

**Paper 352-2011****Scale Measurement: Comparing Factor Analysis and Variable Clustering****Albert Liao, Teck Kiang Tan and Angeline Khoo****Psychological Studies, and Centre for Research in Pedagogy and Practice****Nanyang Technological University, Singapore****ABSTRACT**

In the social sciences, factor analytical methods are commonly used in the scale measurement in examining the structure of scales. This paper provides an alternative way of examining the structure of scale measurement using variable cluster analysis. Three analytical methods are illustrated in this paper, namely exploratory factor analysis, confirmatory factor analysis and variable clustering analysis. As variable cluster analysis is rarely used and not widely discussed in the literature, the advantages, benefits and limitations of using this method are discussed. Comparison of variable clustering to the factor analytical approach is illustrated using the personal strength scale. PROC VARCLUS, PROC FACTOR, and PROC CALIS are the three SAS procedures for these three analytical methods. The syntaxes of the three procedures are given and the results are interpreted and compared.

**INTRODUCTION**

In measurement literature, exploratory and confirmatory factor analyses are two common standard statistical tools for developing measurement scales. SAS<sup>®</sup> PROC FACTOR and PROC CALIS are the two main procedures cater for these two statistical approaches. The SAS<sup>®</sup> VARCLUS procedure, borrows from the factor analysis literature, combined with hierarchical clustering, gives an alternative approach to the measurement procedure. The advantages of the PROC VARCLUS approach over the traditional factor analysis approach are discussed in details in the paper.

**Scale Measurement**

Social science research generally creates scales as constructs by using several items in a questionnaire to capture the various aspects of a construct in order to detain its essence. For example, an educational test might contain 50 items that intends to measure student's academic performance of a subject domain. Psychological measurement almost always relies on using items to create psychological scales. For instance, an inventory of depression that makes up of 40 items to create the depression inventory. This paper uses the personal strength scale proposed by Liao (2007) and Liao et al (2010) as an example to illustrate the scale measurement of the personal strength of students. This scale is consistent with the positive psychology movement that focusing on positive qualities of students, an instrument intends to measure the well-being and development of students and it is of wide practical application (Liao et al 2010). The current personal strength inventory is a shorter version of Liao et al (2010) which makes up of 18 items subdivided into four subscales. The four subscales are emotion awareness (5 items), social competence (4 items), empathy (4 items), and goal setting (5 items).

**Data Source and Sample**

The data are obtained from a three-year longitudinal project examining the effects of video gaming on children and adolescents' development.

## Exploratory Factor Analysis

Although there are many different kinds of factor analysis available in the literature, all have the common goal of uncovering a larger set of variables to form a reduced set of the underlying factors or constructs that could be used to approximate these observed variables. These factors in measurement are called scales and the variables to form the scale are the items. In scale measurement, one of the major objectives is to determine whether the items could form into the scale by examining certain statistical properties using exploratory factor analysis. We shall discuss these properties in detail shortly. The literature distinguished between principal component analysis and factor analysis as they are mathematically different but the approaches and procedures of examining the statistical properties to form the structure of scale are similar. As such, we shall concentrate on principal component analysis (PCA) in this paper.

The SAS code to run PCA is as follows:

```
Proc Factor Data = A;  
  Var EmotionAwareness1-EmotionAwareness5  
      SocialCompetence1-SocialCompetence4  
      Empathy1-Empathy4  
      GoalSetting1-GoalSetting5;  
Run;
```

The PROC FACTOR invokes the factor analysis procedure. By default, principal component analysis is the method for analysis. "DATA = A" specifies the dataset A is used for the analysis. The VAR statement specifies the list of variables for carrying out exploratory factor analysis. The above statement specifies 18 items to be included for analysis.

## How Many Factors?

As exploratory factor analysis seeks to reduce a set of items to a scale or a smaller set of composite scales, the first question we would want to know is the number of the factors that would turn out in the analysis. Conceptually, we should establish the relationship of the items to the construct that they belong to so that we have expectation on the number of factors beforehand. Since there are 4 subscales for the personal strength scale, we would expect the items under the 4 subscale would fall under their respective subscales. That is, the 5 items of emotional awareness would fall under a factor, the 4 items of social competence would fall under a second factor and so on.

Although researchers would normally establish the number of factors beforehand, empirical data may not turn out as expected. Sometimes, the number of items of a new scale could be out of expectation, and under different research setting, the number of factors could be unanticipated as well. In these circumstances or even without these conditionings, we would like to use certain statistical approaches to inform us about the statistical outcomes and to confirm the number of factors as we expected. The measurement literature has proposed quite a number of approaches to determine the appropriate number of factors. We shall focus four of the commonly used methods, namely Kaiser's criterion, the scree test, the proportion of variance explained and parallel analysis.

## Kaiser's Criterion

As the number of factors to retain is critical to produce insight into the structure of latent variables, this gives rise to Kaiser (1960) to propose the "eigenvalue greater than one" rule. Because in principal component analysis, the sum of the eigenvalues is equal to the number of variables, the Kaiser's criterion considers the number of factors to be retained is with eigenvalues that is greater than one. The fundamental idea behind the Kaiser's criterion is that

choosing the number of factors that are above the “average” proportion of variance (which normally considered as 1) is justifiable as the information content is greater than the item level of information with value of eigenvalue one. As there are 18 items for this study, the total eigenvalues is 18. The initial factor solution as shown below lists the eigenvalues from largest to smallest. The number of factors retained is four since the first four factors are with eigenvalues greater than one. For most software, including SAS PROC FACTOR, the number of factors is set by Kaiser’s criterion as the default. So without further specification, the default output would be based on the Kaiser’s criterion. It could be used as a starting point in examining the structure of measurement scale. Kaiser’s criterion is also known as “Kaiser” rule, the “Kaiser-Guttman” rule, and the “K1” rule.

The FACTOR Procedure  
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 18 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	6.78261098	5.36120832	0.3768	0.3768
2	1.42140266	0.19044846	0.0790	0.4558
3	1.23095420	0.15950344	0.0684	0.5242
4	1.07145076	0.29669700	0.0595	0.5837
5	0.77475376	0.04391777	0.0430	0.6267
6	0.73083599	0.04890548	0.0406	0.6673
7	0.68193050	0.06406369	0.0379	0.7052
8	0.61786682	0.03918086	0.0343	0.7395
9	0.57868595	0.02424511	0.0321	0.7717
10	0.55444084	0.02974609	0.0308	0.8025
11	0.52469475	0.03272841	0.0291	0.8316
12	0.49196634	0.02108221	0.0273	0.8590
13	0.47088412	0.01129991	0.0262	0.8851
14	0.45958422	0.02452679	0.0255	0.9107
15	0.43505742	0.01689242	0.0242	0.9348
16	0.41816501	0.01964147	0.0232	0.9581
17	0.39852354	0.04233140	0.0221	0.9802
18	0.35619213		0.0198	1.0000

4 factors will be retained by the MINEIGEN criterion.

## Proportion of Variance Explained

As the result of the factor analysis reduces a larger number of items to a smaller set of subscales, we would expect the variance of a scale is less than the original items to form the scale. The question then is what do we expect the minimum variances we would like to accept for the scale to explain the items. As more factors are retained, the variance percentages would be higher so we expect a minimum variance with an appropriate number of factors. There is no standard as to the minimum percentage of variance explained for a given scale. However, it should not be too low (say lower than 50% for social science data) as that would mean that the scale has lost much of the information of the items nor too high (close to 100%) as that would mean that we have selected the number of factors close to the number of items indicating all the items are separate factors. Generally, we would first examine the variance explained in conjunction with the expected number of factors, and look at the structure of factor loadings and vary the number of factors up or down to see whether theoretically the results make sense or not.

## Scree Plot

Cattell’s (1966) suggests a graphical approach to determine the number of factors to retain which is called the scree plot. A scree plot shows a line plot by specifying the variance explained by each factor on the vertical axis and the number of factors on the horizontal axis. The number of factors is determined by inspecting the shape of scree plot where there is an “elbow” in the plot. This is the point where factors explain relatively little additional variance after that point. In other words, below that point, the eigenvalues remain close to each other and reflect a straight line. Above that point, the variances are substantially more as there are big drops in the size of the eigenvalues.

The rationale of the scree plot is that we are keeping the factors that ensure the contribution of the variances are kept before the diminishing returns take place as established by the elbow in the plot. However, there are limitations of this approach as well. Sometimes, the scree plot may not show obvious elbow pattern, and where there are a common second high-order factor that makes up of several subscales, the number of factors may or may not show a one common factor solution. The left hand side scree plot below shows the one common factor solution of the 18 items personal strength scale. The first factor eigenvalue is high with value 6.78. From second factor onwards, the eigenvalues are relatively small. We could draw a straight line by joining these eigenvalues with a gentle slope and conclude a one factor solution. The scree plot on the right hand however shows that it is possible to view it in another way that there are in fact two elbows or two drops and hence conclude a four factors solution. The scree plot is an optional for PROC FACTOR, we have to specify SCREE after the DATA= statement to generate the plot.

```
Proc Factor Data = Video Scree;
```

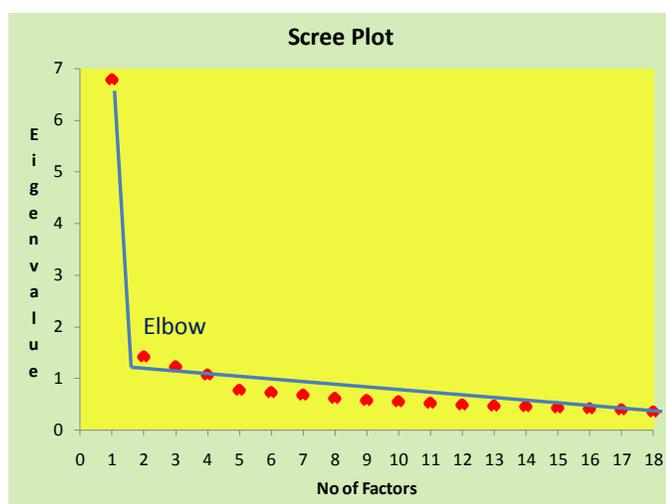


Figure 1 Scree Plot – One Factor Solution

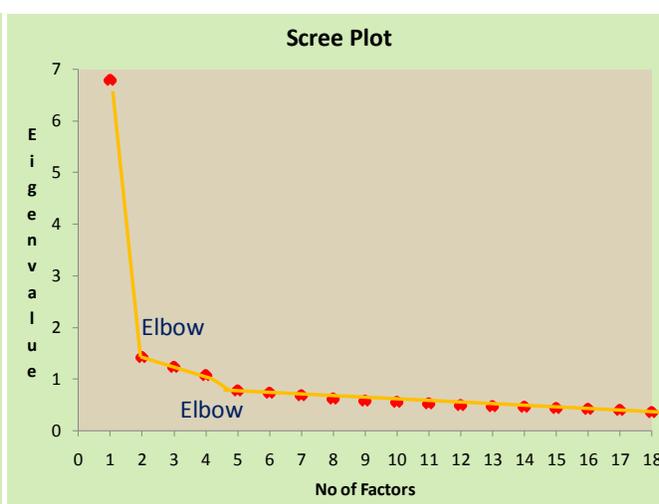
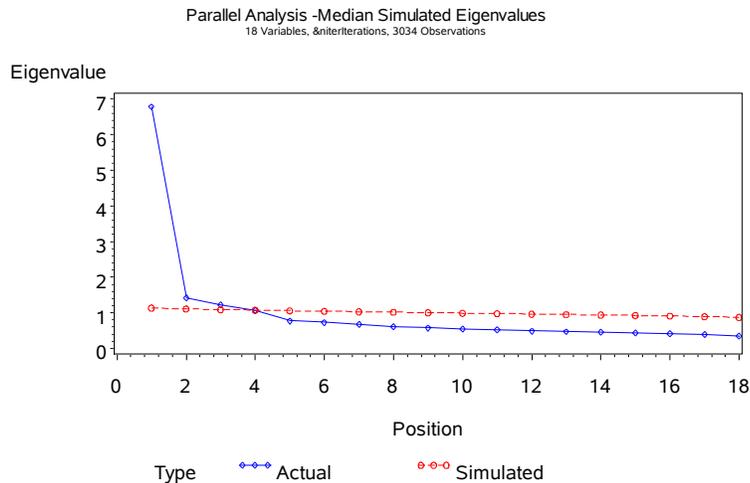


Figure 2 Scree Plot – Four Factors Solution

### Parallel Analysis

Kaiser's (1960) criterion was heavily criticized in the literature. Cliff (1988) provides evidence that Kaiser's criterion of minimum eigenvalue greater than one criterion is mathematically incorrect. Horn's (1965) parallel analysis, an alternative solution, is increasing in usage. There is a growing body of review papers and simulation studies that advocate Horn's parallel analysis (Hayton, Allen, and Scarpello, 2004; Patil, Singh, Mishra, and Todd Donovan, 2008). Practical experiences have also shown that the Kaiser's criterion may not be a good indicator as it generally overstates or understates the number of expected factors. The main idea of parallel analysis is to locate the number of factors that account for more variance than the factors derived from random data. The values of eigenvalues expected by chance are compared with values of observed eigenvalues to determine the justifiable number of factors. Factors with eigenvalues greater than those expected by chance are extracted. We use Kabacoff's (1990) parallel macro to generate parallel analysis with 1000 iterations with median eigenvalues. The parallel analysis shows a 4-factor solution.



In summary, the personal strength scale shows a four-factor solution according to the Kaiser's criterion and parallel analysis. The percentage explained for a four-factor solution is about 58.37%. The scree plot suggests possibilities of a general second-order scale for all the 18 items. This gives rise to the question that whether we should examine further for second-order factor models.

## Rotation

By default, PROC FACTOR produces an initial factor solution. The initial factor solution displays the first set of factor loadings output. However, this set of initial factor solution generally does not produce results that are easily interpretable as the items may not fall nicely to the scale they belong to. The initial factor solution output of the personal strength scale below shows that the factor loadings are high on the first factor but relatively low for the rest of the factors. Mathematically, there are infinite factor solutions for exploratory factor analysis. As such, we could use a method called rotation to change the values of the factor loadings that could possibly give us a solution that is much easier to interpret while retain the same amount of variance. That is, we could generate a factor solution that gives us a simple structure. A simple structure is a factor loading solution that one can easily assigned the items to scale with high factor loadings that close to one or minus one and those items not belong to the scale with low or close to zero loadings. There are dozens of rotation methods available in the literature. They could be generally divided into two main types of rotation methods: the oblique and the orthogonal. The former assumes factors are correlated while the latter assumes that the correlations between factors are zero. In the FACTOR procedure, the orthogonal rotation methods including varimax, quartimax, equamax, and parsimax while the oblique rotation methods include promax, the Harris-Kaiser case II orthoblique rotation, and the Procrustes method. We shall show the output of varimax and promax, the two commonly used rotation methods of the orthogonal and oblique rotations respectively. The rotated solutions using varmax and promax are shown below with clear structure for the four subscales. The specification of the keywords VARIMAX and PROMAX have to be inserted after the ROTATE= statement in order to output these two rotation methods.

```
Proc Factor Data = Video Scree Rotate=Promax;
  Var EmotionAwareness1-EmotionAwareness5
      SocialCompetence1-SocialCompetence4
      Empathy1-Empathy4
      GoalSetting1-GoalSetting5;
Run;
```

The FACTOR Procedure  
Initial Factor Method: Principal Components

## Factor Pattern

	Factor1	Factor2	Factor3	Factor4
EmotionAwareness1	0.54273	0.53627	0.15121	0.02601
EmotionAwareness2	0.61884	0.46706	0.11286	-0.00191
EmotionAwareness3	0.63497	0.39372	0.06324	-0.00441
EmotionAwareness4	0.62694	0.30567	0.14095	-0.19568
EmotionAwareness5	0.64260	0.32998	-0.10709	0.08321
SocialCompetence1	0.65426	-0.06255	0.08908	-0.45501
SocialCompetence2	0.62264	-0.17887	-0.04230	-0.53466
SocialCompetence3	0.66159	-0.26766	-0.04922	-0.34256
SocialCompetence4	0.68740	-0.29497	-0.13685	-0.18181
Empathy1	0.59850	-0.09628	-0.51896	0.19222
Empathy2	0.64904	0.15121	-0.44795	0.16594
Empathy3	0.60874	-0.14418	-0.44788	0.26751
Empathy4	0.55190	0.01505	-0.05609	0.07096
GoalSetting1	0.58305	-0.15182	0.46307	0.20643
GoalSetting2	0.56107	-0.18561	0.39284	0.28193
GoalSetting3	0.58721	-0.26025	0.12787	0.12545
GoalSetting4	0.56240	-0.40738	0.14953	0.15120
GoalSetting5	0.63203	-0.13939	0.21899	0.28366

## Varimax Rotation Output

## Rotated Factor Pattern

	Factor1	Factor2	Factor3	Factor4
EmotionAwareness1	0.75878	0.13950	0.05835	0.08408
EmotionAwareness2	0.73592	0.17372	0.14303	0.14708
EmotionAwareness3	0.67487	0.18453	0.17978	0.20103
EmotionAwareness4	0.61816	0.17497	0.35563	0.07372
EmotionAwareness5	0.59209	0.15927	0.14201	0.37962
SocialCompetence1	0.32711	0.20062	0.70334	0.07081
SocialCompetence2	0.19238	0.12316	0.79658	0.14353
SocialCompetence3	0.13264	0.26740	0.69166	0.24825
SocialCompetence4	0.10267	0.31691	0.58826	0.39280
Empathy1	0.13716	0.13829	0.19412	0.77338
Empathy2	0.38232	0.08371	0.15366	0.70394
Empathy3	0.11391	0.24046	0.15174	0.75477
Empathy4	0.29365	0.27531	0.21110	0.32621
GoalSetting1	0.26441	0.72610	0.15129	-0.00402
GoalSetting2	0.20950	0.72450	0.09431	0.07339
GoalSetting3	0.11757	0.54558	0.26933	0.24619
GoalSetting4	-0.01388	0.62214	0.28505	0.24285
GoalSetting5	0.25319	0.64114	0.11971	0.24042

## Promax Rotation Output

The FACTOR Procedure  
Rotation Method: Promax (power = 3)

## Target Matrix for Procrustean Transformation

	Factor1	Factor2	Factor3	Factor4
EmotionAwareness1	1.00000	0.00674	0.00050	0.00151
EmotionAwareness2	0.89417	0.01276	0.00716	0.00791
EmotionAwareness3	0.78668	0.01745	0.01622	0.02304
EmotionAwareness4	0.63409	0.01560	0.13169	0.00119
EmotionAwareness5	0.56406	0.01191	0.00849	0.16473
SocialCompetence1	0.07257	0.01817	0.78690	0.00082
SocialCompetence2	0.01291	0.00368	1.00000	0.00594
SocialCompetence3	0.00505	0.04486	0.78039	0.03666
SocialCompetence4	0.00244	0.07796	0.50124	0.15159
Empathy1	0.00503	0.00560	0.01557	1.00000
Empathy2	0.10938	0.00125	0.00775	0.75649
Empathy3	0.00295	0.03012	0.00761	0.95113
Empathy4	0.15603	0.13951	0.06322	0.23699
GoalSetting1	0.04085	0.91797	0.00835	-0.00000
GoalSetting2	0.02228	1.00000	0.00222	0.00106
GoalSetting3	0.00591	0.64129	0.07755	0.06017
GoalSetting4	-0.00001	0.73591	0.07115	0.04470
GoalSetting5	0.04326	0.76207	0.00499	0.04103

The FACTOR Procedure  
Rotation Method: Promax (power = 3)

## Inter-Factor Correlations

	Factor1	Factor2	Factor3	Factor4
Factor1	1.00000	0.41251	0.41279	0.38538
Factor2	0.41251	1.00000	0.49836	0.43004
Factor3	0.41279	0.49836	1.00000	0.47021
Factor4	0.38538	0.43004	0.47021	1.00000

Both the varimax and promax rotation results of the personal strength scale show a simple structure that could identify the items to the scale easily. For instance, the factor loadings of the first four items have high coefficients that loaded on the first factor for both the varimax and promax solution, highlighted by the red rectangular box. Similarly, the factor loadings also fall nicely for the rest of the three factors. For oblique rotation, an inter-factor correlation table is also provided to show the degree of correlations for the latent factors. The correlation coefficients of the four factors are moderately high that range from .38 to .50. These findings indicate the possibility of a high-order factor.

In summary, we would expect a satisfactory exploratory factor analysis if the followings are fulfilled.

- (1) Each factor has at least 3 items with moderate high loadings (say greater than 0.4) to be loaded on the factor
- (2) The rotated factor solution shows a simple structure that is easily interpretable without cross loadings. That is, only one high factor loading load on one factor.
- (3) The items that load on the factor are theoretically expected with high loadings.

### Confirmatory Factor Analysis

Exploratory factor analysis (EFA) is concerned about the structural pattern by relating all the items to the factors, while confirmatory factor analysis (CFA) indicates the structure a priori by specifying zero factor loadings to those items that do not belong to the factor and those belong to the scale with nonzero factor loadings. For instance, the items that load to the empathy subscale have nonzero loadings (Var10 to Var 13) while the rest of the items (Var 1 to Var9, and Var 14 to Var 18) have zero loadings. On the contrary, for exploratory factor analysis, we expect those items that load to the scale with high factor loadings and those items with low factor loadings. For instance, we expect high loadings for Var 10 to Var 13 to load to the scale empathy and low loadings for the rest of the items. As confirmatory factor analysis considers those items that do not belong to the subscale empathy by specifying zero factor loadings, we would not get an infinite number of factor solutions. However, we need to have at least three items for a scale in order for confirmatory factor analysis to have a solution. Since we state the model beforehand for CFA, the question is no longer of how low or high loadings are for the items for that of EFA, but we would like to know how well the model is being specified. That is, we are concerned with the fit of the model. Similar to EFA, we could have orthogonal and oblique for CFA. Figure 3 and 4 show the oblique and orthogonal CFA respectively of personal strength scale using path diagram. The difference between the two models is that there are red lines with double arrows of oblique CFA (Figure 3) showing the correlation between the four subscales while the orthogonal CFA (Figure 4) assumes no relationship between the subscales without the double arrows. From the path diagram, we could also see that each item load to their respective scale. The emotion awareness scale only contains variable 1 to 5. Variable 6 to 9 load to subscale social competence, variable 10 to 13 load into empathy subscale and variable 14 to 18 into subscale goal setting.

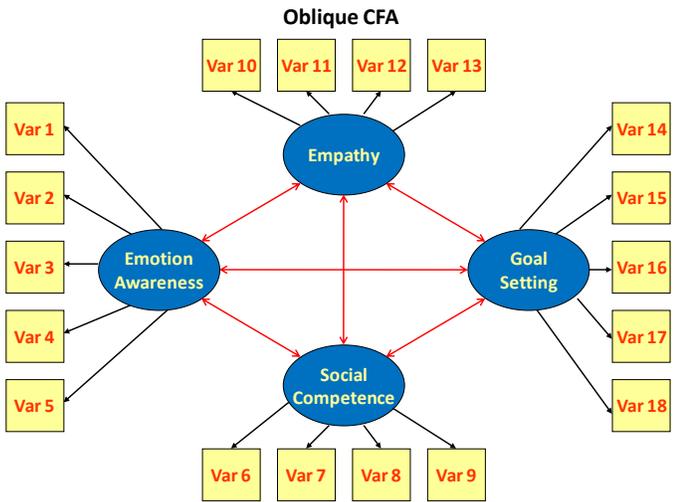


Figure 3 Oblique Confirmatory Factor Analysis

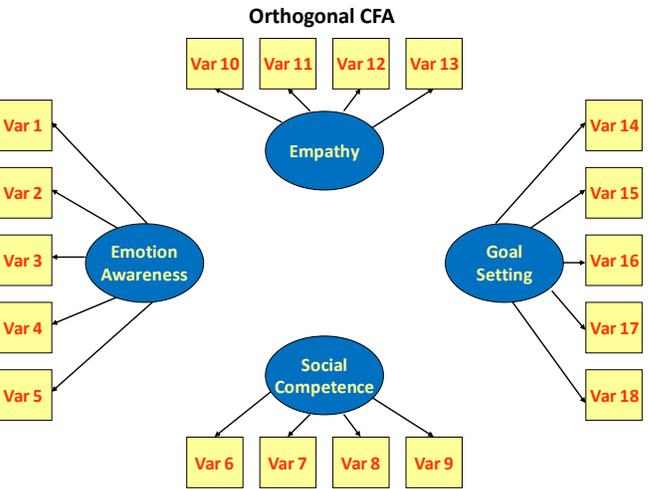


Figure 4 Orthogonal Confirmatory Factor Analysis

As the latent correlations of subscales are high, we would expect a high-order factor that could group these subscales together to form a single construct. Figure 5 shows the second-order CFA model. The 18 items load onto the respective four subscales (blue ovals) and these subscales further load onto a factor called personal strength scale (red oval). The four subscales are called the first-order factor whereas the personal strength scale is referred to as the second-order factor. The second-order model assumes that there exists a higher order scale that grouped the first-order factors together that represents and summarizes the four dimensions and aspects of student strength onto a single scale. Alternatively, we could also specify a one factor solution as shown in Figure 5 below. The one-factor model for personal strength scale loads all the 18 items onto one factor called personal strength scale. This model assumes that there is no subdivision of the 18 items into four sub-dimensions of student's personal strength but there is only one dimension of student personal strength measured by the 18 items.

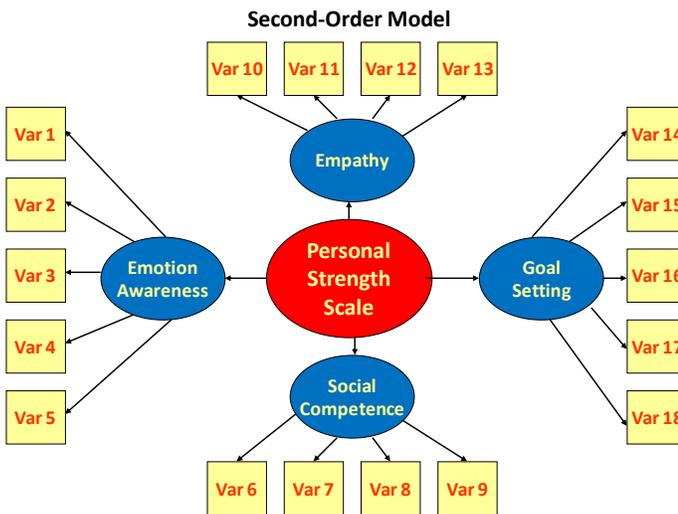


Figure 5 Second-Order Confirmatory Factor Model

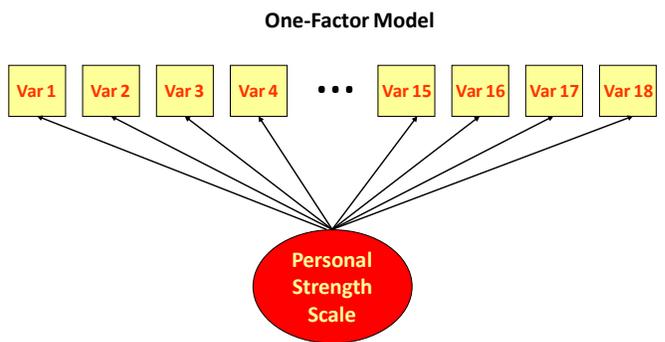


Figure 6 One Factor Model

We have briefly discussed four CFA models but which one is a better model and how do we establish and select the best model? Generally, we compare the fit indices of all the proposed models to select a best model. The fit statistics of the above models are tabulated below. Orthogonal and one-factor CFA have bad fit for GFI, NFI, RMSEA, and SRMR while oblique and second-order CFA have similar fit. This leads us to examine the correlation of the four subscales.

Table 1 Fit Indices of Confirmatory Factor Models

Fit Indices	Chi-Square	df	p-value	GFI	NFI	RMSEA	SRMR
Orthogonal CFA	4442.33	135	<.0001	0.7962	0.7027	0.1184	0.2664
One-Factor CFA	3113.80	135	<.0001	0.8411	0.7916	0.0985	0.0655
Oblique CFA	1270.26	129	<.0001	<b>0.9370</b>	<b>0.9150</b>	<b>0.0624</b>	<b>0.0445</b>
Second-Order CFA	1292.58	131	<.0001	0.9366	0.9135	0.0624	0.0454

The correlation coefficients show that these four subscales are moderately associated. The correlations are extracted from the oblique CFA as shown below. The output is generated using PROC CALIS. The syntaxes of PROC CALIS of the four CFA models are given in the appendix.

#### Correlations Among Exogenous Variables

Var1	Var2	Parameter	Estimate
fEmotionAwareness	fSocialCompetence	theta12	0.67045
fEmotionAwareness	fEmpathy	theta13	0.69661
fSocialCompetence	fEmpathy	theta23	0.71496
fEmotionAwareness	fGoalSetting	theta14	0.66963
fSocialCompetence	fGoalSetting	theta24	0.74900
fEmpathy	fGoalSetting	theta34	0.65781

#### PROC VARCLUS

We have so far discussed two approaches of scale construction using exploratory and confirmatory factor analysis. An alternative way is variable cluster analysis. PROC VARCLUS is the procedure for clustering variables. VARCLUS stands for variable clustering. As the procedure name suggests, it aims to classify variables into clusters of similarity by dividing them into disjoint or hierarchical clusters. The syntax of PROC VARCLUS that applies to the personal strength scale is shown below. The PROC VARCLUS statement states the VARCLUS procedure. The list of variables after the VAR statement states the variables to be used for variable clustering.

```
Proc VarClus Data = A;
  Var EmotionAwareness1-EmotionAwareness5
      SocialCompetence1-SocialCompetence4
      Empathy1-Empathy4
      GoalSetting1-GoalSetting5;
Run;
Proc VarClus Data = A Centroid Cov;
  Var EmotionAwareness1-EmotionAwareness5
      SocialCompetence1-SocialCompetence4
      Empathy1-Empathy4
      GoalSetting1-GoalSetting5;
Run;
```

There are two VARCLUS procedures specified above, one with defaults (the first procedure) and one with the CENTROID and COV options. With the default options, the correlation matrix is analyzed while the COV option specifies the covariance matrix to be used in analysis. The stopping criterion for the default is when the second

eigenvalue is less than one for each group of variables while with the CENTROID and COV options, the stopping criterion is when 75% of the variance is explained by the first centroid component. Similar to principal component analysis, either the correlation or the covariance matrix can be analyzed and when the correlation is specified all variables are treated as equally important. If covariance is specified, variables with larger variances would have more importance in the analysis. The VARCLUS procedure maximizes the variance that is explained by the cluster components by summing over all the clusters with linear combination of variables. The default is that the linear combination using the first principal component approach and the centroid component by specifying the CENTROID option. The first principal component is based on the weighted average of the variables that specified as much variance as possible while the centroid components are unweighted averages. With the COV option, the unweighted sum of the individual variables is used. Hence, it is not scale-invariant such that the different measurements are in not in consideration, i.e. the items are unstandardized variables. Without the COV specification, the sum would be of the standardized variables and it is scale-invariant.

The cluster components are oblique, not orthogonal, even when the cluster components are specified as the first principal components. In an ordinary principal component analysis, all components are computed from the same variables, and the first principal component is orthogonal to the second principal component and to every other principal component. On the contrary, under the VARCLUS procedure, each cluster component is computed from a different set of variables than all the other cluster components. As such, the first principal component of one cluster might be correlated with the first principal component of another cluster. Hence, we could view the VARCLUS algorithm as a type of oblique component analysis.

### VARCLUS Procedure – Personal Strength Scale

The following show the default output specification. The VARCLUS procedure displays the cluster process with cluster summary. It starts with one cluster solution. As the “second eigenvalue” value is greater than one, it splits the 18 items into two clusters. The cluster 1 of the two cluster solution has “second eigenvalue” greater than one, so the procedure proceeds to the third cluster solution. As the “second eigenvalue” of the 4-cluster solution have all less than one, the process stop at 4-cluster solution. These cluster summaries are listed below.

Cluster Summary for 2 Clusters						
Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained	Second Eigenvalue	
1	12	12	4.912382	0.4094	1.1167	
2	6	6	3.161343	0.5269	0.7398	
Cluster Summary for 1 Cluster						
Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained	Second Eigenvalue	
1	18	18	6.782611	0.3768	1.4214	
Cluster Summary for 2 Clusters						
Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained	Second Eigenvalue	
1	12	12	4.912382	0.4094	1.1167	
2	6	6	3.161343	0.5269	0.7398	
Cluster Summary for 3 Clusters						
Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained	Second Eigenvalue	
1	8	8	3.778008	0.4723	1.0788	
2	5	5	2.81133	0.5623	0.6204	
3	5	5	2.600607	0.5201	0.7405	
Cluster Summary for 4 Clusters						
Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained	Second Eigenvalue	
1	4	4	2.514108	0.6285	0.5816	
2	5	5	2.81133	0.5623	0.6204	
3	5	5	2.600607	0.5201	0.7405	
4	4	4	2.289596	0.5724	0.7682	

If we specify the CENTRIOD and COVA option, the stopping rule is that the proportion explained is more than 0.75. As such, 15 cluster solutions are given as shown below for the first two solutions and the fifteenth solution.

Cluster Summary for 1 Cluster					Cluster Summary for 2 Clusters				
Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained	Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained
1	18	9.022807	3.370606	0.3736	1	9	4.675787	2.100487	0.4492
					2	9	4.34702	1.910574	0.4395
...									
Cluster Summary for 15 Clusters									
Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained					
1	1	0.552419	0.552419	1.0000					
2	1	0.449839	0.449839	1.0000					
3	2	1.079169	0.838032	0.7766					
4	2	1.026604	0.798225	0.7775					
5	1	0.490744	0.490744	1.0000					
6	1	0.566919	0.566919	1.0000					
7	1	0.447975	0.447975	1.0000					
8	1	0.570854	0.570854	1.0000					
9	2	1.000062	0.763742	0.7637					
10	1	0.446488	0.446488	1.0000					
11	1	0.555805	0.555805	1.0000					
12	1	0.47092	0.47092	1.0000					
13	1	0.429978	0.429978	1.0000					
14	1	0.406079	0.406079	1.0000					
15	1	0.52895	0.52895	1.0000					

Total variation explained = 8.316971 Proportion = 0.9218

The following output displays the respective variables that belong to the cluster separated by dotted lines. The items fall nicely to the respective subscale of personal strength scale. For instance, all the five social competence items fall under cluster 1. The proportion of variance explained by the four-cluster solution is 56.75%. Two R-squared correlations are calculated for each cluster. The R-square with Own Cluster and next closest cluster display on the third and fourth columns of the output below. Own Cluster R-Square values should be higher than the R-square with any other cluster. The higher the own cluster R-squared the better the item falls under the cluster. The "Next Closest" contains the next-highest squared correlation of the variable with a cluster component. This value is low if the clusters are well separated. "1-R\*\*2 Ratio" is the ratio of one minus the value in the Own Cluster column to one minus the value in the Next Closest column. For instance, for social competence 1, the 1-R\*\*2 ratio is  $(1-0.6254)/(1-0.2244)=0.4829$ . A low ratios indicates well-separated clusters. Care should be taken if there are low values of R-squared with its own cluster and high values of the R-squared with the next closest cluster and high value for "1-R\*\*2 Ratio". It is noted that Empathy 4 has a high ratio. The latter dendrogram show that this item does not fall nicely under the hierarchical graph under empathy subscale.

#### Oblique Principal Component Cluster Analysis

Total variation explained = 10.21564 Proportion = 0.5675

4 Clusters		R-squared with		
Cluster	Variable	Own Cluster	Next Closest	1-R**2 Ratio
Cluster 1	SocialCompetence1	0.6254	0.2244	0.4829
	SocialCompetence2	0.6542	0.1709	0.4170
	SocialCompetence3	0.6349	0.2361	0.4780
	SocialCompetence4	0.5995	0.2791	0.5555
Cluster 2	EmotionAwareness1	0.5322	0.1236	0.5338
	EmotionAwareness2	0.6172	0.1726	0.4626
	EmotionAwareness3	0.5788	0.1855	0.5172
	EmotionAwareness4	0.5408	0.2212	0.5897
	EmotionAwareness5	0.5423	0.2467	0.6076
Cluster 3	GoalSetting1	0.5543	0.1733	0.5392
	GoalSetting2	0.5292	0.1468	0.5518
	GoalSetting3	0.4751	0.2023	0.6580
	GoalSetting4	0.4777	0.2037	0.6559
	GoalSetting5	0.5643	0.1894	0.5375
Cluster 4	Empathy1	0.6555	0.1932	0.4269
	Empathy2	0.6474	0.2683	0.4818
	Empathy3	0.6315	0.1906	0.4552
	Empathy4	0.3551	0.1662	0.7735

The intercluster correlations are analogous to interfactor correlations; they are the correlations among cluster components. Similar to factor analysis results, the variable clustering correlation among the subscales are moderately correlated. The intercluster correlation coefficients range from .52 to .59. These moderate high intercluster correlations again indicate the possibility of an overall personal strength scale instead of separate scales that explained diverse student's strength and ability.

## Inter-Cluster Correlations

Cluster	1	2	3	4
1	1.00000	0.53864	0.58862	0.56908
2	0.53864	1.00000	0.52453	0.55334
3	0.58862	0.52453	1.00000	0.51984
4	0.56908	0.55334	0.51984	1.00000

VARCLUS output is similar to PROC FACTOR procedure using oblique rotation as the VARCLUS algorithm is a type of oblique component analysis. The scoring coefficients have the same meaning in both PROC VARCLUS and PROC FACTOR. They are coefficients applied to the standardized variables to compute component scores. The cluster structure is analogous to the factor structure containing the correlations between each variable and each cluster component. A cluster pattern is not displayed because it would be the same as the cluster structure, except that zeros would appear in the same places in which zeros appear in the scoring coefficients.

## Standardized Scoring Coefficients

Cluster	1	2	3	4
EmotionAwareness1	0.000000	0.259499	0.000000	0.000000
EmotionAwareness2	0.000000	0.279456	0.000000	0.000000
EmotionAwareness3	0.000000	0.270605	0.000000	0.000000
EmotionAwareness4	0.000000	0.261577	0.000000	0.000000
EmotionAwareness5	0.000000	0.261951	0.000000	0.000000
SocialCompetence1	0.314561	0.000000	0.000000	0.000000
SocialCompetence2	0.321725	0.000000	0.000000	0.000000
SocialCompetence3	0.316933	0.000000	0.000000	0.000000
SocialCompetence4	0.307983	0.000000	0.000000	0.000000
Empathy1	0.000000	0.000000	0.000000	0.353625
Empathy2	0.000000	0.000000	0.000000	0.351434
Empathy3	0.000000	0.000000	0.000000	0.347092
Empathy4	0.000000	0.000000	0.000000	0.260247
GoalSetting1	0.000000	0.000000	0.286283	0.000000
GoalSetting2	0.000000	0.000000	0.279732	0.000000
GoalSetting3	0.000000	0.000000	0.265041	0.000000
GoalSetting4	0.000000	0.000000	0.265767	0.000000
GoalSetting5	0.000000	0.000000	0.288858	0.000000

## Oblique Principal Component Cluster Analysis

Number of Clusters	Total Variation Explained by Clusters	Proportion of Variation Explained by Clusters	Minimum Proportion Explained by a Cluster	Maximum Second Eigenvalue in a Cluster	Minimum R-squared for a Variable	Maximum 1-R**2 Ratio for a Variable
1	6.782611	0.3768	0.3768	1.421403	0.2946	
2	8.073725	0.4485	0.4094	1.116744	0.2987	0.8577
3	9.189945	0.5106	0.4723	1.078814	0.3184	0.8174
4	10.215640	0.5675	0.5201	0.768162	0.3551	0.7735

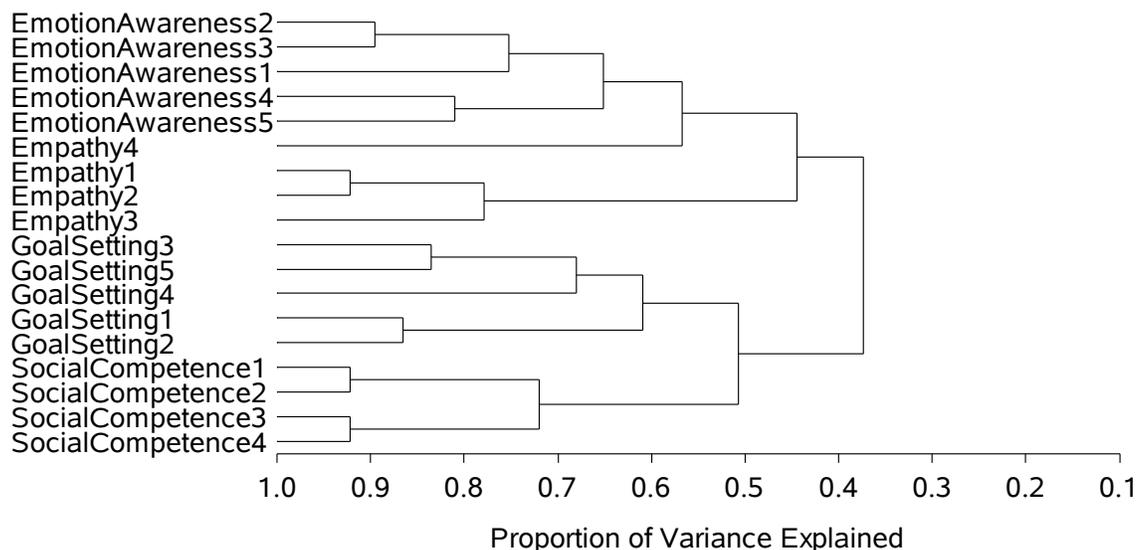
## Cluster Structure

Cluster	1	2	3	4
EmotionAwareness1	0.323771	0.729538	0.338605	0.351633
EmotionAwareness2	0.389488	0.785644	0.393267	0.415473
EmotionAwareness3	0.422173	0.760759	0.417364	0.430661
EmotionAwareness4	0.470368	0.735380	0.409650	0.379749
EmotionAwareness5	0.414188	0.736431	0.407212	0.496695
SocialCompetence1	0.790839	0.473696	0.454050	0.407560
SocialCompetence2	0.808852	0.409975	0.413399	0.404651
SocialCompetence3	0.796804	0.407058	0.485942	0.466966
SocialCompetence4	0.774301	0.417960	0.515542	0.528267
Empathy1	0.433526	0.377287	0.387603	0.809659
Empathy2	0.449882	0.517984	0.379185	0.804643
Empathy3	0.436614	0.379711	0.427265	0.794700
Empathy4	0.399643	0.407639	0.388933	0.595861
GoalSetting1	0.416320	0.404447	0.744509	0.324906
GoalSetting2	0.383167	0.366776	0.727473	0.350646
GoalSetting3	0.449746	0.372062	0.689268	0.388329
GoalSetting4	0.451355	0.307866	0.691156	0.378176
GoalSetting5	0.426129	0.435220	0.751205	0.433819

The results of variable clustering could be output to a dataset as input to PROC TREE to draw the dendrogram. The dendrogram of personal strength scale below show clearly the subscale structure probably with the exception of empathy 4. We could also observe that the emotional awareness subscale is closer to empathy subscale and the social competence subscale is closer to goal setting subscale.

## Personal Strength Scale

### Variable Clustering



The syntax of generating the dendrogram is given below.

```
Proc VarClus Data = Video OutTree = Tree;
  Var EmotionAwareness1-EmotionAwareness5
      SocialCompetence1-SocialCompetence4
      Empathy1-Empathy4
      GoalSetting1-GoalSetting5;
Run;
Title h=10pct 'Personal Strength Scale';
Title2 h=5pct 'Variable Clustering';
GOptions htext=4pct;
Axis1 order=(0.1 to 1 by 0.1);
Axis2 label=none;
Proc Tree Data = Tree Horizontal haxis=axis1 vaxis=axis2;
  Height _propor_;
  ID _label_;
Run;
```

### Comparison of Results of Factor analysis and variable clustering

It is useful to compare the factor analysis results with the VARCLUS results. The inter-factor correlation coefficients from PROC FACTOR and inter-cluster correlation coefficients from PROC VARCLUS are similar with moderately high association although the factor analysis shows slightly lower estimates. The correlation coefficients of confirmatory factor analysis have the highest correlation coefficients. This is expected as the measurement errors are removed. Both factor analysis and variable cluster analysis show similar factor structure of the personal strength scale.

## Advantages of Variable Cluster Analysis over Factor Analysis

One of the major limitations of traditional factor analysis is that the resulting factors are generally not sharply distinct. This is specifically obvious when cross loadings appear in the factor solution where there are items contributing to more than one of the scales in terms of factor loading. This difficulty is often circumvented by applying a multidimensional rotation. However the solution may not always end up with a simple structure. More importantly, the decision of creating a scale under EFA to assign weights of as zero to items that do not belong to the scale is inherently arbitrary and setting some weights to zero means the factor scores are no longer uncorrelated. Variable clustering, on the other hand, provides a more direct way of creating scales and could be highly successful in practice. The obvious advantage over modifying a rotated factor analysis solution is that it always accounts for all of the original variables. In addition, the form of the scale for each group of variables is rather flexible. We can use the first principal component of assuming an optimally weighted sum or the first centroid component of assuming using the simple sum. This provides options for researchers to decide whether an equally weighted is preferred or optimal weighted is ideal.

The VARCLUS procedure could be useful as a reduction method for a large set of variables. These variables can often be replaced by the set of cluster components with little loss of information. A given number of cluster components does not generally explain as much variance as the same number of principal components on the full set of variables, but the cluster components are usually easier to interpret than the principal components, even if the latter are rotated.

## CONCLUSION AND SUGGESTIONS

This paper compares the factor analytic technique and variable clustering analysis in measuring scale using personal strength scale as an example. The four-factor solution from exploratory factor analysis and the variable cluster analysis led to similar results. The personal strength scale has a clear 4 subscales. The inter-factor and inter-cluster are moderately correlated with each other. This leads us to examine the confirmatory factor analysis. The oblique CFA and second-order CFA give satisfactory results. The author suggests using variable clustering as an alternative method in practice to confirm the exploratory factor analysis results and in circumstance the exploratory factor analysis does not produce as intended, it serves as an option to examine the scale measurement.

## ACKNOWLEDGEMENTS

This paper acknowledges the Psychological Studies, National Institute of Education, Nanyang Technological University, Singapore for making the data available from the Video Game Research Project which was funded by the Ministry of Education, and Media Development Authority, Singapore.

## REFERENCES

- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, 103, 276-279.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7, 191-205.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.

Kabacoff, R. I. (2003). SUGI 28, Paper 90-28, *Determining the Dimensionality of Data: A SAS® Macro for Parallel Analysis*.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151.

Liau, A. K. (2007). *Promoting children's personal strengths: Positive psychology goes to school*. In A. G. Tan (Ed.), *Creativity: A handbook for teachers*. Singapore: World Scientific.

Liau, A. K., Chow, D., Tan, T. K., and Senf, K. (2010). Development and Validation of the Personal Strengths Inventory Using Exploratory and Confirmatory Factor Analyses. *Journal of Psychoeducational Assessment*.

Patil, V., Singh, S., Mishra, S., & Todd Donovan, D. (2008). Efficient theory development and factor retention criteria: Abandon the 'eigenvalue greater than one' criterion. *Journal of Business Research*, 61, 162–170.

SAS Institute Inc., *SAS/STAT® User's Guide, Version 9.1*. Cary, N.C.: SAS Institute, Inc., 2004.

#### **AUTHOR CONTACT INFORMATION**

Albert Liau Kien Fie  
Psychological Studies  
Nanyang Technological University  
1 Nanyang Walk, Block 3, Level 3  
Singapore 637616  
Work Phone: 065-67903224  
Email: [albert.liau@nie.edu.sg](mailto:albert.liau@nie.edu.sg)

Teck Kiang Tan  
Centre for Research in Pedagogy and Practice  
Nanyang Technological University  
1 Nanyang Walk, Block 5, Basement 3  
Singapore 637616  
Work Phone: 065-62196277  
Email: [teckkiang.tan@nie.edu.sg](mailto:teckkiang.tan@nie.edu.sg)

Angeline Khoo Cheok Eng  
Psychological Studies  
Nanyang Technological University  
1 Nanyang Walk, Block 3, Level 3  
Singapore 637616  
Work Phone: 065-67903212  
Email: [angeline.khoo@nie.edu.sg](mailto:angeline.khoo@nie.edu.sg)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.

## Appendix – Confirmatory Factor Analysis Syntax

\* Orthogonal CFA;

**Proc Calis** Data = Video;

linesq

```

EmotionAwareness1 = a1 fEmotionAwareness + e1,
EmotionAwareness2 = a2 fEmotionAwareness + e2,
EmotionAwareness3 = a3 fEmotionAwareness + e3,
EmotionAwareness4 = a4 fEmotionAwareness + e4,
EmotionAwareness5 = a5 fEmotionAwareness + e5,
SocialCompetence1 = a6 fSocialCompetence + e6,
SocialCompetence2 = a7 fSocialCompetence + e7,
SocialCompetence3 = a8 fSocialCompetence + e8,
SocialCompetence4 = a9 fSocialCompetence + e9,
Empathy1          = a10 fEmpathy          + e10,
Empathy2          = a11 fEmpathy          + e11,
Empathy3          = a12 fEmpathy          + e12,
Empathy4          = a13 fEmpathy          + e13,
GoalSetting1     = a14 fGoalSetting       + e14,
GoalSetting2     = a15 fGoalSetting       + e15,
GoalSetting3     = a16 fGoalSetting       + e16,
GoalSetting4     = a17 fGoalSetting       + e17,
GoalSetting5     = a18 fGoalSetting       + e18;
```

std

```

fEmotionAwareness = 1,
fSocialCompetence = 1,
fEmpathy          = 1,
fGoalSetting      = 1,
e1 = ee1,
e2 = ee2,
e3 = ee3,
e4 = ee4,
e5 = ee5,
e6 = ee6,
e7 = ee7,
e8 = ee8,
e9 = ee9,
e10 = ee10,
e11 = ee11,
e12 = ee12,
e13 = ee13,
e14 = ee14,
e15 = ee15,
e16 = ee16,
e17 = ee17,
e18 = ee18;
```

```

Var EmotionAwareness1-EmotionAwareness5
    SocialCompetence1-SocialCompetence4
    Empathy1-Empathy4
    GoalSetting1-GoalSetting5;
```

**Run;**

\* Oblique CFA;

**Proc Calis** Data = Video;

linesq

```

EmotionAwareness1 = a1 fEmotionAwareness + e1,
EmotionAwareness2 = a2 fEmotionAwareness + e2,
EmotionAwareness3 = a3 fEmotionAwareness + e3,
EmotionAwareness4 = a4 fEmotionAwareness + e4,
```

```

EmotionAwareness5 = a5 fEmotionAwareness + e5,
SocialCompetence1 = a6 fSocialCompetence + e6,
SocialCompetence2 = a7 fSocialCompetence + e7,
SocialCompetence3 = a8 fSocialCompetence + e8,
SocialCompetence4 = a9 fSocialCompetence + e9,
Empathy1          = a10 fEmpathy          + e10,
Empathy2          = a11 fEmpathy          + e11,
Empathy3          = a12 fEmpathy          + e12,
Empathy4          = a13 fEmpathy          + e13,
GoalSetting1     = a14 fGoalSetting      + e14,
GoalSetting2     = a15 fGoalSetting      + e15,
GoalSetting3     = a16 fGoalSetting      + e16,
GoalSetting4     = a17 fGoalSetting      + e17,
GoalSetting5     = a18 fGoalSetting      + e18;

std
fEmotionAwareness = 1,
fSocialCompetence = 1,
fEmpathy          = 1,
fGoalSetting      = 1,
e1 = ee1,
e2 = ee2,
e3 = ee3,
e4 = ee4,
e5 = ee5,
e6 = ee6,
e7 = ee7,
e8 = ee8,
e9 = ee9,
e10 = ee10,
e11 = ee11,
e12 = ee12,
e13 = ee13,
e14 = ee14,
e15 = ee15,
e16 = ee16,
e17 = ee17,
e18 = ee18;

Cov
fEmotionAwareness fSocialCompetence = theta12,
fEmotionAwareness fEmpathy          = theta13,
fEmotionAwareness fGoalSetting      = theta14,
fSocialCompetence fEmpathy          = theta23,
fSocialCompetence fGoalSetting      = theta24,
fEmpathy          fGoalSetting      = theta34;

Var EmotionAwareness1-EmotionAwareness5
    SocialCompetence1-SocialCompetence4
    Empathy1-Empathy4
    GoalSetting1-GoalSetting5;

Run;
* One Factor CFA;
Proc Calis Data = Video;
linesq
EmotionAwareness1 = a1 fPersonalStrength + e1,
EmotionAwareness2 = a2 fPersonalStrength + e2,
EmotionAwareness3 = a3 fPersonalStrength + e3,
EmotionAwareness4 = a4 fPersonalStrength + e4,
EmotionAwareness5 = a5 fPersonalStrength + e5,
SocialCompetence1 = a6 fPersonalStrength + e6,
SocialCompetence2 = a7 fPersonalStrength + e7,

```

```

SocialCompetence3 = a8 fPersonalStrength + e8,
SocialCompetence4 = a9 fPersonalStrength + e9,
Empathy1          = a10 fPersonalStrength + e10,
Empathy2          = a11 fPersonalStrength + e11,
Empathy3          = a12 fPersonalStrength + e12,
Empathy4          = a13 fPersonalStrength + e13,
GoalSetting1      = a14 fPersonalStrength + e14,
GoalSetting2      = a15 fPersonalStrength + e15,
GoalSetting3      = a16 fPersonalStrength + e16,
GoalSetting4      = a17 fPersonalStrength + e17,
GoalSetting5      = a18 fPersonalStrength + e18;
std
fPersonalStrength = 1,
e1 = ee1,
e2 = ee2,
e3 = ee3,
e4 = ee4,
e5 = ee5,
e6 = ee6,
e7 = ee7,
e8 = ee8,
e9 = ee9,
e10 = ee10,
e11 = ee11,
e12 = ee12,
e13 = ee13,
e14 = ee14,
e15 = ee15,
e16 = ee16,
e17 = ee17,
e18 = ee18;
Var EmotionAwareness1-EmotionAwareness5
    SocialCompetence1-SocialCompetence4
    Empathy1-Empathy4
    GoalSetting1-GoalSetting5;
Run;
* Second-Order CFA;
Proc Calis Data = Video;
lineqs
EmotionAwareness1 = a1 fEmotionAwareness + e1,
EmotionAwareness2 = a2 fEmotionAwareness + e2,
EmotionAwareness3 = a3 fEmotionAwareness + e3,
EmotionAwareness4 = a4 fEmotionAwareness + e4,
EmotionAwareness5 = a5 fEmotionAwareness + e5,
SocialCompetence1 = a6 fSocialCompetence + e6,
SocialCompetence2 = a7 fSocialCompetence + e7,
SocialCompetence3 = a8 fSocialCompetence + e8,
SocialCompetence4 = a9 fSocialCompetence + e9,
Empathy1          = a10 fEmpathy          + e10,
Empathy2          = a11 fEmpathy          + e11,
Empathy3          = a12 fEmpathy          + e12,
Empathy4          = a13 fEmpathy          + e13,
GoalSetting1      = a14 fGoalSetting      + e14,
GoalSetting2      = a15 fGoalSetting      + e15,
GoalSetting3      = a16 fGoalSetting      + e16,
GoalSetting4      = a17 fGoalSetting      + e17,
GoalSetting5      = a18 fGoalSetting      + e18,
fEmotionAwareness = a1 fPersonalStrength + d1,
fSocialCompetence = a2 fPersonalStrength + d2,

```

```
fEmpathy          = a3 fPersonalStrength + d3,
fGoalSetting      = a4 fPersonalStrength + d4;
std
fPersonalStrength = 1,
e1 = ee1,
e2 = ee2,
e3 = ee3,
e4 = ee4,
e5 = ee5,
e6 = ee6,
e7 = ee7,
e8 = ee8,
e9 = ee9,
e10 = ee10,
e11 = ee11,
e12 = ee12,
e13 = ee13,
e14 = ee14,
e15 = ee15,
e16 = ee16,
e17 = ee17,
e18 = ee18,
dl-D4 = ZETA1-ZETA4;
Var EmotionAwareness1-EmotionAwareness5
    SocialCompetence1-SocialCompetence4
    Empathy1-Empathy4
    GoalSetting1-GoalSetting5;
Run;
```