FACTOR ANALYSIS

Factor analysis is a method of data reduction. It does this by seeking underlying unobservable (latent) variables that are reflected in the observed variables (manifest variables). There are many different methods that can be used to conduct a factor analysis (such as principal axis factor, maximum likelihood, generalized least squares, unweighted least squares), There are also many different types of rotations that can be done after the initial extraction of factors, including orthogonal rotations, such as varimax and equimax, which impose the restriction that the factors cannot be correlated, and oblique rotations, such as promax, which allow the factors to be correlated with one another.

Factor analysis is a technique that requires a large sample size. Factor analysis is based on the correlation matrix of the variables involved, and correlations usually need a large sample size before they stabilize. Tabachnick and Fidell (2001, page 588) cite Comrey and Lee's (1992) advise regarding sample size: 50 cases is very poor, 100 is poor, 200 is fair, 300 is good, 500 is very good, and 1000 or more is excellent. As a rule of thumb, a bare minimum of 10 observations per variable is necessary to avoid computational difficulties.

FACTOR ANALYSIS ON SPSS

Analyze ➔ Data Reduction ➔ Factor ➔ Descriptive ➔ Initial Solution (under Statistics), Coefficient, Determinant, KMO & Bartlett Test (Correlation Matrix) ➔ Continue ➔ Extraction ➔ Principal Axis factoring from methods ➔ Unclick unrotated factor solution (under display) ➔ Check number of factors ➔ Continue ➔ Rotation ➔ Varimax (make sure rotation solution is checked)➔ Continue ➔ Options ➔ Sorted by size ➔ Suppress absolute values ➔ OK

<u>Table 1</u>

High correlations mean these items probably will be in the same factor.

Indicates how each question is associated with each other.

Correlation Matrix a

		item01 motivation	item02 pleasure	item03 competence	item04 low motiv	item05 low comp	item06 low pleas	item07 motivation	item08 low motiv	item09 competence	item10 low pleas	item11 low comp	item12 motivation	item13 motivation	item14 pleasure
Correlation	item01 motivation	1.000	.484	.628	-.305	-.745	-.165	.461	-.340	.209	.071	-.441	.186	.187	.040
	item02 pleasure	.484	1.000	.389	-.168	-.547	-.312	.261	-.176	.219	-.389	-.401	.118	.038	.475
	item03 competence	.628	.389	1.000	-.348	-.743	-.209	.423	-.248	.328	.027	-.513	.185	.170	.088
	item04 low motiv	-.305	-.168	-.348	1.000	.383	.323	-.596	.576	-.120	.102	.388	-.361	-.334	-.083
	item05 low comp	-.745	-.547	-.743	.383	1.000	.280	-.538	.276	-.361	.130	.605	-.187	-.169	-.188
	item06 low pleas	-.165	-.312	-.209	.323	.280	1.000	-.288	.192	-.131	.217	.418	-.044	.001	-.469
	item07 motivation	.461	.261	.423	-.596	-.538	-.288	1.000	-.606	.228	-.169	-.331	.347	.361	.180
	item08 low motiv	-.340	-.176	-.248	.576	.276	.192	-.606	1.000	-.243	.067	.370	-.392	-.308	-.117
	item09 competence	.209	.219	.328	-.120	-.361	-.131	.228	-.243	1.000	-.109	-.407	.406	.286	-.020
	item10 low pleas	.071	-.389	.027	.102	.130	.217	-.169	.067	-.109	1.000	.250	-.059	-.062	-.447
	item11 low comp	-.441	-.401	-.513	.388	.605	.418	-.331	.370	-.407	.250	1.000	-.148	-.008	-.238
	item12 motivation	.186	.118	.185	-.361	-.187	-.044	.347	-.392	.406	-.059	-.148	1.000	.607	.088
	item13 motivation	.187	.038	.170	-.334	-.169	.001	.361	-.308	.286	-.062	-.008	.607	1.000	-.030
	item14 pleasure	.040	.475	.088	-.083	-.188	-.469	.180	-.117	-.020	-.447	-.238	.088	-.030	1.000

a. Determinant = .001

Low correlations will not be in the same factor.

Should be greater than .0001. If close to zero, collinearity is too high. If zero, no solution is possible.

Table 2

Tests of assumptions.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.770
Bartlett's Test of Sphericity	Approx. Chi-Square	433.486
	df	91
	Sig.	.000

Should be greater than .70 indicating sufficient items for each factor.

Should be significant (less than .05), indicating that the correlation matrix is significantly different from an identity matrix, in which correlations between variables are all zero.

Table 3

Communalities

	Initial
item01 motivation	.660
item02 pleasure	.542
item03 competence	.598
item04 low motiv	.562
item05 low comp	.772
item06 low pleas	.382
item07 motivation	.607
item08 low motiv	.533
item09 competence	.412
item10 low pleas	.372
item11 low comp	.591
item12 motivation	.499
item13 motivation	.452
item14 pleasure	.479

Extraction Method: Principal Axis Factoring.

These communalities represent the relation between the variable and all other variables (i.e., the squared multiple correlation between the item and all other items).

Table 4

Eigenvalues refer to the variance explained or accounted for.

Percent of variance for each component before and after rotation.

Total Variance Explained

Factor	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.888	34.916	34.916	3.017	21.549	21.549
2	2.000	14.284	49.200	2.327	16.621	38.171
3	1.613	11.519	60.719	1.784	12.746	50.917
4	1.134	8.097	68.816			
5	.904	6.459	75.275			
6	.716	5.113	80.388			
7	.577	4.125	84.513			
8	.461	3.293	87.806			
9	.400	2.857	90.664			
10	.379	2.710	93.374			
11	.298	2.126	95.500			
12	.258	1.846	97.346			
13	.217	1.551	98.897			
14	.154	1.103	100.000			

Extraction Method: Principal Axis Factoring.

Half of the variance is accounted for by the first three factors.

Factor Matrix[a]

a. 3 factors extracted. 12 iterations required.

Table 5

Rotated Factor Matrix[a]

	Factor		
	1	2	3
item05 low comp	-.897		
item03 competence	.780		
item01 motivation	.777		
item11 low comp	-.572		.355
item12 motivation		.721	
item13 motivation		.667	
item08 low motiv		-.619	
item04 low motiv		-.601	
item07 motivation	.412	.585	
item09 competence		.332	
item14 pleasure			.797
item10 low pleas			.580
item02 pleasure	.487		-.535
item06 low pleas			.515

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

The items cluster into these three groups defined by high loadings.

Table 6

Factor Transformation Matrix

Factor	1	2	3
1	.747	.552	-.370
2	-.162	.692	.704
3	.645	-.466	.606

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

We will ignore this; it was used to convert the initial factor matrix into the rotated factor matrix.

INTERPRETATION OF OUTPUT

The first table in Output is a correlation matrix showing how each of the 14 items is associated with each of the other 13. Note that some of the correlations are high (e.g., + or - .60 or greater) and some are low (i.e., near zero). The high correlations indicate that two items are associated and will probably be grouped together by the factor analysis.

Next, several assumptions are tested. The determinant (located under the correlation matrix) should be more than .00001. Here, it is .001 so this assumption is met. If the determinant is zero, then a factor analytic solution can not be obtained, because this would require dividing by zero. This would mean that at least one of the items can be understood as a linear combination of some set of the other items. The Kaiser-Meyer-OIkin (KMO) measure should be greater than .70, and is inadequate if less than .50. The KMO test tells one whether or not enough items are predicted by each factor. The Bartlett test should be significant (i.e., a significance value of less than .05); this means that the variables are correlated highly enough to provide a reasonable basis for factor analysis.

The Total Variance Explained table shows how the variance is divided among the 14 possible factors. Note that four factors have eigenvalues (a measure of explained variance) greater than 1.0, which is a common criterion for a factor to be useful. When the eigenvalue is less than 1.0, this means that the factor explains less information than a single item would have explained. Most researchers would not consider the information gained from such a factor to be sufficient to justify keeping that factor. Thus, if you had not specified otherwise, the computer would have looked for the best four-factor solution by "rotating" four factors. Because we specified that we wanted only three factors rotated, only three will be rotated.

For this we will use an orthogonal rotation (varimax). This means that the final factors will be as uncorrelated as possible with each other. As a result, we can assume that the information explained by one factor is independent of the information in the other factors. We rotate the factors so that they are easier to interpret. Rotation makes it so that, as much as possible, different items are explained or predicted by different underlying factors, and each factor explains more than one item.

This is a condition called simple structure. Although this is the goal of rotation, in reality, this is not always achieved. One thing to look for in the Rotated Matrix of factor loadings is the extent to which simple structure is achieved.

The Rotated Factor Matrix table, which contains these loadings, is key for understanding the results of the analysis. Note that the computer has sorted the 14 math attitude questions (item 01 to item 14) into three overlapping groups of items, each which has a loading of |.30| or higher (|.30| means the absolute value, or value without considering the sign, is greater than .30). Actually, every item has some loading from every factor, but there are blanks in the matrix where weights were less than |.30|. Within each factor (to the extent possible), the items are sorted from the one with the highest factor weight or loading for that factor (i.e., item 05 for factor 1, with a loading of-.897) to the one with the lowest loading on that first factor (item 07). Loadings resulting from an orthogonal rotation are correlation coefficients of each item with the factor, so they range from -1.0 through 0 to + 1.0. A negative loading just means that the question needs to be interpreted in the opposite direction from the way it is written for that factor (e.g., item 05 "I am a little slow catching on to new topics in math" has a negative loading from the competence factor, which indicates that the people scoring higher on this item are lower in competence).

Usually, factor loadings lower than |.30| are considered low, which is why we suppressed loadings less than |.30|. On the other hand, loadings of |.40| or greater are typically considered high. This is just a guideline, however, and one could set the criterion for "high" loadings as low as .30 or as high as .50. Setting the criterion lower than .30 or higher than .50 would be very unusual.

The investigator should examine the content of the items that have high loadings from each factor to see if they fit together conceptually and can be named. Items 5, 3, and 11 were intended to reflect a perception of competence at math, so the fact that they all have strong loadings from the same factor provides some support for their being conceptualized as pertaining to the same construct. On the other hand, item 01 was intended to measure motivation for doing math, but it is highly related to this same competence factor. In retrospect, one can see why this item could also be interpreted as competence. The item reads, "I practice math skills until I can do them well." Unless one felt one could do math problems well, this would not be true. Likewise, item 02," I feel happy after solving a hard problem," although intended to measure pleasure at doing math (and having its strongest loading there), might also reflect competence at doing math, in that again one could not endorse this item unless one had solved hard problems, which one could only do if one were good at math. On the other hand, item09, which was originally conceptualized as a competence item, had no really strong loadings. Item 77, as mentioned earlier, had a high loading for the first factor, as expected. However, it also had a moderate loading for

Factor 3, which seems to be a (low) pleasure factor. This item reads, "I have some difficulties doing math as well as other kids my age." Can you think of why this might be related to low pleasure?

Every item has a weight or loading from every factor, but in a "clean" factor analysis almost all of the loadings that are not in the circles that we have drawn on the Rotated Factor Matrix will be low (blank or less than |.40|). The fact that both Factors 1 and 3 load highly on item 02 and fairly highly on item 77, and the fact that Factors 1 and 2 both load highly on item 07 is common but undesirable, in that one wants only one factor to predict each item.