PARAMETRIC VS. NON-PARAMETRIC IRT MODELING OF LIKERT-TYPE PERSONALITY DATA

Albert Maydeu-Olivares, David Gallardo, Uwe Kramp
Faculty of Psychology. University of Barcelona

1. Introduction

Most often personality assessment is performed using Likert-type items. Since these variables are categorical it is in principle more appropriate to use a non-linear latent trait model (an IRT model) than a linear model (factor analysis) as the relationship between a categorical vari able and a continuous latent trait can never be exactly linear (McDonald, 1999). However, when we abandon the linear model for a non-linear IRT model, which IRT model to use?

The only study in which the fit of different IRT models to personality data has been compared is Chernyshenko , Stark, Chan, Drasgow and Williams (2001) . These authors investigated the fit of the most widely used IRT model for Likert-type data (Samejima's logistic graded model) against a non-parametric IRT model (Levine's MFS model) to the scales of two personality questionnaires, the 16PF Questionnaire and to Goldberg's public domain measure of the Big Five personality factor markers. The fit of the non-parametric model was adequate but the fit of Samejima's model was poor. Hence they concluded that the fit of traditional parametric IRT models was "a cause of concern".

We attribute the poor fit of the graded model to small amounts of multidimensionality present in their data. To verify this conjecture, we compare the fit of these models a personality inventory, the Social Problem Solving Inventory-Revised (D'Zurilla , Nezu & Maydeu-Olivares, 2002), whose scales were designed to be unidimensional. Also, we consider additional parametric IRT models.

2. Method

2.1 Participants

We used two samples of undergraduate students from a large public university from the USA, a calibration sample (N = 1053), and a cross validation sample of new observations from the same university (N = 953).

2.2 Measures

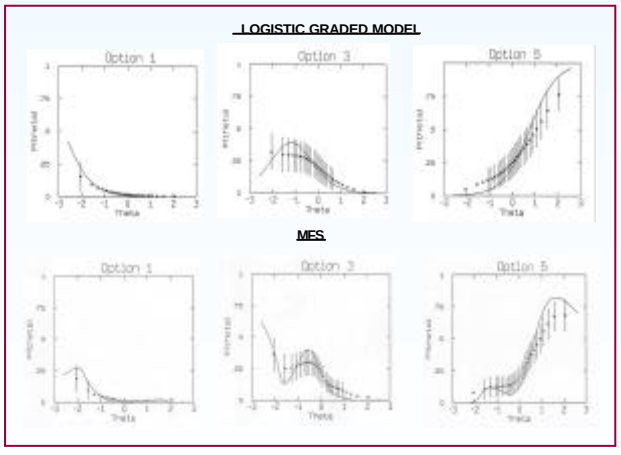
We shall fit PPO scale of the SPSI-R (D'Zurilla et al., 2002). This questionnaire consist of 52 Likert-type items with five ordered categories each aimed at measuring the process by which people attempt to resolve problems they experience in everyday living. The SPSI-R measures two constructive or adaptive problem-solving dimensions -positive problem orientation (PPO) and rational problem solving (RPS)- and three dysfunctional dimensions - negative problem orientation (NPO), impulsivity/carelessness style (ICS), and avoidance style (AS). The number of items comprising the PPO, NPO, RPS, ICS and AS scales are 5, 10, 20, 10 and 7, respectively.

2.3 IRT models, their estimation and goodness of fit assessment

The following unidimensional models will be compared:

1) Samejima's logistic graded model
2) Samejima's normal ogive graded model
3) Bock's nominal model
4) Thissen and Steinberg's ordinal model
5) Masters' partial credit model
6) Levine's MFS model

Figure 1.
Fitplot of the CRFs of selected options of item 2 of the PPO scale in the cross-validation sample



The first five models are parametric and the last one is non-parametric. Parametric IRT models can be divided into two large classes, difference models and divide-by-total models (Thissen & Steinberg, 1986). Samejima's graded model is a difference model, and models 3), 4) and 5) are divide-by-total models. These three models are nested: Bock's model (Bock, 1972) is the most general and it has more parameters than Samejima's, Thissen and Steinberg's has the same number of parameters as Samejima's, and Masters' model (Master, 1982) is the most restrictive (a Rasch-type model).

To assess the goodness of fit of the models, we shall compute chi-square statistics for each item, pair of items, and triplet of items (see Drasgow et al, 1995; Chernyshenko et al., 2001).

3. Results

(1) Among parametric models, the graded model consistently outperforms any divide-by-total model in both a calibration and a cross-validation sample. This is true for both the normal graded model estimated by limited information procedure or the logistic graded model estimated by full information procedures (Fig. 1).

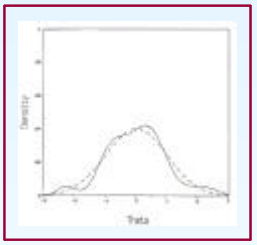
(2) There are only small differences between the logistic graded model estimated by full information ML and the normal graded model estimated by limited information procedures (Fig 2).

(3) Among the divide-by-total models, the more parameters a model has, the better it will fit a calibration sample. When a cross-validation sample is employed the nominal model not always outperforms the ordinal model, although the former has more parameters. Thus, when fitting ordinal data, a model for nominal data does not necessarily yield a better fit. Furthermore, we found that the Rasch-type partial credit model yields the poorest fit to these data. This is not surprising, as the SPSI-R items were not chosen to be equally discriminating (Fig. 1).

(4) The MFS non-parametric model consistently outperforms all parametric model in the calibration sample. This is as it should be as it has a considerably larger number of parameters. However, it needs not necessarily fit better than the graded model in a cross-validation sample. This conflicts with the results of Chernyshenko et al.'s (2001) study. We conjecture that the different results concerning the appropriateness of the unidimensional graded model to fit personality scales are the result of different amounts of multidimensionality present in their data and ours. For scales that are substantially unidimensional –like those considered here- the unidimensional graded model may be an appropriate model. For scales with moderate amounts of multidimensionality, one must resort to a non-parametric unidimensional model which essentially projects the multidimensional space into a unidimensional one (Levine, 1994).

Figure 2.

Plot of the density of the PPO latent trait estimated non-parametrically using the MFS model (solid line) vs. a N(0,1) density (dashed line)



4. Recommendations for modeling personality data using IRT

(1) If it is of interest to fit a unidimensional model, the best choice seems to be a graded model, regardless of whether it is estimated using limited or full information procedures.

(2) If a unidimensional graded model does not yield a satisfactory fit to the data, and a unidimensional model is of substantive interest, then it may be worth considering employing a non-parametric unidimensional model, such as Levine's (1984) MFS.

(3) If multidimensionality is suspected and it is of substantive interest to model it, then a multidimensional normal graded model is called for. This model can be readily estimated using limited information methods with Lisrel or Mplus.

References

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or morenomial categories. Psychometrika, 37, 29-51.
Chernyshenko, O.S., Stark, S., Chan, K.-Y., Drasgow, F. & Williams, B. (2001). Fitting item response theory modelsto two personality inventories: Issues and insights. Multivariate Behavioral Research, 36, 563-562.
Drasgow, F., Levine, M.V., Tsien, S., Williams, B., & Mead, A.D. (1995). Fittingpolychotomous item response theory models to multiple-choice tests. Applied Psychological Measurement 19, 143-165.
D'Zurilla, T.J., Nezu, A.M. & Maydeu-Olivares, A. (2002). Manual of the Social Problem-Solving Inventory-Revised. North Tonawanda, NY: Multi-Health Systems, Inc.
Levine, M.V. (1994). Every data set that is well fit by a two-dimensional IRT model can be equally well fit by a lower dimensional model. Paper presented at the 59th Annual Meeting of the Psychometric Society . Champaign, IL.
Masters, G.N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.
McDonald, R.P. (1999). Test theory. A unified approach. Mahwah, NJ: Lawrence Erlbaum.
Samejima, F. (1969). Calibration of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement No. 17.
Thissen, D. & Steinberg, L. (1986). A taxonomy of item response models. Psychometrika, 51, 567-577.

UNIVERSITAT DE BARCELONA

Further information: amaydeu@psi.ub.es

Sociedad Española para la Investigación de las Diferencias Individuales (SEIDI)
17th September , 2004.
Universitat de Lleida