

Combining Multilevel and Rasch Model for the Evaluation of a Service

*La Valutazione di un Servizio Utilizzando il Modello di Rasch
Congiuntamente al Modello Multilevel*

Maria Chiara Zanarotti

Università Cattolica, L.go Gemelli, 1, 20123, Italy, e-mail: chiara.zanarotti@unicatt.it

Laura Pagani

Università di Udine, Via Treppo, 18, 33100, Italy, e-mail: laura.pagani@dss.uniud.it

Riassunto: Negli ultimi anni è cresciuta in modo considerevole la consuetudine di rilevare i giudizi degli utenti dei servizi più disparati: dai servizi sanitari a quelli delle telecomunicazioni, dai servizi dei trasporti a quelli della grande distribuzione. Obiettivo comune di queste indagini è, in generale, quello di ottenere delle informazioni relativamente alla qualità del servizio offerto. Il problema che si pone è quello della misurazione della qualità (del servizio) che, come è noto, costituisce un'entità astratta ad esso associata. Dal momento che le risposte agli *items* dei questionari riflettono, oltre alla qualità del servizio, anche le caratteristiche soggettive degli utenti, è necessario scorporare queste due componenti tra loro. Uno strumento utile in questo senso è il modello di Rasch, che viene di seguito utilizzato congiuntamente al modello multilevel.

Keywords: Rasch Models, Multilevel Models, Ordinal Data.

1. Introduction

In the evaluation of a service the main difficulty is how to measure its quality. Quality is not directly observable and measurable, but is a latent variable: to obtain information about its *magnitude* it is necessary to construct a suitable measure tool. Usually in practice researchers carry out a questionnaire formed by a few items relating to different aspects of the service, and customers compile that questionnaire. Typically answers are ordinal categories that vary from “very dissatisfied” (or very insufficient, or strongly disagree) to “very satisfied” (or very good, or strongly agree), and these ordinal data are influenced by the quality level of the particular aspect each item concern, but also by individual, subjective characteristics of the respondents. The statistician so has to elaborate data that not only are expressed through an ordinal categorical scale, but result from all the previous mentioned aspects. A desirable step for similar data is to transform this ordinal and composite information to obtain an universal measure on an interval scale. A possible solution to obtain this kind of measure is represented by the Rasch model. In 1960 Georg Rasch (Rasch, 1960) suggested a model that starting from raw scores resulted from items is able to construct linear continuous measures of both items and subjects. As pointed out by a lot of authors (see, for example, Molenaar and Fisher, 1995; Bond and Fox, 2001), this model represents a very appealing way to obtain universal, objective measures in social

sciences. In this paper, after a brief presentation of the methodology used, the Rasch model is used to obtain items and subjects measures (so called *item and person parameters*) and these measures are introduced in multilevel models to study the effectiveness of university courses in an Italian University.

2. The Rasch measures of item and subjects

The Rasch model was originally introduced in psychometric field for the study of misreading in oral test, but because of its utility, importance, simplicity and flexibility, it has an increasing success in other applied field (in social sciences as well as in medicine). To better understand the meaning of Rasch measures, first consider the dichotomous model, that is the model obtained when each item has only two possible responses (true or wrong, generally coded 1 and 0). Through Rasch model is possible to construct measures based on probabilistic relations between any item's difficulty and any persons' ability. The *idea* underlying the Rasch model is the following: the more difficult one item is, the more probable is only person with high ability gives correct answer to it; the more able one person is, the more probable is that his responses are correct for a lot of items, particularly for the simplest ones. If X_{ij} denote response of person i ($i=1, \dots, I$) to item j ($j=1, \dots, J$), then $X_{ij}=1$ means person i give correct answer on item j , and $X_{ij}=0$ means the contrary, *i.e.* incorrect answer. The Rasch model supposes that the probability $P(X_{ij}=1)$ is a function of the difference between ability of person i and difficulty of item j , *i.e.*: $P(X_{ij}=1)=f(\beta_i-\delta_j)$. The parameters β_i ($i=1, \dots, I$) are called *person parameters*, and they are the Rasch (linear continuous) measures of *person ability*; the parameters δ_j ($j=1, \dots, J$) are called *item parameters*, and they are the Rasch (linear continuous) measures of *item difficulty*. As function $f(\cdot)$, Rasch used the logistic transformation, *i.e.*:

$$P(X_{ij} = 1) = \frac{\exp\{\beta_i - \delta_j\}}{1 + \exp\{\beta_i - \delta_j\}} \quad (1)$$

so that: $\ln\{P(X_{ij}=1)/[1 - P(X_{ij}=1)]\} = \beta_i - \delta_j$. By means of the parameters β_i and δ_j , with Rasch model is possible to represent on an interval-scale both items and persons (their *difficulty* and *ability* measures, respectively) along a unique dimension.

Row and column total scores are used to obtain maximum likelihood estimates of persons and items parameters. The parameters estimation is simple because of the *parameter separation theorem*, that hold for the entire family of Rasch models (see, for example, Fisher et al. 1995). This paper focuses on the ordinal extension of the Rasch model obtained when each items has K ordinate response categories; in particular the so called *Partial Credit Model* (PCM) is used. For this model, probability that person i ($i=1, \dots, I$) chooses response k ($k=1, \dots, K$) for item j ($j=1, \dots, J$) instead of response $k-1$, is:

$$P(X_{ij} = k \mid \beta_i, \delta_j, \tau_{k(j)}) = \frac{\exp\{\beta_i - \delta_j - \tau_{k(j)}\}}{\sum_{k=1}^K \exp\{\beta_i - \delta_j - \tau_{k(j)}\}} \quad (2)$$

In PCM there is a new set of $J \times (K-1)$ parameters $\tau_{k(j)}$ (called *thresholds*) that give a measure of the *difficulty* to endorse each response category over the previous one. The use of Rasch models in quality evaluation of a service, as already made by other authors (for example, Bertoli-Barsotti and Franzoni, 2001), implies a new meaning for the parameters: person parameters β_i ($i=1, \dots, I$), in this context, represent individual satisfaction (instead of person ability, as before). High values of person parameter mean highly satisfied persons, while low values mean the reverse; by item parameters δ_j ($j=1, \dots, J$), on the other hand, is possible to measure and ordinate items from the one with the best quality to the one with the worst quality. Note that the scale for item parameters has to be read on the contrary: high values means that it is very difficult to endorse high categories for these items, and low values mean the reverse.

The PCM formulation of Rasch model is used in this paper to analyse a data set constituted by 345 questionnaires administrated during year 2000 to students of the Faculty of Economy, University of Udine (Italy), to evaluate 59 university courses. The questionnaire was formed by 3 sections: a) information on the students' characteristics (age, gender, type of high school attended, and so on); b) 18 general items about teacher and course characteristics; c) a last item on global satisfaction. The responses to each item are collected through a four points categorical scale (*not at all satisfied, dissatisfied, satisfied, very satisfied*). Data were analysed with RUMM 2010 (RUMM Laboratory Pty Ltd), a standard software for the Rasch analysis.

3. Rasch measures and multilevel models in evaluation of effectiveness of university courses

In this section we use the person parameters (PP) obtained with the PCM Rasch model as a dependent variable in a multilevel model to assess courses effectiveness. There is a large literature on multilevel models (see, for example, Goldstein, 1995) and on multilevel models used to evaluate public sector activities (Goldstein and Spiegelhalter, 1996, among others). The aim here is to evaluate the effectiveness of university courses considering questionnaires ($i=1, \dots, 345$) as level-1 units and courses ($m=1, \dots, 59$) as the level-2 units. Since PP are continuous measures of students satisfaction, after checking PP normality with a Kolmogorov-Smirnov test ($D=0.0614$; p -value 0.066), a standard multilevel model is used. Several models with different sets of explanatory variables and with different assumptions about the random part have been fitted. The final model is a two-level random intercept one with the following hypothesis:

$$Y_{im} = \alpha_0 + \sum_{h=1}^H \alpha_h w_{him} + \sum_{k=1}^K \beta_k z_{km} + U_{0m} + R_{im}, \quad U_{0m} \sim N(0, \sigma_u^2), \quad R_{im} \sim N(0, \sigma_e^2)$$

Results (see table 1), for fixed parameters, show that: the mark “Good” as final grade of secondary school, to be a female student, not to be first year student, to attend a compulsory course and to have a female teacher imply a negative effect on the PP; to attend over the 60% of the lessons, on the contrary, has a positive effect on the PP.

The main result for the random parameters is the significance of the second level variance. This means that PP variability (*i.e.* variability of the level of satisfaction) depends also on the differences among courses (there is a course effect on PP).

Moreover, the intra-course correlation says that the 30% of the residual PP variability is due to the course effect.

Table 1: *Parameters estimation for two-level random intercept model*

Fixed Parameters	Estimate	Std. Err.	Random Parameters	Estimate	Std. Err.
Constant	1.378	0.384	Variance between questionnaires	1.684	0.141
Good	-0.374	0.158	Variance between courses	0.722	0.221
Female Student	-0.427	0.153	Intra-course correlation	0.300	
More than second	-0.364	0.191			
Over 60%	0.741	0.252			
Compulsory	-0.584	0.291			
Female teacher	-0.535	0.303			

4. Concluding remarks

The Rasch Model allowed to obtain a measure, on an interval scale, for every student (PP parameter). This measure, that reflects student satisfaction, has been introduced in a two levels multilevel model to study effects of some covariates on PP. This procedure permits to use all the information included in the questionnaire to measure students satisfaction and then to investigate relations with covariates. In standard multilevel approach, in fact, only one ordinal variable at time can be considered as dependent, so covariate effects are investigated for each response variable separately.

Further research will try to take into consideration also the other parameters obtained by Rasch model, *i.e.* the item parameters, that in the present work have not been considered in the model.

References

- Bertoli-Barsotti L., Franzoni S. (2001). Analisi della soddisfazione del paziente in una struttura sanitaria: un caso di studio. *Università Cattolica di Milano, Serie E.P. N. 104*, 1-17.
- Bond T.G., Fox C.M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates, London.
- Fischer, G.H., Molenaar, I.W. (1995). *Rasch Models. Foundations, Recent Developments, and Applications*. Springer-Verlag, New York.
- Goldstein, H., (1995). *Multilevel Statistical Models*. Edward Arnold, London.
- Goldstein, H., Spiegelhalter, D. J. (1996). League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. *J. R. Stat. Soc. A*, 159, Part 3, 385-443.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. *Danish Institute for Educational Research*, Copenhagen.