized hierarchically" beneath the broad primary factors (DeYoung et al., 2007; Krueger et al. 2011; Soto & John, 2017; Lee & Ashton, 2018). An essential property of personality factor models that hypothesize facets is that facets do not increase dimensionality. A five-factor model with ten facets does not become a ten-factor model but instead divides a single factor into two or more sub-factors. DeYoung et al. (2007) and Soto and John (2017) verified a five-factor structure when all inventory items were subjected to factor analysis. Their results suggest facets are not pure measurements of their factors but instead represent compositions of primary factors; in other words, a frame analytic model.

Although facets represent redundant and not new factor information, we can use frame analysis to determine if the data support the number of hypothesized facets in hierarchical models: DeYoung et al. (2007) hypothesize 10 aspects (facets) in their Big Five Aspect Scale personality inventory. Soto and John (2017) hypothesize 15 facets in their Big Five Inventory 2. Both Costa and McCrae's NEO-PI-R (1992) and Johnson's (2014) IPIP-120 hypothesize 30 facets. Under a frame analysis, facets (and factors) are uniquely expressed frame vectors. Therefore, the number of hypothesized facets should be equivalent to the number of expressed frame vectors; for example, a 30-facet model should have exactly 30 expressed frame vectors.

1.4.2 Archival Personality Inventories

We used two archival personality data collections available at https://osf.io/tbmh5/. Johnson (2014) used the international personality item pool (IPIP) 300 to create a short form, the IPIP-120. This necessarily means that the IPIP-120 is a subset of the IPIP-300. Both inventories were administered online to self-selecting participants. For our analyses, we required participants to have indicated 1.) United States of America as their country of origin, 2.) their age as 18 or older, and 3.) answered all the items.

1.4.2.1 International Personality Item Pool – 300

The international personality item pool 300 is a set of 300 personality items designed to measure 30 facets of the broad five-factor model (Goldberg, 1999). It was intended to be a royalty-free version of the NEO-PI-R (Costa & McCrae, 1992). This scale consists of 300 items on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). We calculated Cronbach's alpha reliabilities for each of the five factors:

neuroticism (α=.96), extraversion (α=.94), openness (α=.91), agreeableness (α=.92), and conscientiousness (α=.95). The total number of collected surveys before the inclusion criteria was 307,313. After we applied the inclusion requirements, the total sample was 79,743 participants (40% male) with an age of *M*=27.2 (*SD*=9.70). Inclusion criterion 3, answering all items, was responsible for the largest drop in participants. Only 145,388 participants completed all items.

1.4.2.2 International Personality Item Pool – 120

International personality item pool - 120 (Johnson, 2014) is an open-source version of the NEO PI-R (Costa & McCrae, 1992). Johnson (2014) designed the inventory to measure personality on five broad factors with six facets each for 30 facets, four questions per facet. This scale consists of 120 items on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Example items include “I worry about things”, “I make friends easily”, and “I enjoy being reckless.” Cronbach’s alpha reliability for the five broad factors were neuroticism (α = .90), extroversion (α = .89), openness (α = .83), agreeableness (α = .86), and conscientiousness (α = .90). The total number of collected surveys before the inclusion criteria was 619,150. Applying the criteria resulted in 243,347 participants (41% male) with an age of *M*=26.6 (*SD*=9.97). Inclusion criterion 3, answering all items, was responsible for the largest drop in participants. Only 410,376 participants completed all items.

1.4.3 Metrics

1.4.3.1 Coverage

In addition to the two metrics discussed in Experiment 1, we define frame dictionary coverage as the ratio of expressed frame vectors to the total number of available frame vectors (i.e., the number of frame vectors in the frame dictionary).

1.4.4 Procedure

We applied exploratory frame analysis to the IPIP-300 personality inventory. We performed a polychoric correlation followed by principal axis factoring and factor loading extraction. Based on the scree test, we used five factors to extract the factor loadings. We comprised a frame dictionary of eight frame vectors as described in Appendix A to determine each item's expressed frame vector.

Next, we ran a constrained frame analysis using the frame vectors determined from the IPIP-300 exploratory frame analysis and applying them to the shared items on the IPIP-120 personality inventory. The root-mean-squared-error captured the error between the extracted loadings and reconstructed loadings.

Finally, we performed a partially constrained frame analysis using one-half of the frame vectors determined from the IPIP-NEO-300 exploratory frame analysis. We measured the average congruence between the unused frame vectors and the recovered frame vectors after alignment.

1.4.5 Results

The results of exploratory and (partially-) constrained frame analysis on archival personality data are shown in Table 6. Applying exploratory frame analysis to the IPIP-300 resulted in an RMSE consistent with the quantization error from the simulated results in Experiment 1. This result suggests the quantization error is independent of the data

Table 6 – Archival Personality Results for Exploratory, Constrained and Partially-Constrained Frame Analyses

Analysis	RMSE	Coverage	Congruence
Exploratory			
IPIP-300	0.052	19.9%	
Constrained			
IPIP-120	0.057		
Partially-Constrained			
Constrained Items	0.059		
Unconstrained Items	0.051		.971

Note. IPIP: International Personality Item Pool

source. The frame dictionary coverage for the 300 items was 171 out of 860 possible frame vectors or encompassed roughly 20% of the partitioned multi-dimensional space.

Applying constrained frame analysis to the IPIP-120 using the item-frame vector associations found from the IPIP-300 also resulted in an RMSE consistent with the replicating simulation results. This result suggests the IPIP-120 model was confirmed using the IPIP-300 frame vectors. Applying partially-constrained frame analysis to half of the IPIP-120 again found RMSEs consistent with that of simulation. The RMSE from the unconstrained items was lower than that from the constrained items which were consistent with expectations. We also observed high levels of average congruence between items common to both the unconstrained set and the IPIP-300 exploratory results.

1.4.6 Discussion

We applied exploratory, constrained, and partially constrained frame analysis to archival personality data with encouraging results. The computed RMSE on the IPIP-300 was similar to the RMSE found during the exploratory frame analysis simulation study suggesting quantization error was the dominant error and is invariant to sample size for large samples. The number of expressed frame vectors was modest relative to the possible number of vectors, indicating only about 20 percent of the five-dimensional personality space was filled by the items. However, the number of expressed frame vectors was more than half of the total number of items; this necessarily requires that some partitions in the five-dimensional vector space were only associated with one item. Furthermore, the IPIP-120 and IPIP-300 were designed to measure 30 facets of the broad five personality markers, but these 30 facets failed to cohere in frame analysis. Instead of 30 facets, exploratory frame analysis provides relative composite information at the item level. Items may group together, contingent on the frame dictionary quantization level, but it is not a requirement.

Additionally, using the frame vectors derived from the IPIP-300 to constrain the frame loadings on the IPIP-120 resulted in an RMSE similar to the quantization error found in the exploratory frame analysis. These results strongly suggest the same relative relationship between items and provide further validation for Johnson's (2014) IPIP-120 short form. Reversing this process also allows the researcher to predict a respondent's score on IPIP-300 items from the IPIP-120. This is possible since both the IPIP-120 and IPIP-300 inhabit the same multidimensional space and the two coordinate systems are aligned through the constrained frame analysis process.

We also demonstrated that partially constrained frame analysis extends beyond simulation and functions on real data. Both constrained and unconstrained (exploratory) RMSEs were similar to both the simulation and the previous personality results. The high average congruence suggests that if two inventories contained separate items, a direct comparison between these unmeasured items, linked by a small common subset, would be accurate.

One application we envision for partially constrained frame analysis is creating a canonical set of five-factor personality items. Researchers could append this small canonical set to orient their new construct or inventory and assess validity. This canonical set also defines the five factors in much the same way the International Earth Rotation and Reference Systems Service Reference Meridian defines 0° longitude. An absolute five-factor personality reference frame still allows for tailored personality inventories, but now a method to create a Rosetta stone between them is established. A future study will identify this canonical set from the IPIP-300 items and apply it to other five-factor personality inventories.

A significant limitation of using online self-selecting personality data is that there is no way to prevent participants from taking the inventory multiple times or taking both IPIP inventories. Johnson's website (http://www.personal.psu.edu/~j5j/IPIP/) provides successive links to the IPIP-120 and the IPIP-300, making it easy for participants to take both. Furthermore, easily shareable online links facilitate snowball sampling schemes. In concert, these issues might explain the similarity found between the two inventories and favorably bias the constrained frame analysis results. A properly controlled study is needed to replicate our findings.

1.5 Discussion

Frame analysis is a modern approach for investigating the relationships between latent variables. It relies on a dense representation of the *n*-dimensional vector space and obviates factor rotations to simple structure. This seemingly minor change provides several benefits over traditional factor analysis. First, frame analysis offers a straightforward interpretation of frame loadings as the square root of the communality and decoupled from the frame vectors. The output of frame analysis is the common variance in the item (frame loadings) and the percent variance common with other items (frame correlation matrix). Therefore, these metrics are relative to the items and are rotationally invariant up to the quantization error.

Second, each item in frame analysis is associated with *one* frame vector. The frame vector defines the item composition relative to the underlying latent variables. These associations afford item-level investigation. In addition, defining a fixed coordinate system requires only n linearly independent frame vectors or selecting the associated items. Third, the item-frame vector pairs make constrained frame analysis, the analog to confirmatory factor analysis, possible within the same algorithmic framework. Frame loadings and frame vectors are both revealed from the extracted loadings during the exploratory frame analysis process. In constrained frame analysis, the frame vectors are supplied, and only the frame loadings that minimize reconstruction error are sought. The estimated parameters in constrained frame analysis are equal to the number of items or the n-dimensional rotation angles: n (n-1) / 2. In a five-dimensional space, the estimated parameters are 10, considerably less than the number of items in many inventories.

Third, partially constrained frame analysis, an amalgam of exploratory and constrained analysis, allows direct expansion of untested or disparate items onto a set of known items. This analysis is helpful for validity considerations, to combat construct proliferation, and link inventories without requiring participants to take both inventories completely. We discussed how a canonical set of items that *anchor* factor orientations would benefit five-factor personality research. Using partially constrained frame analysis and a canonical set of personality items interspersed with the inventory under consideration would provide a mechanism for researchers to translate between their domain-specific personality inventories.

Another advantage frame analysis has is the inherent clustering of related items. Items that express on the same frame vector are defined to be identical. Counting the number of non-zero frame loadings associated with a frame vector informs which direction in the multidimensional space is or is not represented. Summing non-zero counts over neighboring frame vectors provides an efficient method to determine the number and breadth of item clusters.

Finally, frame analysis frees the researcher from setting a factor loading threshold for the purpose of relating items and factors. It does this by giving equal status to cross-loaded items through frame vector expressions. The implications of this minor change are potentially considerable in the construction of multi-factor inventories. Multi-factor inventories no longer need to consist of multiple single factor constructs. Researchers might desire a broad representation of the multidimensional factor space that provides greater flexibility in item creation and less sensitivity to missing items. This property is a novel development in the examination of latent variables.

1.5.1 Limitations and Considerations

A significant limitation in frame analysis is a missing goodness-of-fit metric for (partially-) constrained algorithms. We showed proof-of-concept for both methods, but we need to provide an acceptance criterion to garner confidence and adoption. Our initial thought is to co-opt a root-mean-squared error of approximation-like fit index from confirmatory factor analysis after removing the noise floor from exploratory frame analysis. Another possibility is using Akaike or Bayesian Information criteria in concert with a maximum-likelihood extraction approach. These metrics are only speculation, and we must perform a proper sensitivity analysis to quantify performance.

Closely related to a missing goodness-of-fit metric is frame analysis' lack of an inferential statistic. Traditional inferential metrics inversely scale with sample size; larger sample sizes result in smaller detectable differences. An inferential statistic could help guide a frame dictionary's appropriate quantization level. One possible approach is to use bootstrapping to build confidence intervals for the determined frame vectors.

Out of necessity, we used a legacy factor extraction algorithm, principal axis factoring, to apply frame analysis to the extracted loadings. A complete frame analysis bypasses this step and estimates the correlation or covariance matrix directly, which would require the development of frame analysis-specific extraction algorithms. These undeveloped algorithms' impacts on frame analysis are unknown. However, we speculate the implications would be minimal enough to warrant the continued use of the current factor analysis extraction algorithms.

In addition to factor analysis extraction algorithms, we co-opted other factor analysis methods in developing frame analysis. We also borrowed factor analytic methods to assess dimensionality: Cattell's (1966) scree test, Velicer's (1976) minimum average partial test, and Horn's (1965) parallel analysis. It is possible that under our new redundant paradigm, the appropriate number of factors to extract is not the same as factor analysis. The impact of missing data on the identification and selection of frame vectors also remains unknown. We did not simulate any conditions with missing data, and we also excluded personality assessments that had even one missing item.

In using frame analysis, the researcher must decide the frame dictionary's quantization level. We demonstrated frame analysis using a frame dictionary with a .925 correlation sensitivity. Ideally, measurement reliability and item resolution should guide this decision. It seems implausible that such fine resolution is inherent in self- or observer-report survey data. Furthermore, there is no theoretical justification to create frame dictionaries that are nominally uniform and isotropic in factor space; our decision to apply these constraints embeds convergent and discriminant validity into the measurement space directly. In defining a uniform correlational lower bound, we impose the same similarity criteria on all frame vectors. However, it is possible that different quadrants in the multidimensional space afford different similarity criteria. The selection of the optimal frame dictionary properties remains an area for future investigation.

Finally, the impacts of misspecification in either dimensionality or frame vector expression are unknown. Hypothesizing a dimensionality less than the true dimensionality results in a projection of the higher space into the lower one. How this error propagates across (partially-) constrained frame analyses remains to be determined. Similarly, misspecified frame vectors also impact (partially-) constrained frame analyses results. Under such conditions, frame analysis might fail to provide the positive results found in the present analysis.

1.5.2 Future Applications

1.5.2.1 Nested Vector Spaces

Nested vector spaces are lower-dimensional vector spaces wholly contained within higher-dimensional vector spaces. Figure 11 shows an example of a hypothetical three-factor vector space. The two-factor solution represented by the blue plane captures factors A and B, while factor C is excised. This paradigm is easy to visualize and a common way to imagine factor reduction. However, this is not the only plausible two-factor solution. Another possible solution is given by the red semi-transparent plane. All three factors contribute to this two-factor solution but degenerate due to their vector space orientation.

In the context of personality psychology, this property could couch Eysenck's (Eysenck & Eysenck, 1975) three-factor model as nested within a degenerate five-factor model. Costa and McCrae (1995b, Table 3) provide some evidence in support of this hypothesis. The plethora of five-factor models might be degenerate nestings of Lee and Ashton's (2018) six-factor HEXACO model. Using partially constrained frame analysis, answering these questions is now possible

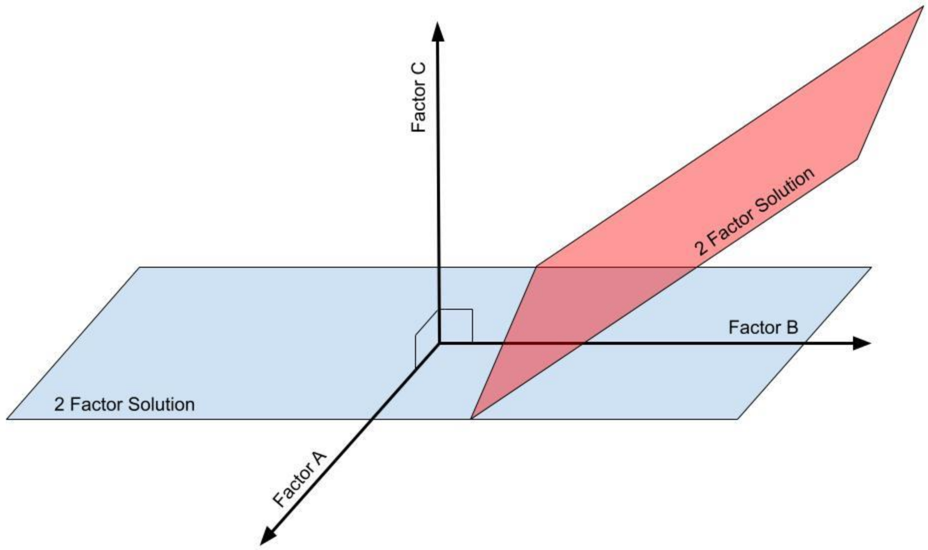

Figure 11- Diagram of Nested Factor Spaces. Two possible two factor solutions in a three factor space.

1.5.2.2 Hierarchical and Bifactor Models

In traditional factor analysis, correlated factors are used to hypothesize hierarchical and bifactor models. Ashton et al. (2009) provided evidence that hierarchical models can also be interpreted as blended orthogonal factors, a frame analysis model. Jennrich and Bentler (2011) developed an exploratory bifactor procedure based on a structured factor loading matrix. The structure they suggest is equivalent to selecting frame vectors with desirable properties. If hierarchical and bifactor models are special cases of frame analysis, then frame analysis is the more economical solution and ought to be preferred.

1.5.3 Conclusion

We introduced and developed a new statistical method for exploring latent structures: Frame Analysis. Frame analysis celebrates cross-loaded items while eschewing rotation to simple structure. We demonstrated that a unified framework for both exploratory and constrained frame analysis emerged by adopting a heterogeneous approach to the item-factor relationship. Finally, we showed frame analysis's efficacy on simulated and collected data with positive results. We envision frame analysis as a drop-in replacement for both exploratory and confirmatory factor analysis that simplifies the researcher's modus operandi when investigating latent structures.

APPENDIX A. FRAME DICTIONARY CREATION

The frame dictionary is a set of frame vectors that partition the *n*-dimensional vector space. This dictionary defines all possible relationships between items in frame analysis. Here, we outline a method to create a quasi-isotropic frame dictionary that approximately creates sections of equal surface area on a hypersphere. This property results in a uniform relative quantization error. Many other criteria can be used to create different frame dictionaries; a bifactor frame dictionary is one example of alternative criteria. Ultimately, the researcher should be the arbiter who defines domain-specific properties tailored to their needs.

A.1 Hyperspherical Vectors

It is necessary to define a hyperspherical vector in *n*-dimensions to populate the frame dictionary; n-1 angles define these vectors. If we allow for hemispherical symmetry, all angles, θ_n, are valid over the semi-open domain $[0, \pi)$. Blumenson (1960) outlines the analytic expression to compute these vectors:

$$\vec{v}_n = \begin{bmatrix} \cos(\theta_1) \\ \sin(\theta_1)\cos(\theta_2) \\ \sin(\theta_1)\sin(\theta_2)\cos(\theta_3) \\ \vdots \\ \sin(\theta_1)\cdots\sin(\theta_{n-2})\cos(\theta_{n-1}) \\ \sin(\theta_1)\cdots\sin(\theta_{n-2})\sin(\theta_{n-1}) \end{bmatrix} \tag{A1}$$

A.2 Frame Dictionary

One parameter, the number of partitions in any one canonical hyperplane, defines the frame dictionary. This parameter is then applied to every other canonical hyperplane, and we note an even number of partitions results in explicit orthogonal frame vectors in the canonical bases. Due to surface curvature, we vary the number of partitions between hyperplanes to maintain an approximately uniform arc length. The number of partitions is proportional to the product of $\sin(\theta_1)$ to $\sin(\theta_{n-2})$. Since the computed number of partitions is often not an integer, we truncate the computed number. Truncating is a conservative approach that results in slightly larger partitions away from the canonical hyperplanes but prevents over-representation at the poles. The frame vectors are computed by Equation A1 and collated into a matrix whose row space represents a frame in the desired dimension. A Python function that computes frame dictionaries is included in the supplemental material. Table A1 shows the number of frame vectors as a function of dimensionality and the number of partitions. For high dimensionality, the frame dictionary becomes large but can be computed offline for easy retrieval.

Table A1 – Number of Frame Vectors in a Frame Dictionary as a Function of Dimensionality and Number of Partitions

Dimensionality	6 Partitions	8 Partitions
2	6	8
3	22	39
4	73	184
5	256	860
6	875	3,934
7	2,666	17,746

APPENDIX B. FRAME LOADING CALCULATION

B.1 Exploratory Frame Analysis

Since frame analysis quasi-isotropically covers the multi-dimensional space, computing the frame loadings in exploratory frame analysis is straightforward. The first step entails calculating the extracted loadings using any factor analysis algorithm, Step 1 in Table B1. Next, the vector dot product between a frame vector and the loadings matrix computes the items' frame loading on the frame vector, Step 2 in Table B1. We use the frame dictionary matrix to calculate all possible frame loadings on all items simultaneously. Lastly, selecting the largest absolute frame loading for each item maximizes the correspondence between the extracted loadings from factor analysis and the quantization scheme used in frame analysis, Step 3a in Table B1. Alternatively, we can refine the solution by applying an *n*-dimensional orthogonal rotation to search for the least-squares answer between the extracted loadings and reconstructed loadings, Step 3b in Table B1. This alternative step, Step 3b, first rotates the frame dictionary and computes a candidate set of frame loadings. These candidate frame loadings are then used to compute the reconstructed loadings. Lastly, the squared error between the reconstructed loadings and extracted loadings is computed to determine if the applied rotation results in minimal squared error.

B.2 Constrained Frame Analysis

In constrained frame analysis, a previous exploratory analysis supplies the item-frame vector associations. We require the constrained frame vectors to correspond with the items such that frame vector 1 defines item 1 and frame vector n defines item n. The task to accomplish is determining the frame coefficients that minimize the error between the reconstructed loadings and extracted loadings. This task is further complicated due to a possible global rotation between the extracted and constrained coordinate systems. Computing the constrained frame loadings is similar to the method outlined for exploratory frame analysis with two crucial differences.

First, an *n*-dimensional rotation must be found that properly orients the two coordinate systems. We seed this rotation with a Procrustes rotation (Schönemann, 1966) to minimize the search space, Step 2 in Table B2. Next, we refine the rotation by minimizing the squared error between the extracted loadings and reconstructed loadings. This step, Step 3 in Table B2, is similar to Step 3b in Table B1. Several essential differences include using the supplied constrained frame vector matrix, **C**, instead of a frame dictionary, **D**, and selecting the frame loading associated with the constrained frame vector instead of keeping the largest absolute frame loading. The number of parameters estimated equals the number of angles required in the rotation matrix in our algorithm. Finally, Step 4 in Table B2, we compute the constrained frame loadings by using the set of angles found in Step 3.

B.3 Partially-Constrained Frame Analysis

Partially-constrained frame analysis applies the constrained frame analysis algorithm to the known items and uses the determined rotation matrix to apply the exploratory analysis algorithm to the unknown items. We do not require the two frame dictionaries to be common between the known and unknown items, but this will most likely be the case in practice.

B.4 Rotations

Orthogonal rotations in two and three dimensions are familiar to most researchers, and it is evident that one and three angles define two- and three-dimensional rotations, respectively. Rotations in higher dimensions are less obvious, and we use Givens rotations (Ford, 2015) to apply successive rotations to hyperplanes. This rotation formulation requires estimating $n(n-1)/2$ angles.

Table B1 – Frame Loading Algorithm Outline for Exploratory Frame Analysis

---

Step 1: Extract loadings using factor analysis algorithm, **L**

Step 2: Compute all possible frame loadings, $\mathbf{F}_{\mathbf{all}}$, from the frame dictionary, **D**, and extracted loadings

$$\mathbf{F^T_{all} = D} \times \mathbf{L^T}$$

Step 3a: Select the largest absolute frame loading for each item

$$\mathbf{F^T_{final} =} \textbf{\textit{column max}} \mathbf{(F^T_{all})}$$

Step 3b: (Optional) Perform *n*-dimensional rotation, **R**, by estimating *n(n-1)/2* angles, **φ**, that minimize the squared error between extracted loadings and reconstructed loadings

$$\boldsymbol{\varphi^* =} \textbf{\textit{min}}\ \boldsymbol{\varphi} \sum[\textbf{\textit{column max}}(\mathbf{D} \times \mathbf{R}(\boldsymbol{\varphi}) \times \mathbf{L^T})^\mathbf{T} \times \mathbf{D} \times \mathbf{R}(\boldsymbol{\varphi}) - \mathbf{L}]^2$$

$$\mathbf{F^T_{final} =} \textbf{\textit{column max}}(\mathbf{D} \times \mathbf{R}(\boldsymbol{\varphi^*}) \times \mathbf{L^T})$$

Note: column max zeros out all matrix elements in a column not equal to the maximum absolute loading in that column.

Extracted Loading Matrix, **L,** [i x f] : [*number of items* by *number of factors]*

Frame Dictionary Matrix, **D**, [v x f] : [*number of frame vectors* by *number of factors*]

Frame Loadings Sparse Matrix, **F**, [i x v] : [*number of items* by *number of frame vectors*]

Rotation Matrix, **R**, [f x f] : [*number of factors* by *number of factors*]

Table B2 – Frame Loading Algorithm Outline for Constrained Frame Analysis

---

Step 1: Extract loadings using factor analysis algorithm, **L**

Step 2: With constrained frame vector matrix, **C**, compute the Procrustes rotation, $\mathbf{R_p}$, using the singular value decomposition algorithm

$$\mathbf{U, S, V} = \boldsymbol{SVD}\ [\ \mathbf{C^T} \times \mathbf{L}]$$
$$\mathbf{R_p} = \mathbf{U} \times \mathbf{V^T}$$

Step 3: Perform *n*-dimensional rotation, **R**, by estimating *n(n-1)/2* angles, **φ**, that minimize the squared error between extracted loadings and reconstructed loadings

$$\boldsymbol{\varphi^*} = \boldsymbol{min}\ \varphi \sum[\boldsymbol{ddiag}[\mathbf{C} \times \mathbf{R_p} \times \mathbf{R}(\boldsymbol{\varphi}) \times \mathbf{L^T}]\ \times \mathbf{C} \times \mathbf{R_p} \times \mathbf{R}(\boldsymbol{\varphi}) - \mathbf{L}]^2$$

Step 4: Extract the constrained frame loadings, $\mathbf{F_{const}}$

$$\mathbf{F^T_{const}} = \boldsymbol{ddiag}[\mathbf{C} \times \mathbf{R_p} \times \mathbf{R}(\boldsymbol{\varphi^*}) \times \mathbf{L^T}]$$

Note: ddiag zeros out all matrix elements not on the main diagonal

Extracted Loading Matrix, **L**, [i x f] : [*number of items* by *number of factors*]

Constrained Frame Vectors, **C**, [v x f] : [*number of frame vectors* by *number of factors*]

Frame Loadings Sparse Matrix, **F**, [i x v] : [*number of items* by *number of frame vectors*]

Rotation Matrix, **R**, [f x f] : [*number of factors* by *number of factors*]

REFERENCES

Ashton, M. C., Lee, K., Goldberg, L. R., & Vries, R. E. (2009). Higher Order Factors of Personality: Do They Exist? *Personality and Social Psychology Review, 13*(2), 79-91. doi:10.1177/1088868309338467

Blumenson, L. E. (1960). A Derivation of n-Dimensional Spherical Coordinates. *The American Mathematical Monthly*, *67*(1), 63. https://doi.org/10.2307/2308932

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Browne, M. W. (2001). An Overview of Analytic Rotation in Exploratory Factor Analysis. *Multivariate Behavioral Research, 36*(1), 111-150. doi:10.1207/s15327906mbr3601_05

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources

Costa, P. T., & Mccrae, R. R. (1995a). Domains and Facets: Hierarchical Personality Assessment Using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*(1), 21-50. doi:10.1207/s15327752jpa6401_2

Costa, P. T., & McCrae, R. R. (1995b). Primary traits of Eysenck's P-E-N system: Three- and five-factor solutions. *Journal of Personality and Social Psychology*, *69*(2), 308–317. https://doi.org/10.1037/0022-3514.69.2.308

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, *113*(3), 492–511. https://doi.org/10.1037/pspp0000102

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93(5), 880–896. doi: 10.1037/0022-3514.93.5.880

DeYoung, C. G., Peterson, J. B., Séguin, J. R., & Tremblay, R. E. (2008). Externalizing behavior and the higher order factors of the Big Five. *Journal of Abnormal Psychology, 117*(4), 947-953. doi:10.1037/a0013742

Duckworth, A. L., & Quinn, P. D. (2009). Development and Validation of the Short Grit Scale (Grit–S). *Journal of Personality Assessment*, *91*(2), 166–174. https://doi.org/10.1080/00223890802634290

Eysenck, H. J., & Eysenck, S. B. G. (1975). Manual of the Eysenck Personality Questionnaire. San Diego, CA: EdITS

Ford, W. (2015). Implementing the QR Decomposition. In *Numerical Linear Algebra with Applications* (pp. 351–378). Elsevier. https://doi.org/10.1016/b978-0-12-394435-1.00017-x

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.). *Personality Psychology in Europe* (Vol. 7, pp. 7-28)

Harman, H. A. (1967). *Modern Factor Analysis*. Chicago, IL: The University of Chicago Press.

Hofstee, W. K., Raad, B. D., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, *63*(1), 146–163. https://doi.org/10.1037/0022-3514.63.1.146

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2, 41–54.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185. doi:10.1007/bf02289447

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jennrich, R. I. (2007). Rotation methods, algorithms, and standard errors. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 315-335). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory Bi-Factor Analysis. *Psychometrika, 76*(4), 537-549. doi:10.1007/s11336-011-9218-4

John, O. P., Robins, R. W., & Pervin, L. A. (2008). *Handbook of personality theory and research* (3rd ed.). New York, NY: Guilford Press.

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, *51*, 78–89. doi: 10.1016/j.jrp.2014.05.003

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*(2), 183-202. doi:10.1007/bf02289343

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2011). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42*(9), 1879-1890. doi:10.1017/s0033291711002674

Kovacevic, J., & Chebira, A. (2007). Life Beyond Bases: The Advent of Frames (Part I). *IEEE Signal Processing Magazine, 24*(4), 86-104. doi:10.1109/msp.2007.4286567

Kovacevic, J., & Chebira, A. (2007). Life Beyond Bases: The Advent of Frames (Part II). *IEEE Signal Processing Magazine*, *24*(5), 115-125. doi:10.1109/msp.2007.904809

Leary, T. (1957). *Interpersonal diagnosis of personality: A functional theory and methodology for personality evaluation*. Wipf and Stock Publishers.

Lee, K., & Ashton, M. C. (2018). Psychometric Properties of the HEXACO-100. *Assessment*, *25*(5), 543-556. doi:10.1177/1073191116659134

Ludeke, S. G., Bainbridge, T. F., Liu, J., Zhao, K., Smillie, L. D., & Zettler, I. (2019). Using the Big Five Aspect Scales to translate between the HEXACO and Big Five personality models. *Journal of Personality*, *87*(5), 1025–1038. https://doi.org/10.1111/jopy.12453

Markey, P. M., & Markey, C. N. (2006). A spherical conceptualization of personality traits. *European Journal of Personality, 20*(3), 169-193. doi:10.1002/per.582

Morgenshtern, V.I., & Bölcskei, H. (2011). A Short Course on Frame Theory. *ArXiv*, https://www.arxiv.org/abs/1104.4300

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460. https://doi.org/10.1007/bf02296207

Shaffer, J. A., Degeest, D., & Li, A. (2015). Tackling the Problem of Construct Proliferation. *Organizational Research Methods*, *19*(1), 80–110. doi: 10.1177/1094428115598239

Schönemann, P. H. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika, 31*(1), 1–10. https://doi.org/10.1007/bf02289451

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology, 15*(2), 201. doi:10.2307/1412107

Spearman, C. (1927). *The Abilities of man*. London: Macmillan.

Thurstone, L.L. (1935). *The Vectors of Mind, Multiple Factor Analysis for the Isolation of Primary Traits*. Chicago, IL: The University of Chicago Press.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321-327. doi:10.1007/bf02293557

Yong, A. G., & Pearce, S. (2013). A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. https://doi.org/10.20982/tqmp.09.2.p079