

Statistics Corner

Questions and answers about language testing statistics:

How are PCA and EFA used in language research?

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’s 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 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: Those are interesting questions and imply at least five sub-questions: (a) What are principal components analysis (PCA) and exploratory factor analysis (EFA), how do they differ, and how do researchers decide which to use? (b) How do investigators determine the number of components or factors to include? (c) What is rotation, and the most common rotation types, and how do researchers decide which to use? (d) How are PCA and EFA employed in language research? And, (e) how are PCA and EFA used in language test and questionnaire development? I addressed the first three questions (a, b, & c) in previous columns (Brown, 2009a, b, c). I’ll attend to the fourth one (d) here, and the fifth one (e) in the next column.

So how are PCA and EFA used in language research? I have found at least three uses for these forms of analysis in my research:

1. Reducing the number of variables in a study
2. Exploring patterns in the correlations among variables
3. Supporting a theory of how variables are related

Let’s consider each of these issues individually.

Reducing the Number of Variables in a Study

One of the primary uses of factor analyses is to reduce the number of variables in a study. In second language research, we are often dealing with large numbers of variables. Unfortunately, large sets of variables tend to reduce the statistical power of a study (i.e., reduce the possibility of finding statistically significant results even if such results exist in the population). We can often strengthen a study by eliminating redundant variables that are doing pretty much the same thing as other variables.

The fact that PCA and EFA are often used for reducing redundancy among variables is evident in the design of the SPSS statistical software, where PCA and EFA are found in the menu system in version 16 or earlier under **Analyze** then submenus **Data reduction** and **Factor** (or in version 17 under **Analyze** then submenus **Dimension reduction** and **Factor**).

An example of how such data reduction can be applied is found in Brown (1998), where I used PCA to help reduce 44 variables (various linguistic characteristics of the blanks in 50 cloze passages) to what turned out to be the four most important and relatively orthogonal (i.e., independent, or non-redundant) variables. As I put it in Brown (1998, pp. 19-20, 24):

Factor analysis techniques, including principal components analysis and Varimax rotation, were used to investigate the degree to which variables were orthogonal (independent of each other). ... A large number of linguistic variables were also examined for relationship to EFL Difficulty. Four of these variables were selected on the basis of factor analysis as being orthogonal: syllables per sentence, average frequency elsewhere in the passage of the words that had been deleted, the percent of long words of seven letters or more, and the percent of function words. When combined, they proved to be the best predictors of observed EFL Difficulty.

The mechanics of reducing the number of variables in a study can be accomplished in several ways: (a) by going factor-by-factor and using that variable that loads highest on the first factor to represent all the other variables that load heavily on that factor, then turning to the second factor and doing the same thing, and then turning to the third factor, etc., or (b) by saving and using the component or factor scores (that are produced during the PCA or EFA analyses) as variables to represent the components or factors in the study. Clearly then, one way to use factor analyses is for reducing the number of variables in a study and thereby increasing the power of the study.

Exploring Patterns in the Correlations Among Variables

Correlational analysis is very common in second language studies. Research articles often present correlation matrices of all intercorrelations among 5, 10, 20, or more variables, replete with asterisks showing which were significant at $p < .05$, or $p < .01$, or both. Interpreting overall patterns in such matrices by simply eye-balling them is difficult for at least three reasons: (a) each correlation coefficient only represents the degree of relationship between two variables, (b) the p values in large sets of correlation coefficients are accurate for any one pair of variables, but not for the entire set, (c) the underlying sample sizes, distributions, and reliabilities can differ substantially among variables and sometimes dramatically affect the magnitude of the resulting correlation coefficients.

PCA and EFA provide tools that can help explore a correlation matrix and find overall patterns that may exist among the correlations. For example, in Brown (2001), I analyzed (based on data first gathered and analyzed for Yamashita, 1996, by permission) six types of pragmatics tests Written Discourse Completion Tasks (WDCT), Multiple-choice DCTs (MDCT), Oral DCTs (ODCT), Discourse Role Play Tasks (DRPT), Discourse Self Assessment Tasks (DSAT), and Role Play Self Assessments (RPSA) when they were administered in Japan to native-speakers of English learning Japanese as a second language (JSL). All possible correlations for these six pragmatics tests are shown below the diagonal line of 1.00s in Table 1. Coefficients of determination, i.e., the squared correlation coefficients, are shown above the diagonal.

Table 1. *Correlation Coefficients (Below the Diagonal) and Coefficients of Determination (Above the Diagonal) for the JSL Data*

JSL	WDCT	MDCT	ODCT	DRPT	DSAT	RPSA
WDCT	1.00	0.15	0.45	0.34	0.24	0.16
MDCT	0.39*	1.00	0.14	0.06	0.01	0.03
ODCT	0.67*	0.37*	1.00	0.62	0.31	0.24
DRPT	0.58*	0.24	0.79*	1.00	0.37	0.28
DSAT	0.49*	0.11	0.56*	0.61*	1.00	0.47
RPSA	0.40*	0.18	0.49*	0.53*	0.68*	1.00

* $p < .01$

Notice that the patterns of correlations in Table 1 are not very interesting. Sure, 12 of the correlation coefficients in Table 1 are significant at $p < .01$. But even with all these significant correlation coefficients, each coefficient only tells us about the degree of relationship between whatever two variables are involved. No amount of staring at Table 1 leads to any interesting pattern of overall relationships (except, perhaps, that the MDCT didn't correlate well with any other measure). In addition, there is no way of knowing from Table 1 how much differences in the sample sizes, distributions of scores, and test reliabilities of the variables may have affected the relative values of these correlation coefficients, or the degree to which the number of correlation coefficients has distorted the meaning of the p values.

However, a factor analysis of the same data can be much more revealing (see Table 2). Notice in Table 2 that the highest loadings for each variable (in bold type) indicate that the ODCT, DRPT, DSAT, and RPSA all load most heavily on the first factor, while the WDCT and MDCT load more heavily on the second factor. Because the ODCT, DRPT, DSAT, and RPSA are all tests of oral abilities, the first factor could be labeled as an oral-language factor. In contrast, the WDCT and MDCT can both be considered written-language tests, so factor two might appropriately be labeled a written-language factor. These oral-language and written-language categories were interpreted in the original paper as test method factors. Note, however, that this argument is somewhat undermined by the fact that consideration of all loadings above .30 (i.e., those with asterisks) shows a pattern that was not quite so clear because the WDCT, ODCT, and DRPT are complex, that is, they all load to some meaningful degree on both factors. Nonetheless, the original paper interpreted these patterns as indications of test-method effects—an interpretation that would not have been possible based on the correlation matrix alone.

Table 2. *Factor Analysis of JSL Data (after VARIMAX rotation)**

VARIABLE	FACTOR 1	FACTOR 2	h^2
WDCT	*.56	*.61	.68
MDCT	-.02	*.90	.82
ODCT	*.70	*.54	.78
DRPT	*.78	*.36	.74
DSAT	*.89	.03	.79
RPSA	*.82	.04	.68
Proportion of Variance	.48	.27	.75

*Note: Bold-faced type indicates highest loading for each variable. Asterisks show all loadings over .30.

Clearly then, while it is often difficult to detect patterns in correlation matrices, factor analysis techniques can reveal interesting and interpretable patterns among those same correlation coefficients. However, researchers must take great care in interpreting such patterns, especially insuring that they have looked for complexity and that they only name components or factors very carefully and tentatively based on what they *think* is going on.

Supporting a Theory of How Variables Are Related

Sometimes researchers have a theory of how variables should be related to each other. Such was the case in Brown, Rosenkjar, and Robson (2001), where among other things, we examined the *Y/G Personality Inventory* (Y/GPI) (Guilford & Yatabe, 1957), which assesses twelve traits (social extraversion, ascendance, thinking extraversion, rathymia, general activity, lack of agreeableness, lack of cooperativeness, lack of objectivity, nervousness, inferiority feelings, cyclic tendencies, and depression) with ten items per trait.

Previous research had consistently shown that these twelve traits fall into two general categories labeled neuroticism and extraversion (the first six traits representing extraversion and the last six representing neuroticism). The results shown in Table 3 are for Brazilian university students taking the Y/GPI. With the exception of *Thinking extraversion*, the bold-faced italics loadings are in exactly the pattern of relationships that theory would predict.

However, there is also some minor evidence that these data do not fit the theory. First, *Thinking extraversion* does load strongly on either factor, perhaps because the participants were Brazilian, or for some other reason altogether. In addition, *Rhathymia* and *Inferiority feelings* load at higher than .30 on both factors. So these variables are complex for this data set.

Table 3. *Two-Factor Analysis (with Varimax rotation) of the 12 Variables of Question on the Y/G Personality Inventory*

Variables	Rotated 2 Factors		
	Factor 1	Factor 2	<i>h</i> ²
Social extraversion	-0.108	<i>0.668</i>	0.458
Ascendance	-0.086	<i>0.553</i>	0.314
Thinking extraversion	-0.064	-0.019	0.005
Rhathymia	0.405	<i>0.573</i>	0.493
General activity	-0.191	<i>0.692</i>	0.515
Lack of agreeableness	0.139	<i>0.527</i>	0.297
Lack of cooperativeness	<i>0.468</i>	0.013	0.219
Lack of objectivity	<i>0.607</i>	0.018	0.368
Nervousness	<i>0.754</i>	-0.199	0.608
Inferiority feelings	<i>0.656</i>	-0.494	0.675
Cyclic tendencies	<i>0.792</i>	0.077	0.633
Depression	<i>0.773</i>	-0.257	0.664
Proportion of Variance	0.255	0.183	0.437

Whether or not the patterns found in the data are 100% clear, the underlying complexities, and discrepancies from theory can be very interesting. Such complexities and discrepancies can even serve as the basis for carefully revising or refining the theories that are being examined.

Conclusion

Many researchers use factor analysis for one purpose or another without realizing the rich variety of other purposes this form of analysis can serve. I've shown here that EFA and PCA have applications in language

research work that include at least reducing the number of variables in a study, exploring patterns in the correlations among variables, and supporting a theory of how variables are related. In the next column, I will discuss three ways EFA and PCA are often used in test or questionnaire development projects.

If you are currently using EFA and PCA, consider expanding the ways you apply these analyses. If you are *not* currently using EFA and PCA, you might want to ask yourself, why not?

References

- Brown, J. D. (1998). An EFL readability index. *JALT Journal, 20* (2), 7-36.
- Brown, J. D. (2001). Six types of pragmatics tests in two different contexts. In K. Rose & G. Kasper (Eds.), *Pragmatics in Language Teaching* (pp. 301-325). Cambridge: Cambridge University.
- Brown, J. D. (2009a). 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. Also retrieved from the World Wide Web at http://www.jalt.org/test/bro_29.htm
- Brown, J. D. (2009b). Statistics Corner. Questions and answers about language testing statistics: Choosing the right number of components or factors in PCA and EFA. *Shiken: JALT Testing & Evaluation SIG Newsletter, 13* (2), 19-23. Also retrieved from the World Wide Web at http://www.jalt.org/test/bro_30.htm

Brown, J. D. (2009c). Statistics Corner. Questions and answers about language testing statistics: Choosing the right type of rotation in PCA and EFA. *Shiken: JALT Testing & Evaluation SIG Newsletter, 13* (3), 20 - 25. Also retrieved from the World Wide Web at http://www.jalt.org/test/bro_31.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.

Guilford, J. P., & Yatabe, T. (1957). *Yatabe Guilford Personality Inventory*. Osaka: Institute for Psychological Testing.

Lee, J. S., & Kim, H. Y. (2008). Heritage language learners' attitudes, motivations, and instructional needs: The case of postsecondary Korean language learners. In K. Kondo-Brown & J. D. Brown (Eds.), *Teaching Chinese, Japanese, and Korean Heritage Language Students*. New York: Lawrence Erlbaum Associates.

Yamashita, S. O. (1996). *Six measures of JSL Pragmatics* (Technical Report #14. Second Language Teaching and Curriculum Center). Honolulu, HI: University of Hawai'i Press.