Investigating Social Integration of Second-Generation Foreign Children in Italy: a Model-Based Indicator

Alessio Buonomo, Stefania Capecchi and Rosaria Simone

Abstract Social inclusion and integration are key topics in social sciences, and they are particularly relevant in the educational field to counteract the experience of inequality conditions.

With specific reference to migrants educational integration, in 2008 the European Commission's Green Paper established several guidelines, stressing the importance of involving migrant communities in schools. As a matter of fact, the educational path of children with a migratory background, especially of those born abroad, can present specific problems (as an instance: the linguistic ones).

Our contribution concerns indicators computed on the basis of information collected as ordinal data, as generally arising from surveys where respondents are asked to manifest their perception about multiple aspects of a latent construct. The interest lays in synthesyzing opinions/evaluations of a complex phenomenon with a rating based on a set of ordered discrete choices.

In order to build a synthetic measure able to compare and cluster results, by individual characteristics as well as by citizenship, we focus on ratings to educational integration questions from the national survey on "Integration of the second generation" (ISG).

Keywords Indicators · Second Generation · Social Integration · Heterogeneity · CUB models

1 Data and questions of interest

Data of interest stem from the "Integration of the second generation" survey (ISG), carried out in 2015 lower and upper secondary schools by the Italian National Institute of Statistics (ISTAT) in collaboration with the Ministry of

Department of Political Sciences
University of Naples Federico II Italy

Education, University and Research (MIUR). Data were released in 2018 and the overall unweighted sample amounted to 68,127 individuals (36.440 Italian citizens and 31.687 foreigners), 48% of which were students in lower secondary schools.

The ISG survey is one of the most reliable sources of information on conditions and educational aspirations of students with migratory backgrounds. This cross-sectional survey is based on a sample of state (public) secondary schools with at least five foreign students. The survey involved mainly young people with a migratory background, but it also allows for comparisons with national students. Immigrant status is based on student citizenship at the time of the interview.

Items of interest for this study are those of question C14 of the ISG questionnaire (20 items), all referred to students' self-assessed behaviours, trust towards classmates and teachers, attitudes towards relatives and study habits. For each item, respondents are asked to express their level of agreement on a 5-point Likert scale, ranging from strong agreement to strong disagreement. Many respondents' covariates are available in the dataset (both ordinal and socio-demographic).

2 Methodology

In building indicators, the risk to aggregate data or indices without saving essential information is always pervasive. Classical proposals often include average operations (arithmetic, geometric, harmonic, etc.) of the expressed ratings, or the selection of the first variable of the principal component analysis (PCA) performed on the data matrix.

Nevertheless, especially in surveys focused on multifaceted and sensitive topics, such as perceptions/opinions on multidimensional issues, the additive indices may disregard to account for inherent heterogeneity/uncertainty in response patterns. This latter feature is meant as both intrinsic indeterminacy in any decision process, and as heterogeneity derived by the presence of subgroups whose response behaviour is not homogeneous.

This study aims at proposing an indicator of second-generation foreign children integration at school in light of the procedure recently introduced by Capecchi and Simone (2019). The paper would provide a synthetic measure able to synthesize subjective evaluations while accounting for both *agreement* towards the items and establishing a direct control for *heterogeneity/uncertainty* in response patterns, relying on the CUB models framework (Piccolo, 2003). For updated references on the class of CUB models, see Piccolo and Simone (2019).

When dealing with this kind of ordinal data, we consider a questionnaire designed to measure a latent trait (as an opinion, for instance) via K observable response variables $R_1, \ldots, R_K$ defined on an ordinal scale with supports the first $m_1, m_2, \ldots, m_K$ integers, respectively. Then, we let: $m_1 = m_2 = \cdots = m_K = m$.

Briefly, for a given m, a CUB model for the random variable R, consisting in the responses to a given item and defined over the support $\{1, 2, \ldots, m\}$, is the probability mass distribution:

$$Pr\left(R = r \mid \boldsymbol{\theta}\right) = \pi \binom{m-1}{r-1} \xi^{m-r} \left(1 - \xi\right)^{r-1} + \left(1 - \pi\right) \frac{1}{m}, \quad r = 1, 2, \ldots, m.$$

The model is well defined when the parameter vector $\boldsymbol{\theta} = (\pi, \xi)'$ is such that: $\pi \in (0, 1]$ and $\xi \in [0, 1]$; thus, the parameter space is the (left open) unit square: $\Omega(\boldsymbol{\theta}) = \Omega(\pi, \xi) = \{(\pi, \xi) : 0 < \pi \leq 1; \ 0 \leq \xi \leq 1\}$. Therefore, each model may be depicted in the unit square by means of a point with coordinates $(1 - \pi, 1 - \xi)$, which are immediately interpretable as measures of uncertainty/heterogeneity and feeling/agreement, respectively.

We assume that a CUB model fits the ordinal responses given to each item in an effective parametric way. We denote by $R_k \sim \text{CUB}\,(\hat{\pi}_k, \hat{\xi}_k)$, for $k = 1, \ldots, K$, the random variable R_k associated to the $k - th$ item whose probability distribution is:

$$Pr(R_k = r | \pi_k, \xi_k) = \pi_k \binom{m-1}{r-1} \xi_k^{m-r} (1 - \xi_k)^{r-1} + (1 - \pi_k) \frac{1}{m}, \quad r = 1, \ldots, m.$$

In this study we implement a *weighted* CUB *model* $\tilde{R} \sim \text{CUB}\,(\tilde{\pi}, \tilde{\xi})$:

$$\tilde{\pi} = \sum_{k=1}^{K} w_k \hat{\pi}_k, \qquad \tilde{\xi} = \sum_{k=1}^{K} w_k \hat{\xi}_k, \tag{1}$$

as a 2-dimensional composite indicator denoted as CI-CUB (Composite Indicator CUB) for the latent trait of interest, that is the educational/social integration of second-generation immigrant children. This choice accounts for both uncertainty/heterogeneity and feeling/agreement; the weighting system can be chosen in order to assign higher weights to the most important items.

Results will be illustrated in detail in the paper, with reference to both individual characteristics and citizenship clusters.

In presence of significant covariates (or when estimating CUB models for subgroups) the impacts of subjective characteristics on the components of the estimated models will be represented and those effects will be discussed. Effective graphical representations in the parameter space would enhance the readability of results also for comparative purposes.

References

- Capecchi, S. and Simone, R. (2019). A Proposal for a Model-Based Composite Indicator: Experience on Perceived Discrimination in Europe. *Social Indicators Research* 141(1), 95-110.
- Piccolo, D. (2003). On the moments of a mixture of uniform and shifted binomial random variables. *Quaderni di Statistica* 5, 85-104

- Piccolo, D. and Simone, R. (2019). The class of CUB models: statistical foundations, inferential issues and empirical evidence. *Statistical Methods & Applications* 28(3), 389-435 (with discussion and rejoinder).