Rasch Measurement in Language Education Part 6:

Rasch Measurement and Factor Analysis

by James Sick (International Christian University, Tokyo)

Previous installments of this series have provided an overview of Rasch measurement theory, reviewed the differences among the various Rasch models, and discussed the assumptions and requirements that underlie Rasch measurement theory (RMT) and item response theory (IRT). In this installment, I will compare RMT with factor analysis, another technique frequently used in the validation of questionnaire data.

Question: Can you elaborate on the difference between principal component or factor analysis, and Rasch analysis? Aren't they both be used to validate questionnaires? Also, how does principal components analysis of Rasch residuals differ from ordinary principal component analysis?

Answer: Both factor analysis and Rasch analysis are frequently employed in the validation of tests and questionnaires, sometimes independently and sometimes in conjunction. For an online explanation of factor analysis and principle components analysis, see Brown (2001) or the *Statsoft Electronic Textbook* (2001). For this article, I'll assume a basic familiarity with factor analysis and focus on the differences between Rasch and factor analysis. I will also use factor analysis as a blanket term for both principle factor analysis and the closely related principle components analysis (PCA).

Rasch Analysis versus Factor Analysis

Both Rasch analysis and factor analysis are used to evaluate the dimensionality of a data set: that is, in identifying the number of abilities, attitudes, or traits that are influencing the response pattern. Although factor analysis *can* be used to test whether a data set is indicative of a single latent trait, it is more commonly used to identify *multiple* sources of variance. For example, a factor analysis might be applied to a motivation questionnaire in order to identify more specific components of that construct, such as integrativeness, desire to learn English, or enjoyment of language study. Components may be designated *a priori,* referred to as confirmatory factor analysis, or *post hoc*, referred to as exploratory factor analysis. In classical approaches to questionnaire validation, a factor analysis is first conducted in order to confirm or create subscales. The subscales are then given descriptive labels that indicate the traits they are hypothesized to measure. The total scores from each subscale can then be used as person measures for each trait, and Cronbach alpha or a similar statistic used to estimate the reliability of the subscale scores.

Rasch analysis, in contrast, takes as its starting point the assumption that a set of items is intended to measure a single construct. If a questionnaire contains items that are hypothesized to measure multiple traits, subscales must be designated *a priori* and separate Rasch analyses conducted for each subscale. In other words, Rasch analysis is not designed to identify multiple constructs. That is deemed to be the responsibility of the instrument designer, who groups items into subscales in advance based on experience or theory.

In addition, the Rasch model requires a different set of criteria for the items to fit the model. Factor analysis is a correlational model. For items to load on a factor, they must correlate with the other items that designate that factor. Endorsing one item thus implies that a respondent is likely to endorse *all* of the items making up that subscale. Rasch analysis, in contrast, is a hierarchical implicational model. Difficult items are expected to be endorsed only by respondents who possess a greater amount of the trait. Persons with low degrees of the trait are expected to endorse only the *easy* items. Put another way, the items form a hierarchical structure where positive responses to difficult items imply positive responses to easier items, but the reverse is not true. That is, positive responses to easy items do not necessarily imply positive responses to difficult items. The Rasch implicational structure is directional, from hard to easy, while factor analysis is non-directional.

In fact, wide differences in item difficulty, or endorsability in the case of Likert style items, can be problematic for factor analysis. When an item is difficult to endorse, it may not correlate strongly with items that are easy to endorse, even if these items are indicative of the same trait. In some instances, easy items and difficult items may not load together, forming "difficulty" factors, a misleading result that is considered a nuisance by the factor analyst. Designing a questionnaire that works well with factor analysis thus requires that the survey designer avoid items that are either very easy or very difficult to endorse. The Rasch approach is not affected by this restriction, and in fact functions best when items vary in difficulty. From a Rasch perspective, a well-designed questionnaire employs items with a range of difficulty that matches the range of person measures in the target audience. Moreover, the hierarchical ranking of the items can be employed as an empirical test of the validity of the construct. If the items measure a single, coherent latent trait, the Rasch fit statistics should indicate that items fit the model. In addition, the ranking of the items should "makes sense" to the analyst qualitatively, in light of what is understood about the ability or construct being measured.

Another difference of interest is that Rasch theory requires some degree of probabilistic uncertainty in the responses. That is, a response to an item or a set of items should never predict the responses to another item *perfectly*. When there is little or no stochastic variation in responses, an item is said to *overfit* the Rasch model. Two types of questionnaire items that tend to overfit the Rasch model are negative restatements and summary items. A negative restatement would be something like "I like English" and "I hate English," with the second item reverse scored. A summary item is one that summarizes the construct or other items. For example, a questionnaire intended to measure the construct "liking coffee" contains the items "I drink coffee in the morning," "I drink coffee for lunch," "I drink coffee in the evening," and "I like coffee." The final item would be a summary item. Respondents endorsing the first three items almost by necessity, like coffee. Items such as these two examples tend to perform very well in a factor analysis because they load strongly on common factors. In a Rasch analysis, however, they become candidates for deletion because they are too predictable and thus overfit the model. According to Rasch theory, overfitting items do not degrade the quality of measurement, but are inefficient because they provide no unique information about the respondents. Moreover, their high correlations tend to artificially inflate estimates of reliability, tricking us into believing that we are measuring more accurately than we truly are (Wright, Linacre, Gustafson, & Martin-Loff, 1994).

Although it is not uncommon to validate questionnaires by using factor analysis to identify subscales followed by Rasch analysis to assess the quality of the subscales and construct measures, differences in the requirements of these two approaches can cause problems when they are used in conjunction. Factor analysis tends to favor items that fall within a narrow range of difficulty, as well as items that are redundant or are lack item independence. Prescreening items for Rasch analysis by first employing factor analysis, or applying Rasch analysis to a questionnaire that was originally developed using factor analysis, can result in scales with overfitting items and restricted ranges of item difficulty. It is recommended that the analyst consider these effects if using Rasch and factor analysis in conjunction, especially if items are deleted based on the results of a factor analysis before they are tested using Rasch.

Principal Component Analysis of Rasch Residuals

Principle component analysis of the Rasch residuals is an extension of Rasch fit analysis used to confirm whether the Rasch difficulty dimension adequately accounts for all of the non-random variance in the data. Unlike conventional factor analysis, it is not usually used in an exploratory manner: that is, to search for and identify multiple constructs within a data set. A brief explanation of Rasch residual PCA was given in the previous installment of this series (Sick, 2010). Because the interpretation of residual PCA is a substantial topic, I will take it up in greater detail in the next installment.

References

Brown, J. D. (2001). What is an eigen value? *Shiken, 5*(1), 15-19. Retrieved March 17, 2011 from
     http://jalt.org/test/bro_10.htm

Sick, J. R. (2010). Rasch measurement in language education Part 5: Assumptions and requirements
     of Rasch measurement. *Shiken, 14*(2), 23-29. Retrieved March 17, 2011 from
      http://jalt.org/test/sic_5.htm

Statsoft Inc. (2011). Electronic statistics textbook: Principal components and factor analysis.
     Retrieved March 17, 2011 from http://www.statsoft.com/textbook/principal-components-factor-
     analysis/

Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Loff, P. (1994). Reasonable mean-square
     fit values. *Rasch Measurement Transactions, 8*(3), 370. Retrieved March 17, 2011 from
    http://www.rasch.org/rmt/rmt83b.htm

HTML: http://jalt.org/test/sic_6.htm   /   **PDF**: http://jalt.org/test/PDF/Sick6.pdf