Using Item Response Models to investigate Attitudes towards Divorce

Liberato Camilleri, Lara Emma Pace
Department of Statistics and Operations Research
University of Malta
Msida (MSD 06) Malta
E-mail: liberato.camilleri@um.edu.mt

KEYWORDS
Item response theory, Rasch model, Partial-credit model, Rating scale model, Generalized linear latent and mixed model, Adaptive quadrature.

ABSTRACT

Item Response Theory (IRT) is a form of latent structure analysis that is used to analyze binary or ordinal response data. IRT models are used to evaluate the relationships between the latent trait of interest and the items measuring the trait. Several IRT models will be fitted to assess the factors that lead to divorce in the Maltese Islands. The 1-PL and 2-PL logistic Rasch models are used for dichotomous responses, whereas the 1-PL rating scale and 1-PL partial-credit models are used for polytomous responses. All the models are fitted using the generalized linear latent and mixed modeling (GLLAMM) framework. The `gllamm` directive estimates parameters by maximum likelihood using adaptive quadrature (Rabe-Hesketh, Skrondal, and Pickles 2002; 2005).

In the 1-PL Rasch model, the probability that a person agrees with a divorce-related item is modeled as a function of subject ability and item difficulty parameters. The major weakness of this model is that the items have the same discrimination parameter. In the 2-PL Birnbaum model, an item-specific weight is added so that the slope of the item response function varies between the items. The 1-PL rating scale model specifies that the items share the same rating scale structure, while the 1-PL partial credit model specifies a distinct rating scale structure for each item.

1. Introduction

A sample of 755 Maltese university students was selected randomly to investigate their perception about factors that may lead to divorce. These twelve factors include: for no reason, long illness, financial problems, conflicts with each other's relatives, inability to have children, unsatisfactory sexual relationships, incompatible personalities, partner dependence on alcohol, lack of love, adultery, partner's homosexual tendencies and domestic violence. These factors were rated on a 4-point Likert scale (strongly agree, agree disagree, strongly disagree. The middling category 'neither agree nor disagree' was purposely excluded to simplify the conversion from an ordinal scale to a dichotomous scale (agree, disagree).

IRT models are appropriate to analyze rating responses to items, which are measured on the same Likert scale. In order to test the items, the rating scores are considered as 'indirect measures of latent ability' (Zheng and Hesketh, 2007), where the test items can either be measured on a dichotomous or an ordinal scale. The data was fitted using four IRT models, which included the 1-PL and 2-PL logistic models for binary responses and the 1-PL rating scale and 1-PL partial-credit models for polytomous responses. All IRT models were fitted using the facilities of GLLAMM, which is a subroutine of STATA.

To satisfy the Rasch model assumptions, item homogeneity and unidimensionality were tested for the twelve items. The Andersen (1973) conditional likelihood ratio test was used to test for item homogeneity and the splitter-item technique by Molenaar (1983) was applied to test the assumption of unidimensionality. The STATA directive `raschtest` was used to conduct the tests and produce graphical displays.

2. Theory

Let Y_{ik} denote the binary response of the i^{th} person on item k for $i = 1,...,n$, where Y_{ik} takes the values 1 or 0 to signify agreement/disagreement. The IRT model intercept θ_i is the person parameter since it varies randomly across all persons. It is assumed that θ_i is normally distributed with mean zero and variance σ_θ^2. Subject parameters provide a measurement of any latent variables such as achievement levels, abilities, skills, cognitive processes, developmental phases, attitudes, motivations, and personality traits (De Boeck and Wilson, 2005). The item predictor denoted by Z_{kh} for $k = 1,...,K$ and $h = 1,...,H$ represents the value of item k on predictor h. This is a 0-1 indicator, where $Z_{kh} = 1$ when $k = h$ and $Z_{kh} = 0$ when $k \neq h$. The coefficients of the item parameters, denoted by δ_k are fixed since they do not vary across subjects. The item response function (IRF) yields the probability $f_k(\theta_i)$ that the i^{th} person agrees with item k. The main objective is to relate the value of the ability θ_i to the corresponding probability of observing an 'agree' response given the value of δ_k.

One-and two-parameter logistic models

The linear component η_{ik} of the Rasch model is given by:

$$\eta_{ik} = \delta_k + \theta_i$$

where $\delta_k = \sum_{h=1}^{H} \delta_h z_{kh}$

In practice, the sign of the item parameter δ_k is reversed to a negative value to be interpreted as the item difficulty. So, the resulting equation is given by:

$$\eta_{ik} = \theta_i - \delta_k$$

Rasch's probabilistic model for the i^{th} person agreeing with test item *k* is given by:

$$P(Y_{ik} = 1|\theta_i) = \frac{\exp(\theta_i - \delta_k)}{1 + \exp(\theta_i - \delta_k)}$$

Birnbaum (1968) proposed the two-parameter logistic (2-PL) model, by adding a slope parameter, λ_k to the intercept parameter θ_i. The 2-PL model generalizes the 1-PL model by relaxing the fixed discrimination parameter assumption across items and assumes a free slope parameter for each item.

Birnbaum's probabilistic model for the i^{th} person agreeing with test item *k* is given by:

$$P(Y_{ik} = 1|\theta_i) = \frac{\exp[\lambda_k(\theta_i - \delta_k)]}{1 + \exp[\lambda_k(\theta_i - \delta_k)]}$$

Partial Credit Model

The partial credit model (Masters, 1982) is an extension of the Rasch binary response model to a polytomous response model comprising R_k ordered categories for some item *k*. The probability that the i^{th} subject chooses rating score *y* for item *k*, for $r = 0,...,y$ and $j = 0,...,R_k$ is given by:

$$P(Y_{ik} = y) = \frac{\exp[\sum_{r=0}^{y}(\theta_i - \delta_{kr})]}{\sum_{j=0}^{R_k}\exp[\sum_{r=0}^{j}(\theta_i - \delta_{kr})]}$$

where δ_{kr} is the r^{th} threshold location of item *k* on a latent continuum and corresponds to the intersection of the r^{th} and $(r-1)^{th}$ probability curves, while θ_i is the location of the i^{th} subject on the same continuum. A feature shared by both the dichotomous and polytomous Rasch models is that all subject and item parameters (θ_i, δ_{kr}) are locations on the variable being measured. For computational convenience, the value of δ_{k0} is chosen to satisfy $\sum_{r=0}^{0}(\theta_i - \delta_{kr}) = 0$.

Rating scale model

The rating scale model is a special case of the partial credit model under the constraint on the item parameters given by:

$$\delta_{kr} = \delta_r + \tau_r$$

where δ_k are item location parameters, which vary between the items, while the threshold parameter τ_r are kept constant across items and depend solely on the response categories. These threshold parameters define the boundary between the different categories of the rating scale, relative to each item's trait location. The rating scale model is appropriate when all items have the same number of categories *R* and are all equally spaced. The probability that the i^{th} subject chooses rating score *y* for item *k*, for $r = 0,...,y$ and $j = 0,...,R$ is given by:

$$P(Y_{ik} = y) = \frac{\exp[\sum_{r=0}^{y}(\theta_i - (\delta_k + \tau_r))]}{\sum_{j=0}^{R}\exp[\sum_{r=0}^{j}(\theta_i - (\delta_k + \tau_r))]}$$

3. Results of 1-PL and 2-PL Rasch models

The Friedman test was used to compare mean rating scores provided to each factor (item). Domestic violence had the largest mean rating score (3.09) implying higher sentiment in favour of divorce. This is followed by partner's homo-sexual tendency (2.79), adultery (2.78), lack of love (2.46), partner's dependency on alcohol (2.15), incompatible personality (1.83), unsatisfactory sexual relationship (1.22), inability to have children (0.74), conflicts with each other's relatives (0.62), financial problems (0.48), lengthy illness (0.44) and for no reason (0.18). The differences between most of the mean rating scores were significant at the 0.05 level of significance.

	1-PL Rasch Model		2-PL Rasch Model	
For no reason	δ_1	6.67	δ_1	4.13
	λ_1	1	λ_1	1
Lengthy illness	δ_2	5.67	δ_2	4.65
	λ_2	1	λ_2	1.84
Financial problems	δ_3	5.63	δ_3	4.72
	λ_3	1	λ_3	1.91
Conflicts with relatives	δ_4	5.33	δ_4	4.63
	λ_4	1	λ_4	2.05
Inability to have children	δ_5	4.28	δ_5	2.85
	λ_5	1	λ_5	1.36
Unsatisfactory sexual relation	δ_6	2.65	δ_6	2.05
	λ_6	1	λ_6	2.04
Incompatible personalities	δ_7	0.49	δ_7	0.25
	λ_7	1	λ_7	3.15
Alcohol problem	δ_8	-0.21	δ_8	-0.45
	λ_8	1	λ_8	2.34
Lack of love	δ_9	-1.09	δ_9	-2.11
	λ_9	1	λ_9	5.01
Adultery	δ_{10}	-2.07	δ_{10}	-3.04
	λ_{10}	1	λ_{10}	4.04
Homosexual tendencies	δ_{11}	-2.37	δ_{11}	-2.92
	λ_{11}	1	λ_{11}	3.31
Domestic violence	δ_{12}	-3.01	δ_{12}	-6.09
	λ_{12}	1	λ_{12}	5.89

Table 1: Estimates of Item Difficulty and Discrimination Parameters

Table 1 displays the estimates of the item difficulties and the discrimination parameters of the 1-and 2-PL Rasch models. The 1-PL Rasch model has a fixed discrimination parameter of 1 across all items, whereas in the 2-PL Rasch model, each item has its own discrimination parameter to determine how well an item discriminates among different trait levels. Low values of the item parameters imply higher probability that one is in favour of divorce. In fact, both the 1-PL and 2-PL Rasch models show that the estimated coefficients for the first seven items are positive implying that most subjects were against divorce for these seven factors. On the other hand, the estimated coefficients of the last five items are negative implying a higher sentiment in favour of divorce for these five factors.

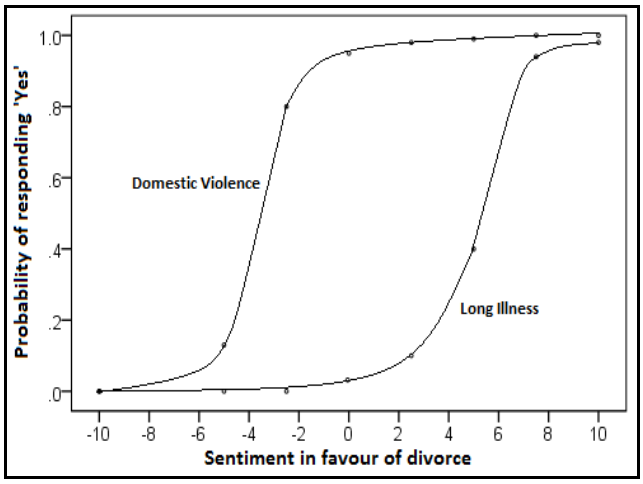

Figure 1: Item-Characteristic Curves for the 1-PL Rasch Model

Figure 1 shows the item-characteristic curves of the 1-PL Rasch model for two selected items, Lengthy illness and Domestic violence. These curves display that respondents have a higher sentiment in favour of divorce in the presence of domestic violence than lengthy illness. Moreover, a rise in sentiment in favour of divorce increases the probability of agreeing with these two factors. The curves are constrained to be parallel since the slope parameters are set to 1.

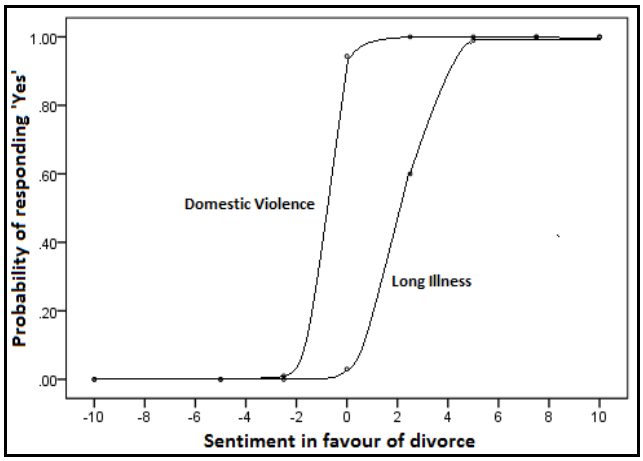

Figure 2: Item-Characteristic Curves for the 2-PL Rasch Model

Figure 2 shows the item-characteristic curves of the 2-PL Rasch model for the same two factors. The curves are not parallel since a slope parameter is estimated for each item. Both models display a higher sentiment in favour of divorce in the presence of domestic violence than lengthy illness.

Given that the 1-PL model is nested within the 2-PL model, a likelihood-ratio chi-square test was used to compare the relative fit of the two models. The 2-PL Rasch model provides a significantly better fit than the 1-PL model since the resulting p-value is less than 0.05 level of significance.

Testing the Rasch Model

The Andersen (1973) conditional likelihood ratio test was used to test for item homogeneity. This test revealed a bad fit for each item since all p-values were less than the 0.05 level of significance. Moreover, the Andersen likelihood ratio test shows that a model with different item difficulty parameters yields a better fit than a model with constrained item parameters. This implies that item homogeneity does not hold for this Rasch model.

The splitter-item technique by Molenaar (1983) was used to check the assumption of unidimensionality. The twelve divorce-related items were tested for the unidimensionality assumption using the splitter-item technique. The procedure involves the division of the sample on the basis of an internal scale criterion called the splitter. Molenaar (1983) predicted that items 'measuring the same latent trait as the splitter will be more difficult for the persons obtaining negative response on the splitter and more easy for persons obtaining a positive response. Items unrelated to this trait should in principle be equally easy for both groups'

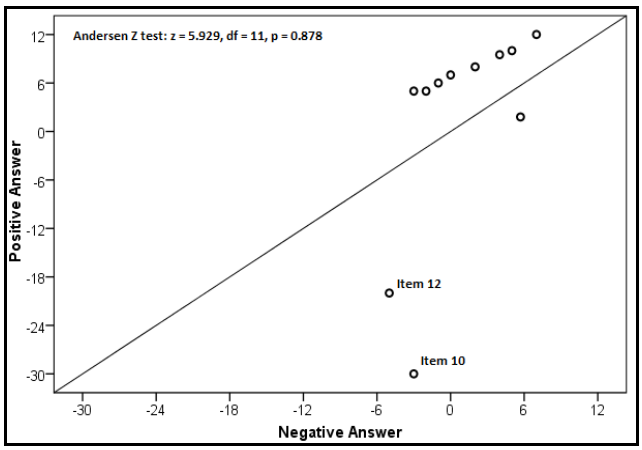

Figure 3: Item difficulty plot for the items with item 3 as the splitter

The splitter in Figure 3 is item 3 (Financial problems). A plot of the item difficulties of one subgroup against the item difficulties of the other shows that items 1 to 9 and item 11 are all grouped together close to the diagonal, indicating that the items were equally difficult for the two subgroups. The Andersen test showed that no deviations were found, providing additional support for the unidimensionality of the set of items $\chi^2(11) = 5.929,\ p = 0.878$. Similar results were produced when using items 1 and 4 as splitters.

Figure 4 shows a contrasting plot when item 10 (Adultery) is used as a splitter. Strong interrelations between the splitter and items 3 and 4 implied that unidimensionality was not present. Similar results were obtained when items 5 to 12 were used as splitters. Thus, the splitter analysis yielded ambiguous results. Some of the plots indicate one underlying dimension, while others indicate two underlying dimensions.

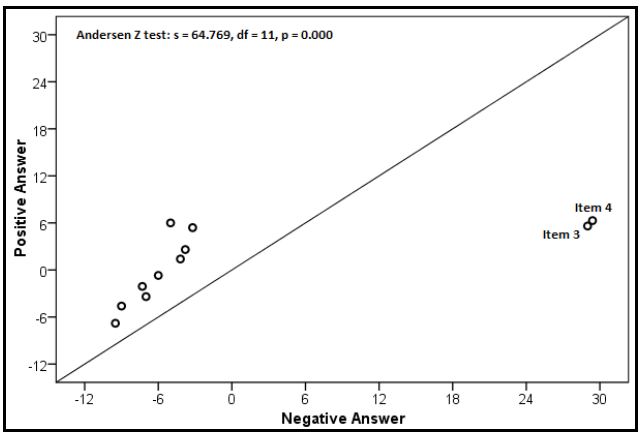

Figure 4: Item difficulty plot for the items with item 10 as the splitter

Unidimensionality was generally revealed when items 1, 3 and 4 were used as splitters, indicating that the items are moderately difficult or relatively easy. The possibility of two dimensions was revealed when items with relatively higher ranking scores, particularly items 10, 11 and 12, were used as splitters. This may be attributed to the fact that difficulty estimates for such items were unreliable. Unidimensionality does not seem to be present for items 5 to12, implying that multidimensional IRT might be a better option.

4. Results of the Partial Credit Model

Figure 5 displays the category probability curves for item 1 (For no reason at all) under the partial credit model. As expected category 1 (Strongly Disagree) is most likely to be observed among low-trait respondents, whereas category 4 (Strongly Agree) is most likely to be observed among high-trait respondents. This implies that most of the students are against adivorce for no reason at all. The response categories 'Strongly Disagree' and 'Disagree' intersect at $\delta_{1,1} = 3.65$ the response categories 'Disagree' and 'Agree' intersect at $\delta_{1,2} = 3.93$ and the response categories 'Agree' and 'Strongly Agree' intersect at $\delta_{1,3} = 3.72.$

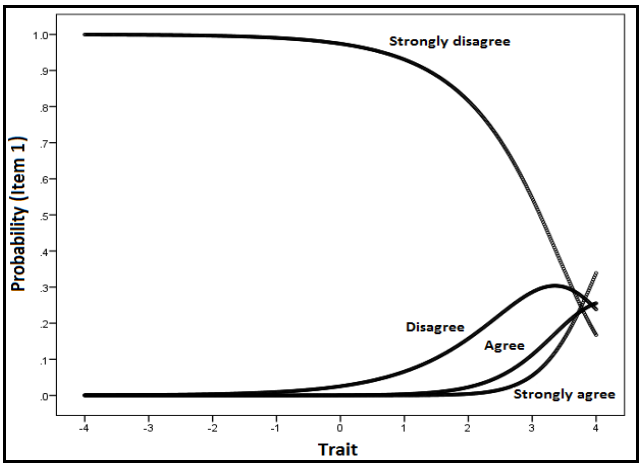

Figure 5: Category Probability Curves for Item 1 (For no reason) under the Partial-Credit Model

Figure 6 represents the category probability curves for item 12 (Domestic violence) under the partial-credit model. Category 1 (Strongly Disagree) is most likely to be observed among very low-trait respondents, while category 4 (Strongly agree) is more likely to be observed among respondents with a high sentiment in favour of divorce. This implies that most students strongly agree with divorce in case of domestic violence. The response categories 'Strongly Disagree' and 'Disagree' intersect at $\delta_{12,1} = -2.86$ the response categories 'Disagree' and 'Agree' intersect at $\delta_{12,2} = -1.42$ and the response categories 'Agree' and 'Strongly Agree' intersect at $\delta_{12,3} = -1.23.$

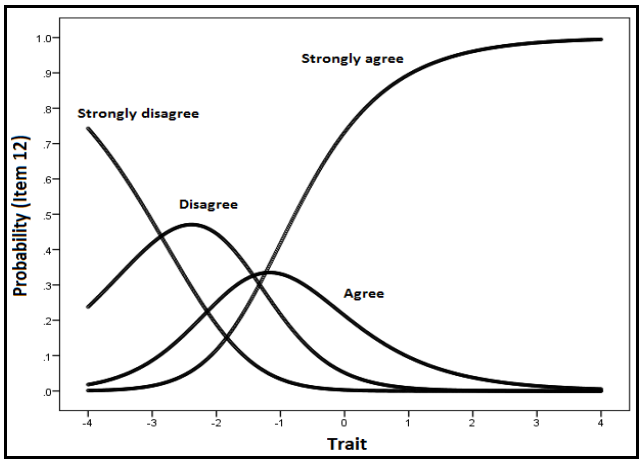

Figure 6: Category Probability Curves for Item 12 (Domestic violence) under the Partial-Credit Model

5. Results of the Rating Scale Model

The category probability curves for items 1 (For no reason) and 12 (Domestic violence) under the rating scale model are given in Figures 7 and 8, respectively.

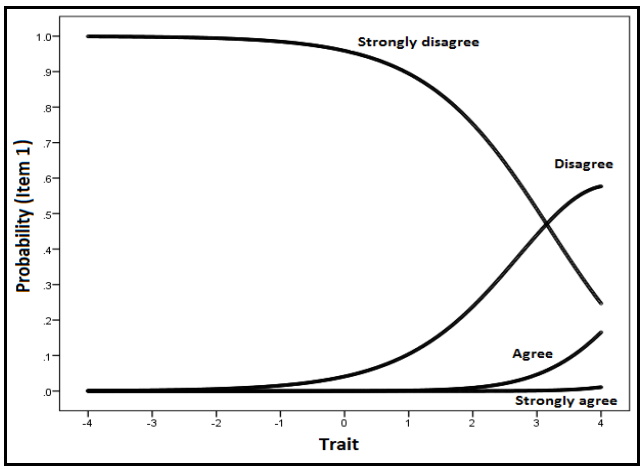

Figure 7: Category Probability Curves for Item 1 (For no reason) under the Rating Scale Model

Figure 7 represents the category probability curves for item 1 (For no reason at all) under the rating scale model. As expected, category 1 (Strongly Disagree) is most likely to be observed among high-trait respondents, whereas category 4 (Strongly Agree) is most likely to be observed among low-trait respondents. This implies that most of the students are not in favour of divorce if it occurs for no reason at all. The value of the latent trait, which is the intersection of the first and second probability curves, is δ_1. The response categories

'Strongly Disagree' and 'Disagree' for item 1 intersect at $\delta_1 = 3.15$ the response categories 'Disagree' and 'Agree' intersect at $\delta_1 + \tau_2 = 5.25$ and the response categories 'Agree' and 'Strongly Agree' intersect at $\delta_1 + \tau_3 = 6.72$.

Figure 8 represents the category probability curves for item 12 (Domestic violence) under the rating scale model. Category 1 (Strongly Disagree) is most likely to be observed among very low-trait respondents, whereas category 4 (Strongly Agree) is most likely to be observed among high-trait respondents. This implies that most of the students strongly agree with divorce in case of domestic violence. The value of the latent trait is the estimated step parameter δ_{12}. In this case, the response categories 'Strongly Disagree' and 'Disagree' intersect at $\delta_{12} = -4.09$, the response categories 'Disagree' and 'Agree' intersect at $\delta_{12} + \tau_2 = -2.00$, and the response categories 'Agree' and 'Strongly Agree' intersect at $\delta_{12} + \tau_3 = -0.53$,

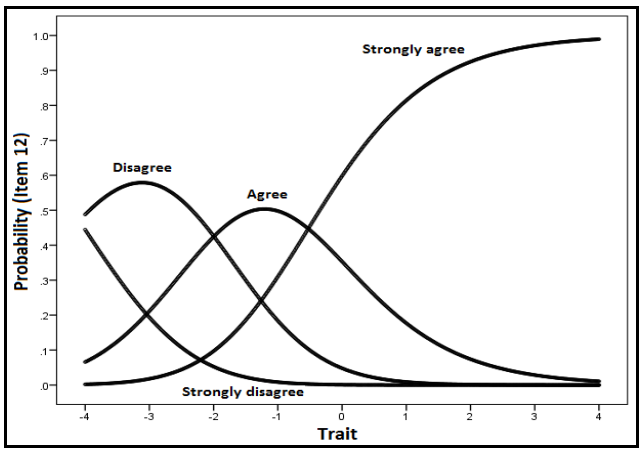

Figure 8: Category Probability Curves for Item 12 (Domestic violence) under the Rating Scale Model

Given that the rating scale model is nested within the partial-credit model, a likelihood-ratio chi-squared test was used to compare the models. The parameter constraints imposed by the rating scale model are clearly rejected. This indicates that the partial credit model provides a better fit than the rating scale model.

6. Conclusion

The 1-PL and 2-PL Rasch models were used to analyze binary response items, while the partial-credit and rating scale models were used to analyze ordered polytomous response items. All models were fitted using the facilities of GLLAMM. The estimated item parameters of all four IRT models were very similar. The estimated coefficients for the first seven items were positive implying that most students were against divorce if there is no sufficient reason, long illness, financial problems, conflicts with each other's relatives, incompatible personalities and inability to have children. On the other hand, the estimated coefficients of the last five items are negative implying that most students are in favour of divorce if domestic violence, adultery, partner's homosexual tendencies, lack of love and alcohol problems is experienced.

References

Andersen, E. B. (1973) A goodness of fit for the Rasch model. Psychometrika 1: 123–140.

Birnbaum, A. (1968) Test scores, sufficient statistics, and the information structures of tests. In Statistical Theories of Mental Test Scores, ed. L. Lord and M. Novick, 425–435. Reading, MA: Addison–Wesley.

De Boeck, P. and Wilson, M.R. (2005) Explanatory Item Response Models: A Generalized linear and nonlinear approach. New York: Springer.

Hambleton, R.K. and van der Linden, W.J. (1982) Advances in Item Response Theory and applications: An introduction. Applied Psychological Measurement, 6, 373-378.

Hambleton, R.K. and Cook, L.L. (1977) Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 14, 75-96.

Hardouin, J. (2007) Rasch Analysis: Estimation and tests with raschtest. The Stata Journal, 7, 22-44.

Molenaar, I. W. (1983) Some improved diagnostics for failure of the Rasch model. Psy-chometrika 48: 49–72.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004) GLLAMM Manual. U.C. Berkeley Division of Biostatistics Working Paper Series, 160, 1-138.

Rabe-Hesketh, S., and A. Skrondal. (2005) Multilevel and Longitudinal Modeling Using Stata. College Station, TX: Stata Press.

Rosenbaum, P.R. (1984) Testing conditional independence and monotonicity assumptions of Item Response Theory. Psychometrika, 49, 425-435.

Skrondal, A. and Rabe-Hesketh, S. (2004) Generalized latent variable modeling: Multilevel, Longitudinal, and Structural Equation Models. Chapman and Hall/CRC.

Skrondal, A., Rabe-Hesketh, S. and Pickles, A. (2004). GLLAMMing with gllamm. Lab Examples for Workshop at the 19th International Workshop on Statistical Modeling Firenze, July 2004.

Skrondal, A. and Rabe-Hesketh, S. (2004) Generalized linear latent and mixed models with composite links and exploded likelihoods. Pages 27-39 in A. Biggeri, E. Dreassi, C. Lagazio and M. Marchi (Eds.), Statistical Modelling. Firenze: Firenze University Press.

Van den Wollenberg, A. L. 1982. Two new test statistics for the Rasch model. Psychometrika 47: 123–140.

Van der Linden, W.J. and Hambleton, R.K. (1996) Handbook of modern Item Response Theory. Springer-Verlag New York, Inc.

Zheng, X., Rabe-Hesketh, S. (2007) Estimating parameters of dichotomous and ordinal Item Response Models with gllamm. The Stata Journal, 7, 313-333.

AUTHOR BIOGRAPHY

LIBERATO CAMILLERI studied Mathematics and Statistics at the University of Malta. He received his PhD degree in Applied Statistics from Lancaster University. His research specialization areas are related to statistical models, which include Generalized Linear models, Latent Class models, Multilevel models and Mixture models. He is an associate professor and Head of the Statistics department at the University of Malta.