

Statistics Corner

Questions and answers about language testing statistics:

Choosing the Right Number of Components or Factors in PCA and EFA

James Dean Brown (University of Hawai'i at Manoa)

QUESTION: In Chapter 7 of the 2008 book on heritage language learning that you co-edited with Kimi Kondo-Brown, a study (Lee and Kim, 2008) compares the attitudes of 111 Korean heritage language learners. On page 167 of that book, a principal components analysis (with varimax rotation) describes the relation of examining 16 purported reasons for studying Korean with four broader factors. Several questions come to mind. What is a principal components analysis? How does principal components analysis differ from factor analysis? What guidelines do researchers need to bear in mind when selecting "factors"? And finally, what is a Varimax rotation and why is it applied?

ANSWER: This inquiry has four sub-questions: (a) What are principal components analysis (PCA) and exploratory factor analysis (EFA), how are they different, and how do researchers decide which to use? (b) How do investigators determine the number of components or factors to include in the analysis? (c) What is rotation, what are the different types, and how do researchers decide which to use? And, (d) how are PCA and EFA used in language test and questionnaire development?

I have addressed the first question (a) in the previous column (Brown, 2009, pp. 26-30). I'll attend to the second one (b) here, and answer the other two in subsequent columns.

Choosing the Number of Components or Factors to Include in a PCA or EFA

So, how do researchers decide on the number of components or factors to include in a PCA or EFA? If the researcher erroneously includes the same number of PCA components as there are variables (say 12 components in the 12-variable Y/GPI Brazilian university student example used in the previous Statistics Corner column), each factor will represent one variable as shown in Table 1. Such situations, where only one variable loads heavily in each column indicates that the factor scores for each factor essentially represent a single variable; the researcher already knew about the single variable, so such single loading "components" or "factors" do not represent any underlying combinations of variables that provide new or interesting information (see the discussion of *trivial factors* below).

Table 1. PCA Results for the 12 Y/GPI Scales Administered in Brazil with 12 Components Based on 12 Variables

Variables	Components											
	1	2	3	4	5	6	7	8	9	10	11	12
Social extraversion	-0.042	-0.003	0.191	-0.085	0.132	0.924	-0.048	0.159	0.196	-0.014	-0.104	-0.003
Ascendance	-0.006	0.029	0.952	-0.044	0.090	0.175	0.010	0.076	0.157	0.008	-0.134	-0.003
Thinking extraversion	0.017	-0.069	0.008	-0.021	0.033	-0.039	0.992	0.047	-0.047	-0.029	0.014	-0.003
Rhathymia	0.054	0.119	0.083	0.061	0.202	0.160	0.059	0.922	0.143	0.163	-0.005	0.003
General activity	0.052	-0.077	0.186	-0.075	0.149	0.213	-0.063	0.154	0.900	-0.056	-0.152	-0.003
Lack of agreeableness	0.056	0.009	0.089	-0.005	0.958	0.119	0.037	0.180	0.124	0.069	-0.059	-0.003
Lack of cooperativeness	0.961	0.103	-0.005	0.107	0.057	-0.035	0.020	0.048	0.044	0.088	0.114	0.003
Lack of objectivity	0.112	0.934	0.031	0.151	0.010	-0.002	-0.082	0.113	-0.068	0.180	0.108	0.003
Nervousness	0.136	0.180	-0.053	0.870	-0.005	-0.097	-0.029	0.067	-0.076	0.262	0.233	0.003
Inferiority feelings	0.183	0.156	-0.213	0.292	-0.092	-0.146	0.026	-0.010	-0.203	0.201	0.814	0.003
Cyclic tendencies	0.116	0.226	0.012	0.271	0.096	-0.013	-0.043	0.199	-0.060	0.862	0.170	0.003
Depression	0.253	0.225	-0.085	0.267	-0.027	-0.103	-0.068	0.091	-0.183	0.220	0.190	0.814

What researchers need instead is some way to determine a smaller number of factors or components (hereafter referred to collectively as *factors*) that account for large amounts of the overall variance without creating any bloated specifics. To that end, a number of "stopping rules" have been proposed to determine when the researcher should stop adding factors (see Bryant and Yarnold, 1995, pp. 102-104; Gorsuch, 1983, pp. 164-174). How does a research know how many factors to use? When should the researcher stop? There are various statistical tests to determine the optimum number of variables (Gorsuch, 1983, p. 143-164), but more commonly these non-statistical strategies are used:

1. Kaiser's stopping rule
2. Scree test
3. Number of non-trivial factors
4. A priori criterion
5. Percent of cumulative variance

"a number of 'stopping rules' have been proposed to determine when the researcher should stop adding factors."

Each of these topics will now be explained and exemplified in turn.

Examples Illustrating the Five Stopping Rules

I will base this discussion on the same example used in the previous column. Recall that the data were based on the 12 subtests of the *Y/G Personality Inventory (Y/GPI)* (Guilford and Yatabe, 1957) which were: social extraversion, ascendance, thinking extraversion, rathymia, general activity, lack of agreeableness, lack of cooperativeness, lack of objectivity, nervousness, inferiority feelings, cyclic tendencies, and depression. The data were based on an English language version of the Y/GPI administered to 259 students at two universities in Brazil for comparative purposes. The descriptive results for these data were shown in Table 2 of the previous column.

Because ample theory and research indicate that the first six subtests are extraversion scales and the remaining six pertain to neuroticism (for more on the Y/GPI, see Guilford and Yatabe, 1957; Robson, 1994; Brown, Robson, and Rosenkjar, 2001), it would probably make sense to perform the EFA instead of PCA (as discussed in the last Statistics Corner column). Let's consider each of the five ways of deciding on the appropriate number of factors when they are applied to the example data.

(1) Kaiser's stopping rule

Kaiser's stopping rule states that only the number of factors with eigenvalues over 1.00 should be considered in the analysis. The initial analysis of the example data indicated that three factors had an eigenvalue of 1.00 or higher (see Table 2 which is taken directly from the initial analysis SPSS output). Notice in Table 2 that the Factors 1, 2, and 3 (labeled in the first column) have eigenvalues of 3.751, 2.492, and 1.115, respectively. Thus all three are above Kaiser's cut-point of 1.00. Factors 4 to 12 are below that cut-point with values of .851 down to .250.

Table 2. Initial EFA for the 12 Y/GPI Scales Administered in Brazil

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3.751	31.257	31.257
2	2.492	20.768	52.025
3	1.115	9.293	61.318
4	.851	7.088	68.407
5	.770	6.413	74.819
6	.605	5.043	79.862
7	.577	4.808	84.670
8	.497	4.144	88.814
9	.444	3.696	92.510
10	.341	2.839	95.349
11	.309	2.571	97.920
12	.250	2.080	100.000

Extraction Method: Principal Axis Factoring.

Now consider the results shown in Table 3 for a three-factor EFA of the example data (with varimax rotation). Notice that the first column contains labels for the 12 scales. Then the next four columns show the results for the EFA including loadings, communalities (on the right), and proportions of variance (across the bottom). Factor 1 appears to have fairly strong loadings from the six neuroticism scales as expected. Factor 2 also has fairly high loadings from the first, second, fourth, fifth, and sixth extraversion scales. Notice that Rhathymia loads on both factors (such variables are referred to as *complex*, in this case, Rhathymia has a poor positive correlation with Factor 1 and a good positive correlation with Factor 2). Inferiority feelings also load on both factors (this variable is also *complex*, but in this case, it has a very good positive correlation with Factor 1 and a fair negative one with Factor 2). Also, surprisingly, the third extraversion scale (Thinking

extraversion) does not load on Factor 1 or 2. Thus this variable does not seem to fit the theory developed in previous research. This result does not mean that the theory was wrong for the types of respondents that participated in the previous research. It does mean, however, that Thinking extraversion does not load on either Factor 1 or 2 for the types of Brazilian university students included in the data analyzed here. Indeed, Thinking extraversion has only one loading worth noting at 0.530, and that loading is just sitting there by itself and not forming a factor that is useful in any way. We will consider this situation further in the ensuing discussion.

Table 3. EFA (with Varimax Rotation) Loadings for the 12 Y/GPI Scales Administered in Brazil

Variables	Rotated EFA Eigenvalues ≥ 1.00			
	Factor 1	Factor 2	Factor 3	<i>h</i> ²
Social extraversion	-0.135	0.665	-0.118	0.474
Ascendance	-0.109	0.548	-0.098	0.321
Thinking extraversion	-0.072	-0.026	0.530	0.287
Rhathymia	0.391	0.606	0.206	0.562
General activity	-0.219	0.680	-0.070	0.515
Lack of agreeableness	0.119	0.540	0.197	0.345
Lack of cooperativeness	0.466	0.029	0.060	0.221
Lack of objectivity	0.614	0.044	-0.187	0.415
Nervousness	0.759	-0.171	-0.021	0.606
Inferiority feelings	0.677	-0.469	0.101	0.689
Cyclic tendencies	0.785	0.106	-0.005	0.627
Depression	0.784	-0.227	-0.076	0.672
Proportion of Variance	0.259	0.182	0.037	0.478

(2) Scree test

Another strategy for examining the eigenvalues is called the scree test.¹ This strategy involves creating a graphic visualization of the relationship between eigenvalues and number of factors as shown in Figure 1.

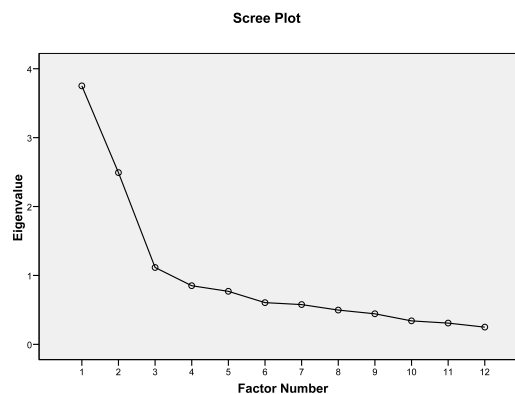


Figure 1. Scree Plot for the EFA for the 12 Y/GPI Scales Administered in Brazil

The scree plot is obviously a graph of the relationship between the relative magnitude of the eigenvalues and the number of factors. The researcher examines the scree plot and decides where the line stops descending precipitously and levels out (for more on scree plot interpretation, see Bryant and Yarnold, 1995, pp. 103-104).

Variables	Factors		
	1	2	<i>h</i> ²
Social extraversion	-0.108	0.668	0.458
Ascendance	0.086	0.553	0.314
Thinking extraversion	0.064	-0.019	0.005
Rhathymia	0.405	0.573	0.493
General activity	-0.191	0.692	0.515
Lack of agreeableness	0.139	0.527	0.297
Lack of cooperativeness	0.468	0.013	0.219
Lack of objectivity	0.607	0.018	0.368
Nervousness	0.754	-0.199	0.608
Inferiority feelings	0.656	-	0.675
Cyclic tendencies	0.794	0.077	0.633
Depression	0.773	-0.257	0.664
Proportion of Variance	0.255	0.183	0.437

In the case shown in Figure 1, that would appear to happen at three factors. The researcher then ignores all of the points along the level part of the line *including the transition point*, and counts the points along the precipitously dropping part of the line. Thus this particular scree plot indicates that a two factor solution would be appropriate.

Table 4 presents just such a two-factor analysis—one in which all variables either load clearly on Factor 1 or 2 (or are *complex*, as explained above, in the case of Rhathymia and Inferiority feelings) except for Thinking extraversion, which loads on neither factor.

Table 4. EFA (with Varimax Rotation) Loadings for 2 Factors Using the 12 Y/GPI Scales Administered in Brazil

(3) Number of non-trivial factors.

Trivial factors are usually defined as those that do not have two or three variables loading above the cut-point (often .30)² on them. Table 5 shows the loadings from an EFA (with Varimax Rotation) for an 11-factor solution using the example data. Notice in Table 5 that only three factors can be said to have three or more loadings above the cut-point of .30. Factors 5 through 9 are clearly examples of single loading factors, and factors 10 and 11 have no loadings worth considering, while Factor 4 only has two variables loading above the cut-point of .30. All of this would seem to argue for either a three or four-factor solution.

¹ Why is it called a *scree* test? Take a look at Figure 1 and try to visualize rocks and debris at the bottom of a cliff. See it? That stuff at the bottom of the cliff is called a *scree* in geology.

² Note that Tabachnick and Fidell (2007, p. 646) suggest a cut-point of .32 and above “...then there is 10% or more overlap in variance among factors.”

Table 5. EFA (with Varimax Rotation) Loadings for 11 Factors³ Using the 12 Y/GPI Scales Administered in Brazil

Variables	Factors										
	1	2	3	4	5	6	7	8	9	10	11
Social extraversion	-0.161	0.397	0.433	-0.086	0.140	-0.140	-0.008	0.433	-0.004	-	0.009
Ascendance	-0.073	0.176	0.704	-0.018	0.106	0.015	0.042	0.044	-0.018	0.003	0.004
Thinking extraversion	-0.058	0.060	-0.013	0.018	-0.040	0.620	-0.045	-0.022	-0.007	-	0.002
Rhathymia	0.236	0.688	0.115	0.007	0.182	0.123	0.237	0.134	0.247	0.001	-
General activity	-0.238	0.376	0.371	0.081	0.637	-0.157	-0.114	0.079	-0.033	-	0.004
Lack of agreeableness	0.008	0.615	0.176	0.089	0.044	0.046	-0.049	-0.007	-0.093	-	-
Lack of cooperativeness	0.256	0.091	-0.025	0.702	0.046	0.037	0.064	-0.021	0.009	0.033	0.006
Lack of objectivity	0.470	0.082	0.063	0.180	-0.129	-0.176	0.501	-0.007	0.038	0.004	0.006
Nervousness	0.820	-0.020	-0.113	0.164	-0.004	-0.026	0.040	-0.064	0.052	-0.117	-
Inferiority feelings	0.675	-0.209	-	0.266	-0.093	0.103	0.111	0.153	-0.071	0.005	0.044
Cyclic tendencies	0.752	0.286	0.008	0.111	-0.107	-0.081	0.143	-0.067	0.001	0.005	0.199
Depression	0.604	-0.009	-0.155	0.439	-0.294	-0.134	0.121	-0.045	0.399	-	-0.002
Proportion of Variance	0.208	0.110	0.092	0.071	0.050	0.043	0.032	0.021	0.020	0.026	0.004

However, there is the possibility that some factors may be trivial. In interpreting trivial and non-trivial factors, it is worth considering that triviality is a matter of degrees. According to Comrey and Lee (1992), loadings of .71 or higher can be considered "excellent", .63 is "very good", .55 is "good", .45 is "fair", and .32 is "poor". So what magnitude is trivial? Clearly, higher loadings indicate variables that are more highly related to whatever the underlying factor is. Hence, variables with high loadings can be considered *purier* measures of the underlying factors. This means that two, three, or more loadings higher than .71 are clearly less trivial than say two loadings of .30 or .40.

In Table 5, based on the number of loadings and their absolute magnitude, it could be argued that Factors 1 and 2 are less trivial than say Factors 3 and 4. The problem is that triviality may be in the eye of the beholder.

(4) *A priori criterion*

If the researcher were replicating previous research wherein a specific number of factors were found, it would make sense to set that same number of factors in the replication research. Similarly, if a researcher has created a set of test or questionnaire items to contain a specific number of subtests or scales, it would make sense to set that same number of factors in the factor analysis of those items. These are known as *a priori criteria* for determining the number of factors. For example, in previous research on the Y/GPI, the 12 subscales were shown to fall into two general categories: extroversion and neuroticism. Thus, a two factor solution for the example data (as shown in Table 4) would make theoretical sense based on a *a priori criteria* drawn from previous research.

(5) *Percent of cumulative variance*

An approach that is closely related to Kaiser's stopping rule and the scree plot is the percent of cumulative variance. However, percents of cumulative variance are harder to interpret than the other two. Clearly, in the example study, a 12 variable solution in a PCA would account for 100% of the variance, but as shown in Table 1, that would tell the researcher nothing. So some smaller number of factors should be sought. Looking down the far right column in Table 2 reveals the percentages of cumulative variance for various numbers of factors in that analysis. The addition of each factor adds some new variance to the cumulative variance. So where should the researcher stop? It is impossible to say.

As a result interpreting the percentages of cumulative variance is more a matter of keeping an eye on the amount of cumulative variance being accounted for by various other stopping rules. Close examination of the far right column in Table 2 indicates that 61.318% of the variance is accounted for if the three-factor solution (based on Kaiser's stopping rule) is used, but, if the two factor solution is adopted (based on the scree test), only 52.025% of the variance is accounted for.

Conclusion

³ Asking for 12 factors in an EFA with varimax rotation, SPSS warns: "You cannot request as many factors as variables with any extraction method except PC. The number of factors will be reduced by one."

"But which method is correct? I guess the safest answer is that no method is correct. Instead, some combination of the five sets of issues must be included in making the decision and explaining it to the readers of the resulting research report."

Clearly, there are a number of different ways to look at the issue of deciding how many components or factors to include in a PCA or EFA. I discussed Kaiser's stopping rule, the scree test, the number of non-trivial factors, a *a priori* criterion, and the percent of cumulative variance as different ways to make such decisions. Each of

these methods indicated that there should be three or two factors. But which method is correct? I guess the safest answer is that no method is correct. Instead, some combination of the five sets of issues must be included in making the decision and explaining it to the readers of the resulting research report. The trick is to make the strongest possible set of arguments for why a particular number of factors were selected in a particular analysis. If an *a priori criterion* argument can be included that may prove the most convincing and useful, but the point is that the argument for the number of components or factors should be based on a combination of information from the five viewpoints explained here.

In the case of the example data used here, there was indeed a theory-based *a priori criterion* of two factors. In addition, the scree plot indicated that a two-factor solution was appropriate, and the percent of variance accounted for by a two-factor solution is about 44% (see bottom right corner of Table 4). Finally, three-factor solutions tended to produce a trivial third factor, while two-factor solutions clearly produced two non-trivial factors. All in all, I am most comfortable with a interpreting the two-factor solution based on all of these considerations.

Two things should be clear in this discussion of how researchers decide on the number of factors to include in a PCA or EFA. First, the decision must be based on the preponderance of evidence from all five perspectives on the issue. Second, this is not a clear-cut decision based on a set of yes/no questions; there is an art to deciding on and explaining why you decided on a specific number of components or factors. And, third, the abilities needed for making such decisions and explaining them to readers improve over time (though they will never be perfect) so don't be afraid to critically read the explanations provided by researchers in second language studies, and indeed if you have testing, questionnaire, or other appropriate data of your own, don't hesitate to dive in and see what you find with a PCA or EFA.

References

- Brown, J. D. (2009). Statistics Corner. Questions and answers about language testing statistics: Principal components analysis and exploratory factor analysis: Definitions, differences, and choices. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 13(1) 26-30. Retrieved March 30, 2009 from http://www.jalt.org/test/bro_29.htm
- Brown, J. D., Robson, G., & Rosenkjar, P. (2001). Personality, motivation, anxiety, strategies, and language proficiency of Japanese students. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 361-398). Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawai'i Press.
- Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Guilford, J. P., & Yatabe, T. (1957). *Yatabe-Guilford personality inventory*. Osaka: Institute for Psychological Testing.
- Kim, J. O., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it*. Beverly Hills, CA: Sage.
- Lee, J. S., & H. Y. (2008). In K. Kondo-Brown & J. D. Brown (Eds.), *Teaching Chinese, Japanese, and Korean heritage language students*. New York: Lawrence Erlbaum Associates.
- Robson, G. (1994). *Relationships between personality, anxiety, proficiency, and participation*. Unpublished doctoral dissertation, Temple University Japan, Tokyo, Japan.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Upper Saddle River, NJ: Pearson Allyn & Bacon.

Statistics Corner Questions and answers about language testing statistics: Choosing the Right Type of Rotation in PCA and EFA. James Dean Brown (University of Hawai'i at Manoa). Question: In Chapter 7 of the 2008 book on heritage language learning that you co-edited with Kimi Kondo-Brown, there is a study (Lee & Kim, 2008) comparing the attitudes of 111 Korean heritage language learners. On page 167 of that book, a principal components analysis (with varimax rotation) describes the relation of examining 16 purported reasons for studying Korean with four broader factors. Several questions c Brown, J. D. (2009). Statistics corner: questions and answers about language testing statistics: choosing the right number of components or factors in PCA and EFA. Shiken: JALT Testing and Evaluation SIG Newsletter, 13, 19-23. Clare, L., Baddeley, A., Moniz-Cook, E. and Woods, B. (2003). A quiet revolution.