Benefits of the Erasmus mobility experience: a discrete latent variable analysis

Benefici dell'esperienza Erasmus: un'analisi a variabili latenti discrete

Silvia Bacci, Valeria Caviezel and Anna Maria Falzoni

Abstract Internationalization of higher education has become a priority in the European education policy. For this reason, research in this area is expanding with the aim of understanding motivations and potential benefits of international mobility. In such a context, an online survey addressed to about 1,600 students with Erasmus mobility experiences was conducted by the University of Bergamo (IT). Two latent traits, that is, the impact of the Erasmus experience on the student's abilities and the student's satisfaction for this experience, are analyzed through a two-dimensional latent class Item Response Theory model under a concomitant variable approach. The twofold issue concerning the choice of the optimal number of latent classes as well as the selection of significant covariates is specially addressed.

Abstract *L'internazionalizzazione della formazione universitaria rappresenta una priorità della politica europea dell'istruzione. Per questo motivo, la ricerca in questo ambito è in espansione, al fine di comprendere le motivazioni e i potenziali benefici della mobilità internazionale. L'Università di Bergamo ha condotto un'indagine online rivolta a circa 1,600 studenti con esperienza di studio all'estero. Due variabili latenti, cioè l'impatto dell'esperienza Erasmus sulle abilità degli studenti e il grado di soddisfazione per l'esperienza vissuta, sono oggetto di analisi tramite un modello Item Response Theory bidimensionale a classi latenti con approccio di variabili concomitanti. Particolare attenzione è posta al duplice problema della scelta del numero di classi latenti e della selezione delle covariate significative.*

Silvia Bacci
Dipartimento di Economia, Università di Perugia, Via A. Pascoli 20, 06123 Perugia e-mail: silvia.bacci@unipg.it

Valeria Caviezel
Dipartimento di Scienze Aziendali, Economiche e Metodi Quantitativi, Università di Bergamo, Via dei Caniana 2, 24127 Bergamo e-mail: valeria.caviezel@unibg.it

Anna Maria Falzoni
Dipartimento di Scienze Aziendali, Economiche e Metodi Quantitativi, Università di Bergamo, Via dei Caniana 2, 24127 Bergamo e-mail: anna-maria.falzoni@unibg.it

Key words: Concomitant variable approach, Credit mobility, Graded Response Model, Item Response Theory, Latent class model

1 Introduction

Internationalization of higher education has become a priority in the European education policy. According to the strategic objectives of Europe 2020, "*EU average of at least 20% of higher education graduates should have had a period of higher education-related study or training abroad, representing a minimum of 15 ECTS credits or lasting a minimum of three months*" (EU Council of Ministers of Education, November 29, 2011). Since it began in 1987/1988, the world's most successful student mobility programme, the Erasmus programme, has provided over three million European students with the opportunity to go abroad and study at a higher education institution or train in a company [5].

In the last years, a number of studies has analysed motivations and potential benefits of international mobility. Research findings seem to show that studying abroad positively affect students personality development and improve their language competencies and soft skills (such as problem-solving and decision-making skills). Relatedly, mobility students seem to have better labour market prospects, in particular abroad ([3], [4]).

To analyse students' motivations and the fullfilment of expectations regarding the study international experience, we asked 1,576 students, enrolled in the University of Bergamo with a credit mobility experience during the a.y. from 2008/09 to 2014/15, to answer a questionnaire. Administrative information at enrolment (e.g., age, gender, field of study, etc.) is also available for each student.

In this contribution we focus on two main aspects related to the Erasmus experience: the impact of the Erasmus experience on the student's skills and the student's fulfilling the expectations for this experience. Both these elements represent latent variables, which are measured through a set of polytmously-scored items. Our aim consists in measuring these latent variables, providing evidence of one or more problematic items. Moreover, we intend to detect individual characteristics that significantly explain the level of latent variables. For these aims, we formulate a bidimensional Latent Class Item Response Theory (LC-IRT) model ([1], [2]) with a concomitant variable approach [6].

2 Erasmus mobility data: description

To assess the international experience of the students of the University of Bergamo an ad hoc questionnaire was prepared, which was organised in three sections: Decision to study abroad, International experience and Coming back. Student's individual characteristics were also collected, such as gender, parents' level of education,

parents' employment status, previous international experiences of him/herself or family, student's current employment status. The survey involved all the 1,576 students, which spent one/two semesters abroad for an Erasmus or Extra EU program from a.y. 2008/09 to a.y. 2014/15; the response rate was 48.6% (766 students).

The sample includes students from all the five fields of study of the University of Bergamo: Foreign Languages (45.7%), Economics (28.6%), Human and Social Sciences (13%), Engineering (9.6%), and Law (3.1%). They are enrolled in a bachelor degree (64%), a master degree (33.8%) and in a five-years degree of study (2.2%). As far as international experience is concerned, about half of the students spent the fall semester abroad (48.2%) and the 20.4% of students the spring semester; for just less than a third of students (31.4%) the experience lasted for the whole academic year. The preferred destination was Spain (28.1%), followed by Germany (17.7%), United Kingdom (15.8%), and France (13.7%); other European countries were overall chosen by the 18.7% of students, whereas the Extra-UE destinations were USA (2.7%), China (2.1%) and Australia (1.2%).

3 The statistical model

Given a set of J ordered polytomous items for the measurement of some latent traits about the Erasmus experience of n students, let $\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_s)'$ be the vector of latent variables that drive the response process and let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)'$ denote one of its possible realizations. Vector $\boldsymbol{\Theta}$ is assumed to have a discrete distribution with k support points, denoted by $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$, and mass probabilities (or weights) $\pi_{i1}, \ldots, \pi_{ik}$ $(i = 1, \ldots, n)$. From an interpretative point of view, each support point detects a group (latent class) of individuals that share a common level of the latent traits, whereas the corresponding mass probability denotes the probability of belonging to each latent class according to the individual characteristics.The mass probabilities are assumed to depend on a set of p individual characteristics $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})'$ $(i = 1, \ldots, n)$, that is, $\pi_{iu} = p(\boldsymbol{\Theta} = \boldsymbol{\xi}_u | \mathbf{X}_i = \mathbf{x}_i)$ $(u = 1, \ldots, k)$, through a multinomial logit model

$$\log \frac{\pi_{iu}}{\pi_{i1}} = \delta_{0u} + \mathbf{x}_i' \boldsymbol{\delta}_{1u}, \quad i = 1, \ldots, n;\ u = 2, \ldots, k,$$

where δ_{0u} and $\boldsymbol{\delta}_{1u}$ denote, respectively, the class-specific intercept term and the class-specific vector of regression coefficients.

As concerns the response process, the probability that student i $(i = 1, \ldots, n)$ answers category y $(y = 1, \ldots, l_j - 1)$ to item j $(j = 1, \ldots, J)$ is modelled through a multidimensional Latent Class Graded Response Model (LC-GRM; see [1] and [2] and references therein for details on the model specification and the estimation):

$$\log \frac{p(Y_{ij} \geq y | \boldsymbol{\Theta} = \boldsymbol{\theta})}{p(Y_{ij} < y | \boldsymbol{\Theta} = \boldsymbol{\theta})} = \gamma_j \sum_{d=1}^{s} (z_{jd} \theta_d - \beta_{jy}), j = 1, \ldots, J;\ y = 1, \ldots, l_j - 1,$$

with z_{jd} dummy variable equal to 1 if item j measures latent trait θ_d and 0 otherwise, and γ_j and β_{jy} item discriminating and difficulty parameters, respectively, having the usual interpretation as in the item response theory context.

The manifest distribution of $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})'$ follows as

$$p(\mathbf{Y}_i = \mathbf{y}_i) = \sum_{u=1}^{k} p(\mathbf{Y}_i = \mathbf{y}_i | \boldsymbol{\Theta} = \boldsymbol{\theta}_u) \pi_{iu}$$

where, in virtue of the local independence assumption,

$$p(\mathbf{Y}_i = \mathbf{y}_i | \boldsymbol{\Theta} = \boldsymbol{\theta}_u) = \prod_{j=1}^{J} p(Y_{ij} = y | \boldsymbol{\Theta} = \boldsymbol{\theta}).$$

In what follow we assume $s = 2$, indicating with Θ_1 the impact of the Erasmus experience on the student's skills and with Θ_2 the fulfilling the expectations of the Erasmus experience. In addition, we consider $J = 12$ items with three ordered response categories ($l_j = 0, 1, 2$) denoting increasing levels of positive impact or satisfaction; Θ_1 is measured by 7 items and Θ_2 by the remaining 5 items.

A relevant aspect with the model at issue and, more in general, with the class of multidimensional LC-IRT models, to which the LC-GRM belongs, is represented by the choice of the number k of latent classes. The main stream of the literature agrees on using the Bayesian Information Criterion (BIC) to select the value of k. In practice, a common approach consists in estimating the model for increasing values of k, being constant all the other elements of the model, and selecting that value of k corresponding to the minimum BIC. However, it has also to be taken into account that, in the presence of several covariates, the likelihood of the model and the value of BIC are affected by the selection process of covariates. In other words, if one or more covariates are eliminated from the model because of their not significance, the sequence of BIC values would be likely different with consequences on the optimal k. In addition, strategy based on the minimization of BIC index often leads to numerous and small latent classes that are not simple to interpret.

To solve the two above problems, we here propose an iterative procedure consisting in the following steps:

Step 1: estimate the multidimensional LC-GRM without covariates and for increasing values of k until the first increasing BIC is obtained. Select that value of k corresponding to the first relative decrease of BIC smaller than a given (small) threshold α, say $\alpha = 0.01$;

Step 2: given k selected at Step 1, estimate the multidimensional LC-GRM with all the plausible covariates of interest. If all the covariates are significant at a given level (e.g., 5%) stop, otherwise go on with the next step;

Step 3: estimate the multidimensional LC-GRM with covariates selected at Step 2 and for increasing values of k until the first increasing BIC is obtained. Select the value of k as in Step 1. If all the covariates are significant stop, otherwise repeat Step 3.

The procedure finishes when there are not any more changes in the value of k and in the set of significant covariates.

4 Application to Erasmus mobility data: main results

According to the iterative procedure described above, we selected a LC-GRM with $k = 3$ components. As shown in Table 1, class 1 collects students that benefited just a little from the international experience (average weight equal to 21.0%), whereas students with significant advantage from this type of experience belong to class 3 (average weight equal to 33.1%). The remaining part of students (45.9%) is allocated in class 2, showing intermediate levels on both the latent variables.

Table 1 Estimates of support points $\hat{\xi}_{ud}$ and averages of weights $\hat{\pi}_{iu}$ ($d = 1,2$; $u = 1,2,3$)

Latent variable d	Class $u = 1$	Class $u = 2$	Class $u = 3$
Impact on skills (Θ_1, $d = 1$)	-0.951	-0.092	1.043
Satisfaction (Θ_2, $d = 2$)	-0.956	-0.084	1.039
Avg. weights	0.210	0.459	0.331

In Table 2 the distribution of the statistically significant covariates is shown. Covariates related to motivations to study abroad allow us for a clear characterisation of the three classes: improving own *curriculum studiorum* (CV), improving foreign language competencies, studying the culture of a foreign country, curiosity about new challenges, and to enhance future employability characterise class 2 and, mainly, class 3, according to an increasing trend. Instead, the impact on the latent class membership of the use of interactive teaching methods is less clear. It is worth to be noted that the host country as well as many other individual characteristics are not enclosed in the subset of significant covariates.

To conclude, Table 3 shows the item parameter estimates. The most problematic aspects related to the impact of international experience are represented by the learning ability and by the motivation to study: in both cases, the difficulty parameters $\hat{\beta}_{j1}$ and $\hat{\beta}_{j2}$ ($j = 2,7$) are definitely greater with respect to the other items; on the opposite, the adaptability ($j = 5$) is not a problem at all. As far as the fulfilling of the expectations, the benefits of the international experience for the Italian labour market ($j = 8$) is perceived as the most critical aspect, whereas the students' expectations in terms of personality development ($j = 10$) are usually completely satisfied.

Table 2 Distribution of students in the latent classes, for each significant covariate (proportions)

Covariate	Class $u = 1$	Class $u = 2$	Class $u = 3$
Motivations			
Improve CV	0.349	0.453	0.640
Improve foreign language skills	0.671	0.897	0.955
Improve host country knowledge	0.221	0.441	0.702
Curiosity about new challenges	0.732	0.915	0.955
Enhance future employability	0.121	0.306	0.566
Use of interactive teaching methods			
less than in Italy	0.013	0.006	0.025
the same as in Italy	0.182	0.062	0.091
more than in Italy	0.805	0.932	0.884

Table 3 Estimates of item difficulty parameters $\hat{\beta}_{jy}$ and item discriminating parameters $\hat{\gamma}_j$ ($j = 1,\ldots,12$; $y = 1,2$)

Item	$\hat{\beta}_{j1}$	$\hat{\beta}_{j2}$	$\hat{\gamma}_j$
Communication skills ($j = 1$)	0.000	3.497	1.000
Learning ability ($j = 2$)	2.533	6.074	0.789
Foreign language skills ($j = 3$)	-0.757	2.864	1.005
Team working skills ($j = 4$)	0.946	5.012	0.646
Adaptability ($j = 5$)	-1.402	2.365	0.928
Problem solving skills ($j = 6$)	0.254	4.211	0.830
Motivation to study ($j = 7$)	2.704	6.612	0.541
Enhance future employability in Italy ($j = 8$)	0.000	2.196	1.000
Enhance future employability abroad ($j = 9$)	-0.514	1.404	1.518
Personality development ($j = 10$)	-1.256	-0.118	3.325
Foreign language skills ($j = 11$)	-0.633	0.303	3.770
Ability to interact with foreign people ($j = 12$)	-0.530	0.419	3.672

References

1. Bacci, S., Bartolucci, F., Gnaldi, M.: A class of multidimensional latent class IRT models for ordinal polytomous item responses. Commun Stat Theory Methods **43**, 787 – 800 (2014)
2. Bartolucci, F., Bacci, S., Gnaldi, M.: Statistical analysis of questionnaires: A unified approach based on R and Stata. Chapman & Hall/CRC Press, Boca Raton (2015)
3. CHE Consult, Brussels Education Services, Centrum fur Hochschulentwicklung, Compostela Group of Universities, Erasmus Student Network: The Erasmus impact study. European Commission, Brussels (2014)
4. CHE Consult GmbH and CHE Consult Prague s.r.o.: The Erasmus impact study. Regional Analysis. European Commission, Brussels (2016)
5. European Commission: Erasmus - Facts, Figures & Trends. Publications Office of the European Union, Luxembourg (2015)
6. Formann, A. K.: Mixture analysis of multivariate categorical data with covariates and missing entries. Comput Stat Data Anal **51**, 5236 – 5246 (2007)