Social Indicators Research (2024) 172:163–190
https://doi.org/10.1007/s11205-023-03261-z

ORIGINAL RESEARCH

An Analysis of the Effect of Streaming on Civic Participation Through a Causal Hidden Markov Model

Francesco Bartolucci[1] · Donata Favaro[2] · Fulvia Pennoni[3] · Dario Sciulli[4]

Accepted: 5 November 2023 / Published online: 1 February 2024

Abstract

We examine the effect of streaming based on ability levels on individuals' civic participation throughout their adult life. The hypothesis we test is that ability grouping influences individuals' general self-concept and, consequently, their civic participation choices across the life course. We employ data from the British National Child Development Study, which follows all UK citizens born during a certain week in 1958. Six binary variables observed at 33, 42, and 51 years of age are considered to measure civic participation. Our approach defines causal estimands with multiple treatments referring to the evolution of civic engagement over time in terms of potential versions of a sequence of latent variables assumed to follow a Markov chain with initial and transition probabilities depending on posttreatment time-varying covariates. The model also addresses partially or entirely missing data on one or more indicators at a given time occasion and missing posttreatment covariate values using dummy indicators. The model is estimated by maximizing a weighted log-likelihood function with weights corresponding to the inverse probability of the received treatment obtained from a multinomial logit model based on pretreatment covariates. Our results show that ability grouping affects the civic participation of high-ability individuals when they are 33 years old with respect to participation in general elections.

Keywords Causal inference · Civic participation · Expectation-maximization algorithm · Generalized propensity score · Voting behavior · Streaming

JEL Classification C31 · C33 · I21

✉ Fulvia Pennoni
fulvia.pennoni@unimib.it

1 Department of Economics, University of Perugia, Via A. Pascoli, 20, Perugia 06123, Italy

2 Department of Economics and Management, University of Padova, Via del Santo, 33, Padua 35123, Italy

3 Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via degli Arcimboldi, 8, Milan 20126, Italy

4 Department of Economic Studies, University of Chieti-Pescara, Viale Pindaro, 42, Pescara 65127, Italy

1 Introduction

Streaming (also known as “tracking" in some countries) is a type of homogeneous grouping that assigns pupils to classes based on an assessment of their general abilities; pupils remain in their streamed class for the majority of studied subjects. Despite comprehensive school reforms that widely reduced or postponed the adoption of streaming during the 1970s, the interest in this practice has significantly rekindled in recent decades. Currently, many countries around the world apply streaming at different stages of the educational pattern.

Educational psychology and social psychology explain how ability grouping in school has effects on the individual’s self-concept formation and how one’s self-concept formed at a young age has repercussions throughout the individual’s life. First, ability grouping has effects on the individual’s academic self-concept. The academic self-concept is a student’s self-perception of her/his academic ability based on personal educational experiences (Marsh & Craven, 2006); it is part of the multidimensional general self-concept, that is, the perception of one’s general ability. Students grouped in high-ability classes experience more instructional inputs, higher self-expectations, and higher expectations from significant others, which allow them to outperform their counterparts in low-ability classes (Pallas et al., 1994). The basking in reflected glory theory (Marsh, 1987; Marsh et al., 1995) supports the idea that a placement on a higher path reinforces self-evaluation. By internalizing the value that society assigns to the group of students they are a part of, students’ self-perception largely depends on the average-ability of their classmates or schoolmates; thus, if a student attends a high-ability class or school, which is generally highly valued by society, that student will have a higher self-concept than a peer who is in a low-ability class. The social dimension of ability grouping suggests that such grouping organizes schools into cohorts of different status and prestige. Placement in low-ability classes is seen as psychologically damaging compared to high-ability classes, which carry more prestige (Gamoran, 1992). However, a different stream of literature shows that one’s self-concept may improve or worsen due to comparison with other students. The “big fish, little pond" theory (Marsh, 1984) suggests that academically gifted students experience worse academic concepts in homogeneous ability groups than in regular, mixed-ability classes. In contrast, educationally disadvantaged students develop higher academic self-concepts in low-ability groups than in regular mixed-ability classes. For a recent analysis of this effect, see Fang et al. (2018).

Empirical analyses mostly confirm the positive effects of belonging to homogeneous ability groups for high-ability students. In contrast, the effect of tracking is negative for low-ability students. This has been confirmed with regard to both student achievement and academic self-concept. High-ability pupils’ achievement in high-ability groups increases compared to that of pupils in heterogeneous groups; in contrast, low-ability pupils’ achievement in low-ability groups decreases when compared to that of low-ability pupils in heterogeneous classes; see Epple and Romano (2011), Wilkinson et al. (2000), Thrupp et al. (2002), Terrin and Triventi (2022), Johnston and Wildy (2016), for a review of the literature. More recently, Francis et al. (2019) showed that grouping is more beneficial for higher-ability students than for lower-ability students. Streaming may also be detrimental to the learning outcomes of specific groups of students, such as minorities and disadvantaged categories (Johnston & Wildy, 2016).

Recent studies in the psychological and sociological literature confirm the influence of ability grouping not only on academic self-concept, but also on students’ social behaviour.

They also show that the effects of tracking are long-term and involve several areas of general self-concept. Mulkey et al. (2005) provide the first study of the long-term effects of ability grouping on academic self-concept, through a comparison of students in tracked schools with similar students in untracked schools. Their results confirm that ability-grouping has persistent instructional benefits for all students. However, high-ability students who are grouped in middle-school suffer considerable losses in terms of self-concept. Dockx et al. (2019) do not find any support for either the "big fish, little pond" hypothesis or the basking in the reflecting glory hypothesis; they suggest that the two theories may not apply to grouping strategies in which the groups have very different characteristics in regard to curriculum and social environments. Palacios et al. (2019) address the need to analyze how ability grouping affects different aspects of self-concept by studying the impact of classroom ability composition on the formation of academic and friendship networks. Their results confirm that ability grouping affects peer relations with the existence of significant differences in the formation and maintenance of academic networks with peers between high- and low-ability classrooms. Francis et al. (2020) support suggestions from the literature that tracking is inequitable with negative impacts on low-achieving pupils that accumulate over time. Kiessling and Norris (2023), by using a longitudinal study of a representative sample of students in the United States, find that individual ability at school and students' relative ranks in their cohorts have an impact on student's mental health and affect long-run outcomes in adulthood, such as employment status, income, as well as, paying bills on time, ever being married, and ever being arrested.

In line with the most recent literature, we propose a study that goes beyond the impact of ability grouping on mere academic self-concept or student achievement. Indeed, we ask whether experiencing ability grouping at school has an effect on the individual's civic participation throughout the individual's adult life. To the best of our knowledge, the first attempt to study the impact of school organization on civic participation is due to Favaro et al. (2020). By analyzing the British context, the authors show that children grouped into homogeneous-ability classes in primary school develop lower levels of civic engagement than their peers who attended non-streamed classes. This effect is particularly important for children grouped into average- and high-ability classes.

The theoretical framework we refer to is the psychological literature, which teaches us that civic participation is an expression of one's self-concept. Self-concept is the image we have of ourselves, which affects how we view our personality traits, and how we see our roles in life, how we feel about our interactions with the world, for example whether we feel we contribute to society. Self-concept influences many aspects of one's life and the hobbies or passions that are important to one's sense of identity, such as being a sports enthusiast, belonging to a certain political party or contributing to society.

Although the development of one's self-concept is never finished, one's self-identity is thought to be primarily formed in childhood (Rogers, 1961) and has a lifelong influence on the individual. Our self-concept develops, in part, through our interaction with others; in addition to our family members and close friends, all the people with whom we relate in our lives can contribute to our self-identity. Thus, the hypothesis that we test is that ability grouping, through the basking in reflected glory effect, influences not only the academic self-concept but also the general self-concept of the individual and, consequently, the individual's civic participation choices throughout her or his life. This theoretical construct allows us to imagine that high-ability students who are clustered in high-ability groups enjoy a higher reputation than low-ability students who are clustered in low-ability groups, and can thus develop a better self-concept than their low-ability peers; furthermore, this higher self-concept translates into greater civic participation in adulthood. We, therefore, expect that participating in streaming for high-ability individuals may positively explains their participation in civic activity in adulthood while negatively explaining participation in civic activity for low-ability individuals.

The data we employ to address our research question are derived from the British National Child Development Studies (NCDS), which is a cohort study concerning UK citizens born during the week of 3–9 March 1958[1] that gathered information on them at different times in their lives from 1965 to 2009.[2] The NCDS is an excellent source of data for this kind of analysis, as it provides information on the type of primary school attended by cohort members (streamed versus non-streamed) and, in cases of streaming, on the ability class (high, medium, or low) to which the individual was assigned. It also provides information on pretreatment covariates collected at the first sweep. Furthermore, the NCDS provides information on the civic engagement of cohort members during adulthood. In this respect, we use information from the sweeps of NCDS when the individuals were 33, 42, and 51 years old. The available measures of civic engagement consider cohort members' participation in (1) political parties, (2) charities, (3) environmental, and voluntary associations, (4) women's organizations, (5) school and residents' associations, and (6) voting in the last general election (Putnam, 2000).

Regarding the methodological approach, in this article, we propose a novel causal hidden (or latent) Markov model (HM) that is tailored to the analysis of binary multivariate longitudinal data and facilitates causal inference with complex data structures that might otherwise be intractable. We extend the proposal in Bartolucci et al. (2016) and in Pennoni et al. (2023), based on potential versions of latent variables that permit a clear definition of the average treatment effects (ATEs), including the effects of posttreatment time-varying covariates. We consider a generalized propensity score (PS) approach (Rosenbaum & Rubin, 2023) to estimate the individual probability of receiving a certain treatment according to pretreatment covariates. Estimation of the HM model is performed by a weighted log-likelihood through the expectation-maximization algorithm (Baum et al., 1970; Dempster et al., 1977) where weights are based on the inverse of the PSs. In this way, we mitigate the confounding due to the dependence of the treatment and latent variables (and subsequently, the responses) on common observable variables in estimating the ATEs. We also account for missing responses, such that individuals with partially or completely missing responses on one or more indicators of participation are included in the analysis on the basis of a missing-at-random (MAR) assumption (Rubin, 1976; Little & Rubin, 2020). In contrast, missing values occurring for the posttreatment covariates are handled by a dummy variable serving as a missing indicator (Dardanoni et al., 2011). A relevant aspect of the proposed approach is that it can be easily applied since the required computational tools are available in package `LMest` (Bartolucci et al., 2017) of the R software (R Core, 2023). An example of the code implemented to estimate the proposed models is available at the Github page https://github.com/penful/Causal_HM_streaming.

The remainder of this paper is organized as follows. Section 2 describes the application that motivates the proposed approach. Section 3 outlines the estimation of the PS weights

[1] The use of cohort studies has pros and cons. The main advantages lie in the possibility of following subjects over a long period and the availability of data for several dimensions of human life; this allows us to analyze a variety of facts and assess how outcomes evolve in the long-term. The main disadvantage is the difficulty of following up with all cohort members along their life-course.

[2] The NCDS database has been used to analyze civic engagement and social participation during adulthood in other studies, such as those of Bowling et al. (2016) and Parry et al. (2021).

and illustrates the causal HM model. Section 4 reports and discusses the empirical results, particularly regarding ATEs. Section 5 provides some concluding remarks. The Appendix reports additional information on the data and results of the application.

2 Longitudinal Cohort Study

The NCDS is a cohort study following the lives of 18,558 individuals born in England, Scotland, and Wales during the week of 3–9 March 1958. It started as the Perinatal Mortality Survey (PMS) and followed as the NCDS in subsequent years: 1965, 1969, 1974, 1981, 1991, 1999–2000, 2004–2005, and 2008–2009.[3] The NCDS gathers information on cohort members' health, education, behavior, parental background, economic conditions, and social and labor market outcomes. These are derived from a variety of sources, including self-reports, participants' parents, medical examinations, and ability and behavioral tests. Figure 1 shown in Appendix 1 illustrates the temporal data structure of the cohort study. The data are particularly well-suited to our goals as they allow measurement of civic engagement during adulthood and include information on the type of school/class attended during childhood; these data also provide substantial information to control for heterogeneity among the cohort members.

To recover all needed information to evaluate the causal effects of streaming, we use six sweeps of the NCDS database. In particular, the 1969 sweep, conducted when cohort members were 11 years old, includes retrospective information that allows us to identify whether an individual attended a primary school that applied streaming and to distinguish between cohort members who participated at a streamed school according to ability class of high, average, or low levels. We estimate the ATEs of streaming on the degrees of civic engagement participation when the individuals were 33 years old and the ATEs on the probability of moving between different levels of engagement when the individuals were 42 and 51 years old.

It is worth noting that, as shown in Fig. 1 reported in Appendix 1, the number of observations in the NCDS has declined over time; for example, it decreased from approximately 15,300 in the second sweep (1969) to approximately 9,800 in the eighth sweep (2008–2009). Indeed, the NCDS presents some attrition from wave to wave, and this may constitute an issue if the representativeness of the sampled individuals is reduced.

2.1 Description of Multiple Treatment and Responses

As described in the introduction, streaming is the practice of grouping students into classes with other students with comparable skills or needs (Plowden, 1967), see Chapter 20.3 for more details); the students were enrolled in low-ability, average-ability, or high-ability classes, or they attended a primary school not applying streaming and were thus enrolled in mixed-ability classes. Cohort members who attended schools not applying streaming correspond to approximately two-thirds of the sample, and the remaining cohort members attended schools applying streaming. About 40% were enrolled in high-ability classes, while other streamed students were equally distributed between

[3] More recently, the ninth wave has been published; for more details, see the website at https://cls.ucl.ac.uk/cls-studies/1958-national-child-development-study/.

Table 1 Number and percentage of individuals according to the available information on the treatments (data collected in the second sweep in 1969)

Treatment	Number	%
Low-ability	1365	7.355
Average-ability	1335	7.194
High-ability	1800	9.699
No stream	9414	50.728
Missing	4644	25.024
Total	18,558	100.000

low-ability and average-ability classes. From the 1969 data (second sweep), we draw information to identify treated and untreated individuals who have attended a primary school applying streaming. The missing observations are not negligible when focusing on the streaming variable, representing 25% of the original cohort-member observations. Our study is based on data from 13,914 individuals who had no missing treatment indicator out of the full sample of 18,558 individuals of the observational cohort study.

Table 1 reports the proportions of treated individuals across groups in the whole sample along with those for whom the observation on the treatment was not available. The group of untreated individuals corresponds to around 51%.

From the fifth, sixth, and eighth sweeps (when the individuals were 33, 42, and 51 years old, respectively), we draw information on different aspects of civic engagement during adulthood. In particular, the outcome consists of a set of multivariate binary variables allowing us to identify whether the surveyed individuals were members of organizations such as political parties, environmental, charity, and voluntary groups, women's groups and institutes, parent–school organizations, and tenants'/residents' associations, as well as whether they voted in the last general election. These forms of engagement are mostly traditional forms of civic participation. Voting is the most straightforward act of citizenship, and political party membership is a measure of direct engagement in politics. The participation in voluntary and charity groups, school-service groups such as parent–school organizations, women's groups/institutes, and tenants'/residents' associations also express traditional forms of civic participation (Putnam, 2000).

Preliminary descriptive statistics on the chosen indicators shown in Table 2 reveal that the most common form of civic participation is represented by voting; in the analyzed years, a percentage varying from 43 to 53% had participated in the last observed general election. Between 5 and 7% of surveyed individuals were members of environmental, charity, and voluntary organizations, while 2–4% participated in parent–school organizations. However, we note that both voting and the membership of organizations, aside from tenants'/residents' associations, declined over the analyzed period. Concerning attrition and selection bias problems in the NCDS data, Dearden et al. (1997) showed that attrition in the NCDS tended to weed out individuals with lower ability and educational qualifications. More recently, Hawkes and Plewis (2006) found that attrition and non-response issues can be associated with only a few significant predictors, supporting the view that the data are still reasonably representative of the reference population. We analyze the data under the MAR assumption for partially or entirely missing values of the response variables.

Table 2 Percentage frequencies of responses for each indicator of civic engagement; percentages of missing values provided in parentheses

Responses	Age 33		Age 42		Age 51	
	No	Yes	No	Yes	No	Yes
Voted in last gen. election	15.775	53.241	15.682	53.392	15.833	43.223
(Missing)	(30.984)		(30.926)		(40.944)	
Political parties	67.529	1.445	67.838	1.380	58.933	0.963
(Missing)	(31.026)		(30.782)		(40.103)	
Volun./environ./charity org.	62.160	6.813	62.16	7.058	55.081	4.815
(Missing)	(31.026)		(30.782)		(40.103)	
Women's org.	67.587	1.387	68.341	0.877	59.372	0.525
(Missing)	(31.026)		(30.782)		(40.103)	
Parent–school org.	64.151	4.822	64.381	4.837	57.633	2.264
(Missing)	(31.026)		(30.782)		(40.103)	
Resident associations	67.191	1.782	67.479	1.739	57.151	2.745
(Missing)	(31.026)		(30.782)		(40.103)	

2.2 Description of Pre and Posttreatment Covariates

From the original 1958 and 1965 sweeps (see Fig. 1 in Appendix 1) we draw information on pretreatment covariates. In particular, we account for gender and region of birth of the surveyed individuals using data from the 1958 sweep. Additionally, using information from the 1965 sweep, we account for the region of living and control for math and reading tests, which were introduced to approximate the role of cognitive skills, and the results of the British Social Adjustment Guide (BSAG, scored at age 7), which is a standardized psychometric test of social maladjustment introduced to measure noncognitive skills. The literature has also stressed that noncognitive skills, which are defined as personality resources linked to motivation in learning, relational capabilities, emotional stability, and autonomy in pursuing personal objectives, may have later effects on civic engagement and political participation; see, among others, Holbein (2017). For example, noncognitive skills may increase an individual's general motivation level, which may, in turn, increase their capacity to participate in politics and empower individuals to follow through once they desire to participate. In addition, the PMS and the 1965 sweep provide some pretreatment parental background covariates. Table 11 shown in Appendix 1 lists the pretreatment covariates that we use in the analysis as potentially influencing the treatment assignment. Since these factors were collected at the beginning of the cohort study, we argue that the effects of unobserved factors are negligible in estimating the treatment-selection probabilities compared to the measured variables. In particular, the BSAG score is measured at age 7 through a standardized psychometric test of social maladjustment that helps to diagnose the extent and nature of this feature among children at school: the higher the score's value, the greater the evidence of behavioral problems. This variable approximates the underlying aspects of the noncognitive skills of the surveyed individuals. The mathematics test score, measured at age 7, along with a reading comprehension score measured at the same age, are introduced as proxies of cohort members' cognitive skills.

We also consider a few variables to account for the role of family background. The social class of parents at the birth of the cohort member (mother and father's social class at birth) is an ordinal variable defined according to five groups, where the first group indicates the highest social class (top), and the fifth group is the lowest social class (bottom). When the cohort member was aged 7 (father working class at age 7), the social class of the father describes the father's socio-economic class according to a working position variable with five categories, that is, professional, managerial/technical, skilled, partly skilled, and unskilled. A dummy variable accounts for accommodation tenure when the cohort member was aged 7 (homeowner). Parental education is considered through two dummy variables indicating whether the parents of the cohort member stayed at school beyond the minimum compulsory education (mother/father with an educational level higher than compulsory education). The information on each cohort member's mother was collected at the surveyed individual's birth, whereas the information on the cohort member's father was collected when the cohort member was aged 7. Among the pretreatment covariates, we also include a set of dummy variables for the time-varying region of living recorded at birth, age 7, and age 11. Their inclusion may be important to account for some heterogeneities, such as regional differences both at the institutional level and in the educational system.

Posttreatment variables were collected during adulthood when the cohort members' ages ranged between 33 and 51 years old. We consider the following control covariates in our analysis: dummy variables for gender and marital status and a categorical variable indicating the work status of the individual. This variable identifies four possible statuses: inactivity, unemployment, part-time work, and full-time work. In the empirical analysis, inactivity and unemployment are collapsed into a single category, indicating a nonemployment condition. Each cohort member's educational level is considered through six dummy variables for the National Vocational Qualifications (NVQs), which are work-based awards in England, Wales, and Northern Ireland. We consider a set of dummy variables indicating the region of living during adulthood. Table 12 shown in Appendix 1 reports the descriptive statistics of the posttreatment covariates and provides additional details.

3 Causal Hidden Markov Model with Posttreatment Covariates

We cast our methodological proposal in the potential outcomes (POs) framework developed in the field of observational studies by Rubin (1974) and later extended by Holland (1986) and Rubin (2005). When a marginal effect of a treatment is of interest, Rosenbaum (1987) proposed the use of the PS method to estimate the effect of treatment versus control in non-randomized studies; see also Imbens (2000), Robins et al. (2000), and Robins (2003). This involves estimating a treatment-selection model to dispose of suitable individual weights corresponding to the probability of receiving treatment given a set of pretreatment covariates. As also recently discussed by Rosenbaum (2020), PS-based methods are widely applied to mitigate bias in treatment-effect estimation incurred by self-selection on observables; see also Stuart (2010).

In this paper, we propose a causal hidden Markov (HM) model for binary multivariate longitudinal data that addresses the problem of estimation of causal effects on a latent variable representing an individual characteristic, namely, civic participation in our application, which is not directly observed but instead measured through several indicators. The main difference with respect to the standard POs approach, which we propose to apply in the current study, is that we define potential versions of latent variables rather than of observable variables; this is in agreement with the proposal found in Bartolucci et al.

(2016). A novelty of the current proposal is that we consider the effect of posttreatment covariates on the initial and transition probabilities of the causal HM model along with the treatment effect, as specified in the next section. Moreover, we account for missing responses, such that individuals with partially or entirely missing responses on one or more indicators are also included in the analysis based on the MAR assumption. We also account for missing values of the posttreatment covariates.

The inferential approach, as detailed in the following section, is developed in two phases. First, we construct a PS model to estimate an individual's probability of receiving the treatment according to the pretreatment covariates. Then, individual weights based on the inverse of the PSs are used for maximum likelihood estimations of the model parameters in order to identify the ATEs on both initial and transition probabilities of the Markov chain. In this way, treated and nontreated individuals are balanced on their confounders at the baseline. The model-building strategy allows us to find a suitable number of latent subpopulations, corresponding to latent states by first estimating the causal HM model only with the PS weights. Once model selection is performed, as illustrated in the following, we estimate the causal HM model with the selected number of latent states with treatment and posttreatment covariates. The ATEs are defined as specified in the following, and they are identified under the usual assumptions required in the causal inference literature. Most importantly, within the proposed framework, these effects are estimated both at the beginning of the period and at each transition between time occasions. Additionally, in comparison with the standard approaches employed for causal inference, we can (1) specify if and how it is possible to characterize distinct groups of individuals in the population according to the observed indicators, (2) identify the most suitable number of these groups, (3) estimate a different ATE for each of the identified subpopulations both at the first time occasion and at subsequent time points, (4) evaluate how individuals move between groups according to the treatment and posttreatment covariates, and (5) cluster individuals dynamically over time, as each individual is assigned to one group at every time occasion with the possibility to predict each single pattern.

3.1 Model Assumptions

Let $\boldsymbol{Y}_{it}$ be the column vector of r response variables defined for every individual i and time point t, with $i = 1, \dots, n$ and $t = 1, \dots, T$. In our application, these variables correspond to voting behavior and individual's participation in the previously noted organizations. We thus refer to six binary response variables observed at three time points. We also admit that some responses are missing at one or more time occasions by relying on the MAR assumption, which is rather standard in the longitudinal literature (Little & Rubin, 2020). This amounts to assuming that the event of a missing response is conditionally independent of the unobserved response given the observed data. That is, the missing pattern's probability, given the observed data, does not depend on the unobserved data (Lu & Copas, 2004; Bojinov et al., 2020). Missing values occurring for the posttreatment covariates are handled by a dummy variable serving as a missing indicator, which is included in the model according to the proposal of Dardanoni et al. (2011).

Let Z_i be a categorical variable indicating the treatment for each individual $i, i = 1, \dots, n$, with levels ranging from 0 to $l - 1$, and let $\boldsymbol{V}_i$ be a column vector of pretreatment covariates, occurring temporally before the treatment decision. In the empirical illustration, these correspond to the socio-demographic characteristics of individuals and include their school performance observed prior to treatment.

The following assumptions are formulated within the proposed approach:

(i) stable unit treatment value assumption (SUTVA), according to which the individual treatments are completely represented without relevant interactions between individuals;
(ii) exogeneity (EXOGEN), requiring that the treatment does not influence the control covariates denoted as $\boldsymbol{V}_i$;
(iii) positivity (POS), requiring that each individual has a non-zero probability of being treated with each type of treatment including control; that is,

$$0 < P(Z_i = z|\boldsymbol{V}_i = \boldsymbol{v}_i) < 1, \quad i = 1, \dots, n, \ z = 0, \dots, l-1,$$

given any configuration $\boldsymbol{v}_i$ of the pretreatment covariates.

3.2 Potential Latent Variables

Potential latent variables are denoted by $H_{it}^{(z)}$, with $i = 1, \dots, n$ and $t = 1, \dots, T$, and these have a discrete distribution with support points $\{1, \dots, k\}$, where k is usually small. These states are related to different clusters of individuals with similar behavior within the clusters. For instance, $H_{i2}^{(1)}$ denotes the potential civic engagement of individual i at the second time point when she/he is exposed to treatment level 1. The individual sequences of potential latent variables $\boldsymbol{H}_i^{(z)} = (H_{i1}^{(z)}, \dots, H_{iT}^{(z)})'$ generate a sequence of latent variables H_{it} collected in the vectors $\boldsymbol{H}_i = (H_{i1}, \dots, H_{iT})'$, $i = 1, \dots, n$, according to the consistency rule: $\boldsymbol{H}_i = \boldsymbol{H}_i^{(z_i)}$ with probability 1, where z_i is the observed treatment of individual i. We also assume that pretreatment covariates are sufficiently informative, so that $Z_i \perp\!\!\!\perp H_{it}^{(z)}|\boldsymbol{V}_i$, for $i = 1, \dots, n, t = 1, \dots, T$, and $z = 0, \dots, l-1$.

The multivariate longitudinal data structure is accounted for by assuming that, for $z = 0, \dots, l-1$, the sequence of latent potential outcomes $\boldsymbol{H}_i^{(z)}$ follows a Markov chain of first-order. Clearly, these latent variables are related to the observable responses. In particular, following a common assumption in the discrete latent variable literature (Bartolucci et al., 2022), the response variables in every vector $\boldsymbol{Y}_{it}$ are assumed to be conditionally independent given $H_{it}^{(z)}$. This conditional distribution is characterized by the success probability parameters

$$\phi_{j|h} = p(Y_{ijt} = 1|H_{it}^{(z)} = h), \quad h = 1, \dots, k, \ j = 1, \dots, r, \tag{1}$$

for every i and t.

Let $\boldsymbol{X}_{it}$ be a column vector of the time-varying posttreatment covariates for $i = 1, \dots, n$ and $t = 1, \dots, T$, with possibly missing values that are dealt with by dummy variables, as already mentioned above. To mitigate selection bias we account for possible factors influencing the underlying propensity towards civic engagement (Leite, 2016; Cela, 2017) considering the distribution of the latent potential outcomes. The initial and transition probabilities are defined as

$$\pi_{h|\boldsymbol{x}}^{(z)} = p(H_{i1}^{(z)} = h|\boldsymbol{X}_{i1} = \boldsymbol{x}), \quad h = 1, \dots k,$$
$$\pi_{h|\bar{h}t\boldsymbol{x}}^{(z)} = p(H_{it}^{(z)} = h|H_{i,t-1}^{(z)} = \bar{h}, \boldsymbol{X}_{it} = \boldsymbol{x}), \quad t = 2, \dots, T, \ \bar{h}, h = 1, \dots, k,$$

and these are collected in the initial probability vector $\boldsymbol{\pi}_{\boldsymbol{x}}^{(z)}$ and in $k \times k$ stochastic transition matrices $\boldsymbol{\Pi}_{t\boldsymbol{x}}^{(z)}$, the elements of which are not constant over time. A multinomial logit parameterization is employed to account for the treatment and covariate effects on the initial probabilities:

$$\log \frac{\pi^{(z)}_{h|\boldsymbol{x}}}{\pi^{(z)}_{1|\boldsymbol{x}}} = \alpha_h + \boldsymbol{d}(z)'\boldsymbol{\beta}_{1h} + \boldsymbol{x}'\boldsymbol{\beta}_{2h}, \quad h = 2, \dots, k, \tag{2}$$

where α_h is the intercept specific to each latent state, $\boldsymbol{\beta}_{1h} = (\beta_{1h2}, \dots, \beta_{1hl})'$ is a column vector of $l-1$ regression parameters referring to the treatment levels, $\boldsymbol{d}(z)$ is a column vector of $l-1$ zeros with the $(z-1)$-th element equal to 1 if $z > 0$, and $\boldsymbol{\beta}_{2h} = (\beta_{2h1}, \dots, \beta_{2hc})'$ is a column vector of regression parameters referring to the posttreatment covariates, with c denoting the number of these covariates. As a reference category, the multinomial logits have the first category corresponding to latent state 1. Since each element β_{1hz} of $\boldsymbol{\beta}_{1h}$ for $z > 1$ is a shifting parameter from the first to the h-th logit, each of these parameters can be interpreted as the ATE. In particular, this parameter summarizes the latent outcome once individuals receive treatment z with respect to the mean outcome once the same individuals receive treatment z'.

For the transition probabilities, we use a parsimonious parameterization which is defined as:

$$\log \frac{p(H^{(z)}_{it} = h | H^{(z)}_{i,t-1} = \bar{h}, \boldsymbol{X}_{it} = \boldsymbol{x})}{p(H^{(z)}_{it} = \bar{h} | H^{(z)}_{i,t-1} = \bar{h}, \boldsymbol{X}_{it} = \boldsymbol{x})} = \gamma_{0\bar{h}h} + \boldsymbol{d}(z)'(\boldsymbol{\gamma}_{1h} - \boldsymbol{\gamma}_{1\bar{h}}) + \boldsymbol{x}'(\boldsymbol{\delta}_{2h} - \boldsymbol{\delta}_{2\bar{h}}), \tag{3}$$

where $\bar{h} = 1, \dots, k$, $h = 2, \dots, k$, and $t = 2, \dots, T$; moreover, $\gamma_{\bar{h}h}$ is the intercept, $\boldsymbol{\gamma}_{1h} = (\gamma_{1h2}, \dots, \gamma_{1hl})'$ are vectors of regression coefficients referring to the treatment levels corresponding to the ATEs, and $\boldsymbol{\delta}_{2h} = (\delta_{2h1}, \dots, \delta_{2hc})'$ are vectors of regression coefficients corresponding to the posttreatment covariates. In this way, there are fewer parameters to be estimated with respect to adopting multinomial logit parameterization; see Bartolucci et al. (2014).

3.3 Weighted Maximum-Likelihood Estimation

For every individual i, where $i = 1, \dots, n$, let $\boldsymbol{v}_i$ be the vector of observed pretreatment covariates and z_i be the realized treatment, and let $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{it}$ be vectors of time-specific posttreatment covariates and responses, respectively, for $t = 1, \dots, T$. We control for differences between the treatment levels by estimating the probability of being assigned to a particular ability class for each individual using the following multinomial logit model:

$$\log \frac{p(Z_i = z | \boldsymbol{v}_i)}{p(Z_i = 0 | \boldsymbol{v}_i)} = \eta_z + \boldsymbol{v}_i'\boldsymbol{\lambda}_z, \quad z = 1, \dots, l-1, \tag{4}$$

where η_z is the intercept specific of each treatment level and $\boldsymbol{\lambda}_z$ is the vector of regression parameters. On the basis of the parameter estimates, we compute the inverse of the PS as follows:

$$\hat{w}_i = n \frac{1/\hat{p}(z_i | \boldsymbol{v}_i)}{\sum_{m=1}^{n} 1/\hat{p}(z_m | \boldsymbol{v}_i)}, \quad i = 1, \dots, n.$$

It is worth mentioning that the PS weights are also estimated for individuals with missing responses at one or more time points. They are rescaled so that their sum is equal to the sample size, that is, $\sum_{i=1}^{n} \hat{w}_i = n$, and as in the proposal of Robins et al. (2000), they can be set up to a maximum value (weights greater than a certain value are set to this value) by employing a trimming method if their variability is too large. In this way, we avoid instability problems in estimating the ATEs (Robins & Rotnitzky, 1995; Stuart, 2010).

Maximum likelihood estimation is based on the weighted log-likelihood

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{w}_i \ell_i(\boldsymbol{\theta}),$$

$$\ell_i(\boldsymbol{\theta}) = \log p(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT} | \boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, z_i),$$

where $\boldsymbol{\theta}$ is the vector of all model parameters arranged in a suitable order, and $p(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT} | \boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, z_i)$ corresponds to the manifest probability of the responses provided by individual i given the treatment and posttreatment covariates. This probability is computed by suitable recursions developed in the HM model literature (Baum et al., 1970; Welch, 2003) according to the conditional probabilities $p(\boldsymbol{y}_{it} | H_{it}^{(z)} = h, \boldsymbol{x}_{it})$ depending on the response probabilities in Eq. (1) and the initial and transition probabilities of the hidden Markov chain parametrized as defined in Eqs. (2) and (3).

The above log-likelihood is maximized with respect to $\boldsymbol{\theta}$ by using the expectation-maximization (EM) algorithm (Baum et al., 1970; Dempster et al., 1977). This algorithm alternates two steps until convergence:

- E-step, which computes the posterior expected value of the complete data weighted log-likelihood given the observed data and the current value of the parameters;
- M-step, which maximizes the posterior expected value with respect to the model parameters.

The complete data log-likelihood is expressed as:

$$\ell^*(\boldsymbol{\theta}) = \sum_{h=1}^{k} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{w}_i a_{hit} \sum_{j=1}^{r} [y_{ijt} \log \phi_{j|h} + (1 - y_{ijt}) \log(1 - \phi_{j|h})]$$
$$+ \sum_{h=1}^{k} \sum_{i=1}^{n} \hat{w}_i a_{hi1} \log \pi_{\boldsymbol{x}_{i1}}^{(z_i)} + \sum_{\bar{h}=1}^{k} \sum_{h=1}^{k} \sum_{i=1}^{n} \sum_{t=2}^{T} \hat{w}_i b_{\bar{h}hit} \log \pi_{\bar{h}|ht\boldsymbol{x}_{it}}^{(z_i)},$$

where a_{hit} is an indicator variable equal to 1 if individual i belongs to latent state h at time point t, and $b_{\bar{h}hit} = a_{\bar{h}i,t-1} a_{hit}$ is an indicator variable equal to 1 if the same individual moves from state $\bar{h}$ to state h at occasion t.

We refer the reader to Bartolucci et al. (2013), Chapter 5, for details about the implementation of the EM algorithm for the HM model. This algorithm requires a proper initialization of the model parameters to account for local maxima in the likelihood function. Several strategies have been proposed to this end; see, among others, Maruotti and Punzo (2021).

Standard errors for the parameter estimates are obtained by exact computation or approximations of the observed information matrix (Bartolucci et al., 2013). We suggest selection of the number of latent states according to the Bayesian information criterion (BIC; Schwarz, 1978). This is based on the index

$$\mathrm{BIC}_k = -2\hat{\ell}_k + \log(n)\#\mathrm{par}_k,$$

where $\hat{\ell}_k$ denotes the maximum of the log-likelihood and $\#\mathrm{par}_k$ denotes the number of free parameters of the model with k latent states.

Another important aspect is related to the fact that the posterior probability of H_{it} can be estimated as a byproduct of the EM algorithm. This is defined as

$$g_{h|it} = p(H_{it} = h|\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, \boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT}, z_i),$$

$h = 1, \dots, k, i = 1, \dots, n, t = 1, \dots, T$, given the posttreatment covariates, response configurations, and the received treatment. Considering these probabilities and using the algorithm of Viterbi (1967) and Juang and Rabiner (1991), we perform global decoding to predict the most *a posteriori* likely sequence of states underlying the individual observed data.

4 Results of the Application

In the following we show the results of the proposed approach. First, we report the estimated PS weights; second, we comment the main results; third, we report additional estimates and prediction of the latent states. Finally, we comment on a robustness check and a sensitivity analysis excluding posttreatment covariates.

4.1 Propensity Score Weights

The individual PSs are estimated on the basis of a multinomial logit model as in Eq. (4) with the pretreatment covariates illustrated in Table 11 shown in Appendix 1 including binary indicators for missing values of the covariates. As previously illustrated, the PS method balances potential confounders between treatment groups in terms of pretreatment covariates. Weights reduce the differences between, for example, individuals allocated to low-ability classes showing a lower average score in reading with respect to other individuals.

Figure 2 shown in Appendix 2 depicts Box plots that display a substantial overlap of the distribution of the PS estimates between the treatment groups. Table 13 shown in Appendix 2 reports the descriptive statistics of baseline covariates by treatment considering the estimated PS weights that have been trimmed to avoid possible instability. Examining the distribution of the covariates, we can observe that the average values across groups of streaming are more similar when compared to those obtained without weights (see Table 11 in Appendix 1).

4.2 Model Selection and Main Parameter Estimates

Model selection concerns, in particular, the choice of the number of latent states; to this end, we first estimate the model with PS weights but without treatment and posttreatment covariates for increasing values of states. The model minimizing the BIC, presented in Sect. 3.3, is selected as the one corresponding to the best compromise between the goodness of fit and parsimony. The results are reported in Table 3, according to which we select $k = 3$ latent states. Note that, in this case, we estimate the model up to four latent states since the BIC index increases with this number of states.

Table 3 Maximum log-likelihood, number of parameters, and BIC index for a number of latent states of the HM model ranging from 1 to 4

k	$\hat{\ell}_k$	$\#\mathrm{par}_k$	BIC_k
1	− 53,520.965	6	106,563.065
2	− 40,323.869	15	80,790.572
3	− 39,424.235	26	79,096.050
4	− 39,403.001	39	79,178.094

Table 4 Estimated conditional probabilities of civic engagement

Responses	Latent state (h)		
	No engagement (1)	Only voting (2)	Participating (3)
Voted in last general election	0.178	0.956	0.918
Political parties	0.001	0.007	0.108
Volun./environ./charity org.	0.025	0.038	0.407
Women's organizations	0.002	0.006	0.068
Parent–school organizations	0.012	0.036	0.222
Resident associations	0.013	0.020	0.105

Table 5 ATEs of streaming at age 33 on civic engagement measured for the second and third latent states with respect to the first latent state (with significance indicated at the †10%, *5%, and **1% level)

ATE estimate	Logit 2	Logit 3
Low-ability	− 0.126	− 0.045
Average-ability	− 0.035	− 0.123
High-ability	0.260**	0.169

The model with $k = 3$ states is then estimated with treatment and posttreatment covariates affecting the latent process as illustrated in Sect. 3, where initial and transition probabilities are parameterized as in Eqs. (2) and (3). At convergence of the EM algorithm, this model has a maximum log-likelihood of −33,682 with 126 free parameters, and the BIC index is equal to 68,565. Thus, we find a substantial decrease in the BIC index with respect to the other models considered in Table 3.

Table 4 lists the estimated conditional success probabilities of the available indicators. Each of the three discovered subpopulations identifies individuals differing in terms of civic engagement: individuals with "no engagement" are in the 1st latent state, individuals participating in voting are in the 2nd, and individuals participating in voting and organizations are in the 3rd. The subpopulation of individuals in the 2nd is labeled as "only voting," as these individual are not involved in any other kind of active civic participation. The 3rd represents the most active subpopulation: they vote and are also involved in some organizations such as voluntary or parent–school organizations, or residents' associations. This group of people is labeled as "participating". In 1991, according to the estimated initial probabilities, the percentages in the population of each group (size of each subpopulation at the beginning of the period) were 26.39%, 58.75%, and 14.36%, respectively.

Table 6 Estimated averages of the initial probabilities (at age 33) according to the treatment level

Treatment	Latent state (h)		
	No engage. (1)	Only voting (2)	Participating (3)
Low-ability	0.329	0.572	0.100
Average-ability	0.280	0.597	0.123
High-ability	0.197	0.607	0.196
No stream	0.265	0.585	0.150

Table 7 ATEs of streaming at ages 42 and 51 on civic engagement measured on the transition probabilities based on the differences between logits (with significance indicated at the †10%, *5%, and **1% level)

ATE estimate	Diff. logit 2	Diff. logit 3
Low-ability	− 0.089	− 0.508
Average-ability	0.091	− 0.779
High-ability	0.028	− 0.560

In the following, we first report the estimated ATEs for streaming in terms of initial and transition probabilities, and then we show other estimates concerning the parameters referring to the posttreatment covariates. In particular, Table 5 concerns the ATEs for the initial probabilities. We notice that only the parameter referring to high-ability classes is significant with respect to the first logit: the estimated odds ratio for the 2nd ("only voting") with respect to the first ("no engagement") is exp(0.260) = 1.30 times the estimated odds for no streaming. Keeping the values of the posttreatment covariates fixed, individuals assigned to high-ability classes tend to participate more in voting than individuals who were not streamed. To make the interpretation of these estimates easier, Table 6 displays the corresponding estimated averages of the initial probabilities according to treatment levels, showing the direction of the effect of the treatment.

Table 7 lists the estimated ATEs on the transition probabilities, revealing that none is significant, while Table 8 displays the estimated average transition matrix for the years 1991–2000 (upper panel) and 2000–2009 (lower panel), where the individuals were 33–51 years old, respectively.

Table 8 Estimated averaged transition probabilities from the first to the second occasion, from age 33 to age 42 (upper section), and from the second to the third occasion, from age 42 to age 51 (lower section)

$\bar{h}$	Latent state (h)		
	No engage. (1)	Only vot. (2)	Particip. (3)
From the first to the second occasion			
1	0.790	0.210	0.000
2	0.203	0.796	0.001
3	0.014	0.110	0.876
From the second to the third occasion			
1	0.815	0.185	0.000
2	0.232	0.767	0.001
3	0.014	0.103	0.883

Table 9 Estimates of the logit regression parameters of the initial probability of belonging to the second or third latent state with respect to the 1st (with significance indicated at the †10%, *5%, and **1% level)

Covariate	Logit 2	Logit 3
Intercept	0.346*	− 1.552**
Married	0.237**	− 0.019
Missing	− 1.111**	− 0.319
Working full-time	− 0.071	− 1.215**
Working part-time	− 0.037	− 0.291
Missing	0.734	− 0.877
Poor health	− 0.388	− 0.684
Missing	− 0.441	1.525**
NVQ1	0.181	0.576
NVQ2	0.589**	1.532**
NVQ3	0.867**	2.643**
NVQ4	0.562**	3.095**
NVQ5/6	0.489**	3.924**
Missing	− 0.009	− 0.420
North	0.060	− 0.472
Yorkshire and Humberside	0.884	0.271
East Midlands	0.011	0.001
East Anglia	0.008	− 0.223
South West	− 0.350**	− 0.262
West Midlands	0.305*	− 0.189
North West	0.087	− 0.468*
Wales	0.704**	0.754**
Scotland	0.590**	0.125

The following categories are assumed as a baseline: not married, unemployed, no qualifications, South East

On the basis of the results in these tables, the probabilities of switching from one state to another over the two time occasions tend to be very similar. For both periods we estimate a general tendency to move from the 3rd ("participating") toward the 2nd ("only voting"), showing that 11% and 10% of individuals withdrew from the associations they were involved in the first and second periods, respectively. Nonetheless, persistence in original states was predominant, thus suggesting that propensity towards civic engagement remains basically crystalized during adulthood. This is consistent with evidence indicating that political and civic identities that are shaped during adolescence and the young-adult years are highly predictive of the views individuals will hold in middle age and late adulthood (Flanagan & Levine, 2010).

4.3 Other Parameter Estimates and Predicted Latent States

Table 9 shows the estimated regression coefficients of the posttreatment covariates in the model for the initial probabilities.

We note that at age 33 there is a general tendency towards the 1st, that of "no engagement," with respect to the third, that of "participating," as suggested by the estimated intercept. Moreover, having a certain educational level increases the probability of civic engagement since the estimated coefficients are always positive and significant

Table 10 Estimates of the logit regression parameters affecting the transition probabilities based on the differences between logits (with significance indicated at the †10%, *5%, and **1% level)

Covariate	Diff. logit 2	Diff. logit 3
Married	0.405**	0.634
Missing	– 1.905**	– 1.782
Working full-time	– 0.144	0.341
Working part-time	– 0.403**	– 0.456
Missing	– 0.306	5.064
Poor health	– 0.855**	2.016
Missing	– 0.622	– 1.457
NVQ1	0.044	– 10.701
NVQ2	0.255**	– 8.985
NVQ3	0.473**	– 8.603
NVQ4	0.870**	– 8.865
NVQ5/6	1.468**	18.216
Missing	0.262	– 3.224
North	– 0.374**	– 1.602**
Yorkshire and Humberside	1.334	2.136
East Midlands	– 0.189	– 1.077
East Anglia	– 0.094	– 0.088
South West	0.106	– 1.163**
West Midlands	– 0.425**	4.293
North West	– 0.678**	– 0.391
Wales	0.154	– 1.288*
Scotland	0.134	1.310

The following categories are assumed as baseline: not married, unemployed, no qualifications, South East

for both logits with respect to the base-category represented by individuals with no qualifications. For example, for the highest educational level (NVQ5/6), the estimated odds ratio for the 3rd ("participating") with respect to the first ("no engagement") is $\exp(3.924) = 50.602$.

Other characteristics appear to be significant in shaping civic engagement. In particular, people who work full time are all less likely to participate in civic activities than those who are unemployed or inactive, probably due to time constraints. Higher levels of civic engagement at age 33 are observed in the West Midlands, Wales, and Scotland, where there is a general tendency towards the 2nd ("only voting") with respect to the 1st ("no engagement") when compared to the base category. Interestingly, individuals living in Wales have a certain propensity to participate, while those living in the South West manifest a low tendency to vote (– 0.350) with respect to people living in the South East.

Table 10 lists the estimated regression parameters of the logit models for the transition probabilities, while the estimates of the intercepts are reported in Table 14 in Appendix 2. For individuals having a certain level of education with respect to individuals with no qualifications, there is a general tendency to transit to the 2nd ("only voting") with respect to staying in the 1st ("no engagement"). People living in the South West tend to move to the 1st with respect to those living in the South East.

4.3.1 Robustness Check and Sensitivity Analysis with Respect to Posttreatment Covariates

The results presented above are robust to the inclusion of the variable controlling for behavioral problems at age 33. We additionally estimate the proposed model, including this posttreatment covariate along with the previous one, through an index that counts the number of problems (depression, anxious/jittery, anxious/scared, overconfident/overexcited, and compelled to repeat actions). We note that also controlling for behavioral problems at age 33 does not change the estimated ATEs; furthermore, its effects on the initial and transition probabilities, when statistically significant, show the expected sign. Having no behavioral problems increases participation on the first occasion, while having more than two behavioral problems reduces the probability of civic engagement. Similar evidence emerges when considering how it affects the transition probabilities. These results are available from the authors upon request.

We also conducted an analysis estimating the causal HM model without posttreatment covariates. The results of the estimated ATEs are reported in Appendix 2. Without including posttreatment covariates, in addition to treatment in the structural part of the proposed model, we overestimate the ATEs. As stated in Rosenbaum (1984) this adjustment is recommended mainly when treatment groups differ in terms of posttreatment covariates, as in the present case. Some remarks on these results are can be found in Appendix 2.

5 Conclusions

In this study, we examined the influence of the practice of streaming at primary schools on civic engagement in adulthood. Streaming is a method used for grouping pupils into classes according to an overall assessment of their general abilities, namely, high, average, and low. The literature shows that ability-grouping affects educational achievement and academic self-concept. Our study evaluates whether ability-grouping affects civic participation due to having an effect on the individual's general self-concept. Our theoretical construct allows us to conjecture that high-ability students clustered in high-ability groups, who enjoy a higher reputation than low-ability students clustered in low-ability groups, can develop a better self-concept than their low-ability peers and that this higher self-concept translates into greater civic participation in adulthood.

Our analysis is novel within the literature on streaming for two reasons. First, the link between streaming and adult civic participation has remained rather unexplored; second, we offer estimations of the long-term causal effects of streaming, while in previous studies, the short-term view has prevailed.

To identify the effect of ability grouping (the treatment) on civic participation, we took advantage of the heterogeneity of British primary schools in terms of streaming during the 1960s. The dataset we employed (coming from the British National Child Development Study) provides information on the type of primary school attended by the cohort members (streamed or not streamed) and, in cases of streaming, on the ability class (high, average, and low) to which each individual was assigned. In addition, the dataset provides information on the civic and political participation of the cohort members during adulthood, together with other information of relevance to the analysis at the individual and family levels.

Given the longitudinal nature of the data, we accounted for multiple treatments and we proposed estimating the averaged treatment effects of streaming through a causal hidden (or latent) Markov model conceiving latent potential outcomes and defining required assumptions to correctly identify the causal effects while also controlling for posttreatment covariates. We handled missing responses under the missing-at-random assumption and for missing values on the posttreatment covariates through dummy indicators. In this model, which extends that proposed by Bartolucci et al. (2016), the generalized propensity score is estimated on the basis of the observed pretreatment covariates. Then, a weighted maximum likelihood estimation of the model parameters is carried out using as weights as the inverse of the propensity score; in addition, average treatment effects are estimated along with the effect of covariates on the initial and transition probabilities of the Markov chain. Another relevant aspect of this approach is that it is easy to apply as the computational tools required to implement it are freely available. This approach has some inferential advantages with respect to a previous proposal made by Favaro et al. (2020) that analyzed similar data because it allows us to account for missing values, and more importantly, the unobserved outcome, civic participation, is not summarized into a score. Due to its flexibility, the proposal is promising for identifying causal effects varying over time in a wide range of applications.

Overall, the study addresses two different research questions. First, we evaluated the average treatment effects of streaming on different degrees of civic engagement at the beginning of the observation period, when individuals were 33 years old by searching for groups of individuals with distinct behaviors. The degrees of civic engagement we defined according to the inferred three clusters are: “no engagement” (low degree of participation), “only voting” (medium degree of participation), “participating” (high degree of participation). Second, we evaluated the influence of streaming on the change in the degree of civic participation over the years (when respondents are 42 and 51 years old) by assessing the averaged treatment effect of streaming on the probability of transitioning (or not) toward clusters with higher or lower degrees of civic engagement. Regarding the first research question, the results of this study show that being enrolled in high-ability classes has a significantly positive effect (relative to the nonstreamed case) on the medium degree of participation, that is, “only voting”. Moreover, individuals who have attended high-ability classes are also twice as likely to have a high degree of civic participation (voting and association participation). Regarding the second research question, we did not detect any significant effect of streaming; this means that ability grouping when young does not change the degree of civic participation after age forty.

Our results are based on a specific cohort of people; however, they can be easily generalized to more recent cohorts of individuals. Indeed, inequality in early-age abilities has persisted during recent decades, and it has even increased because of increasing child poverty and disparities between children born in a country and those who have migrated to that country from elsewhere. Thus, in a broader context, polarization between different ability classes could potentially be greater now, and the related long-term effects could be exacerbated. These considerations appear even more important in light of the renewed interest in the practice of streaming. While comprehensive school reforms promoted during the 1970s questioned the application of streaming, the practice has regained widespread use worldwide in recent decades, albeit through different forms and stages of the educational pattern.

Appendix 1: Data Description

In the following we report some additional details and descriptions of the data presented in Sect. 2. In Fig. 1, we show the temporal data structure of the British National Child Development Study (NCDS), and we highlight the temporal periods of the pre- and posttreatment covariates, as well as of the indicators of civic engagement, along with the sample size. Table 11 reports the descriptive statistics of the pretreatment covariates and Table 12 describes those of the posttreatment covariates.

In particular, among the pretreatment covariates we also consider the time-varying region of living (not shown in Table 11), and we consider two dummy variables, among the posttreatment covariates, namely one for gender and the other for marriage or cohabitating status. A set of dummy variables is also available that indicates the qualifications of the individuals, which are defined according to the competence-based qualifications (NVQs). In particular, no qualification indicates that the individual has not achieved any work based qualification; an NVQ1 at level 1 is broadly equivalent to 3–4 GCSEs at grades D–G; NVQ2 to 4–5 GCSEs at grades A–C; NVQ3 to A levels; NVQ4 to a higher-education certificate; and NVQ5/6 to a higher-education diploma or degree. In terms of the International Standard Classification of Education (ISCED-97), we clarify that NVQ1 and NVQ2 correspond to ISCED level 2, with prevocational and vocational qualifications, respectively. NVQ3 corresponds to ISCED level 3, while NVQ4/5/6 are equivalent to ISCED levels 5 and 6.

We account for individuals' occupational status by considering three dummy variables, one for the inactivity/unemployed status, one for part-time occupation, and one for full-time occupation. We introduce a dummy variable that controls for the individuals' poor health status and we control for the region of residence by introducing time-varying indicators referring to the following regions: North, Yorkshire and Humberside, East Midlands, East Anglia, South East (including London), South West, West Midlands, North West, Wales, and Scotland. Finally, for checking robustness of the results, we also include a posttreatment covariate to the benchmark specification that accounts for the number of behavioral problems at age 33. This may be important in light of the literature suggesting that behavioral problems are malleable during childhood and adolescence and become stable at later ages (Holbein, 2017).

Appendix 2: Additional Results of the Causal Hidden Markov Model

In the following we show some additional results with respect to those presented in Sect. 4. Figure 2 depicts the Box plots referring to the propensity score (PS) estimates according to the three groups. We can notice a substantial overlap of the PS across individuals in each group.

Table 13 shows the descriptive statistics of covariates by treatment once they have been weighted with the estimated PS score weights. Table 14 shows the estimates of the intercepts of the logit models for the transition probabilities. Table 15, shows the proportions of individuals predicted to be in each state according to the estimated posterior probabilities using global decoding as illustrated in Sect. 3.3.

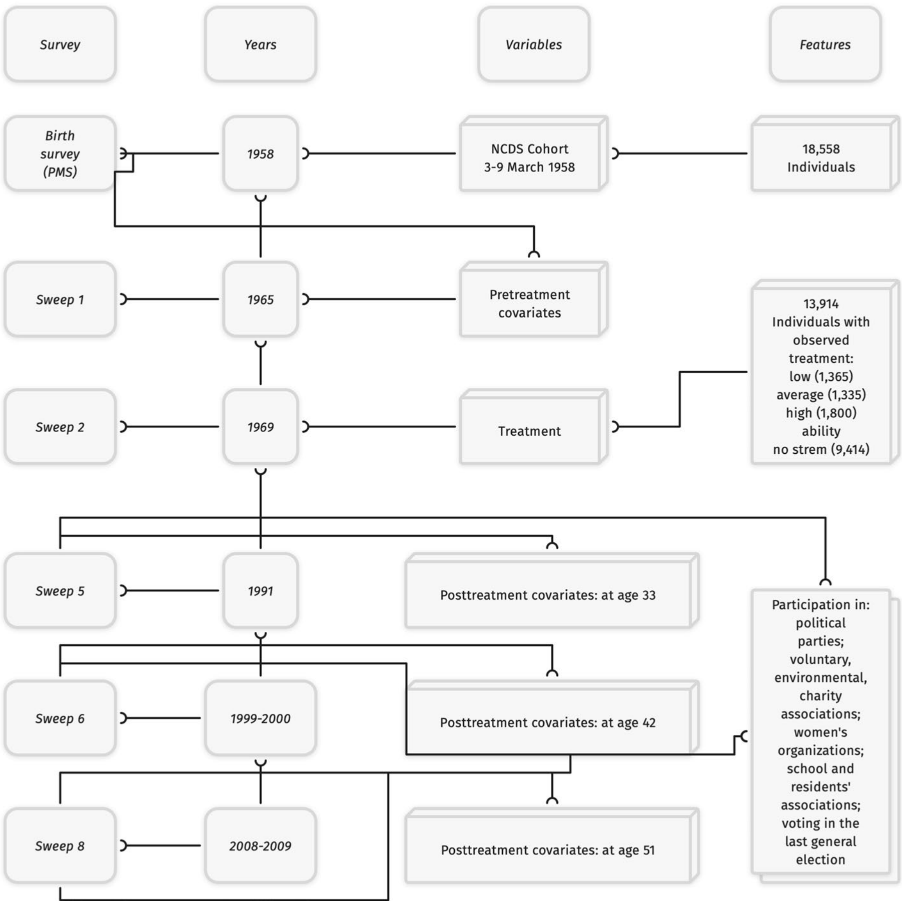

Fig. 1 Flowchart of the cohort study and temporal structure of the data from the British National Child Development Study (NCDS)

We can aslo estimate the most likely sequence of latent states for all sample units, as described in Sect. 3.3. This is the sequence of states that is most likely to have given rise to the observed sequence of responses, and it is estimated according to a recursive scheme provided by the Viterbi algorithm (Viterbi, 1967). Interestingly, in 1991 and 2000, individuals are mainly predicted as "only voting," while in 2009, they are predicted as "no engagement" and "only voting" with similar probability.

To assess the quality of the classification provided in Table 15 a measure may be computed on the basis of the posterior probabilities on all configurations of the latent states, given the observed data for every individual. When the components are well separated, the posterior probabilities tend to define a data partition, assuming values close to one and an entropy close to zero. An entropy measure for the proposed model is as follows:

Table 11 According to the treatment levels, descriptive statistics of the pretreatment covariates (measured at age 7 in 1965, except for the parents' social class, measured at birth)

Pretreatment covariates	Treatment				
	Low ability	Average ability	High ability	No stream	Row total
BSAG score	13.151	8.453	5.216	8.701	35.521
(s.d.)	(9.970)	(8.172)	(6.021)	(8.824)	–
Missing	1.037	0.754	0.931	6.880	9.602
Math score	1.021	4.852	6.393	7.061	19.327
(s.d.)	(2.281)	(2.500)	(2.701)	(2.832)	
Missing	0.812	0.934	1.021	7.065	9.832
Reading score	17.316	23.435	27.681	23.472	91.904
(s.d.)	(8.665)	(8.594)	(7.831)	(9.759)	–
Missing	0.913	6.926	0.980	0.751	9.570
Parents' social class					
I – Top	1.641	2.070	4.704	2.603	11.018
II	10.691	13.442	18.511	16.423	59.067
III	49.231	53.572	52.521	52.110	207.434
IV	18.691	15.983	13.601	15.149	63.424
V – Bottom	19.754	14.941	10.662	13.731	59.088
Missing	2.350	1.953	2.695	14.621	21.619
Father working class					
Managerial/technical	7.464	12.982	17.990	15.121	53.577
Partly skilled	23.183	17.883	12.834	18.328	72.228
Professional	1.931	3.870	7.862	5.191	18.854
Skilled	57.243	59.161	57.671	54.492	228.567
Unskilled	10.183	6.103	3.651	6.881	26.818
Missing	1.620	1.247	1.511	9.901	14.279
Parents' home ownership					
Owner	31.623	43.873	55.801	40.580	171.877
Rented	68.382	56.134	44.205	59.428	228.149
Missing	1.261	0.910	1.164	8.113	11.448
Mother > compulsory edu.					
No	85.713	77.223	66.061	75.081	304.078
Yes	14.291	22.781	33.940	24.922	95.934
Missing	0.601	0.384	0.683	3.541	5.209
Father > compulsory edu.					
No	86.571	81.012	67.560	76.734	311.877
Yes	13.431	18.991	32.448	23.276	88.146
Missing	1.626	1.234	1.463	10.182	14.505
Female					
No	58.974	53.858	47.666	50.647	211.145
Yes	41.026	46.142	52.334	49.353	188.855

Means and standard deviations are given for the continuous variables, proportions are given for categorical variables (%), and missing values are given as percentages

Table 12 Descriptive statistics of the posttreatment covariates (collected in the fifth sweep in 1991)

Posttreatment covariates	%
Female	48.612
Married	54.909
Missing	31.450
Education	
No qualifications	8.261
NVQ1	8.488
NVQ2	23.257
NVQ3	9.789
NVQ4	9.631
NVQ5/6	8.473
Missing	31.192
Employment status	
Full-time	43.237
Part-time	11.578
Unemployed	2.774
Inactive	11.219
Missing	31.192
Poor health	
Yes	1.207
Missing	31.357
Region of residence	
North	4.370
Yorkshire and Humberside	6.410
East Midlands	5.123
East Anglia	2.722
South East (including London)	21.051
South West	6.242
West Midlands	6.531
North West	6.802
Wales	3.892
Scotland	6.321
Missing	30.546

$$\mathrm{EN} = -\sum_i \sum_t \sum_h g_{h|i,t} \log g_{h|i,t}.$$

This quantity is compared with its maximum that is

$$\mathrm{EN}_{\max} = -\sum_i \sum_t \sum_h g^*_{h|i,t} \log g^*_{h|i,t} = nT \log k,$$

where $g^*_{h|i,t} = 1/k$, corresponding to the worst clustering capability of the model. The realized value of such a measure is 21,157, which should be compared with its maximum equal

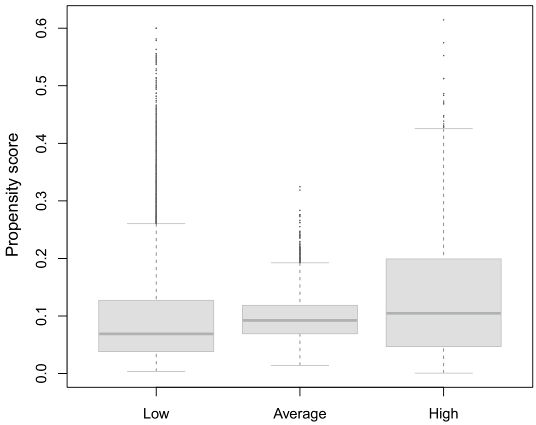

Fig. 2 Box plots showing the overlap of the estimated generalized propensity score over groups of streaming abilities: low, average, and high

to 45,858. When the modeling purpose is not related to the latent states prediction as it is in the present proposal this measure is not so relevant; for additional details, see Celeux et al. (2019).

Appendix 2.1. Results of the Model Without Posttreatment Covariates

The model with $k = 3$ states is estimated with treatments, including weights without posttreatment covariates that affect the latent potential outcomes. At the convergence of the EM algorithm, this model has a maximum log-likelihood of − 34,463, with 36 free parameters; the BIC index is equal to 69,288. Thus, we find an increase in the BIC index with respect to the model estimated with posttreatment covariates.

The subpopulations identified within this model are the same as those illustrated in Sect. 4.2. To compare the results with those described in the main text of the paper, we show the estimated ATEs for streaming in terms of initial and transition probabilities. In particular, Table 16 reports the ATEs for the initial probabilities. We notice that, unlike the previous model, all the coefficients are slightly higher than the previous ones in absolute value, and all of them are statistically significant except the coefficient referring to the logit of the average-ability comparing latent state 1 with latent state 2. Specifically, at the age of 33, having attended low and average-ability classes leads to lower probabilities of belonging to states labeled as “only voting" and “participating" while having attended high-ability classes leads to a higher probability of belonging to the state labeled “participating". These results align with our expectation since we are estimating a treatment effect that is not net of the other posttreatment covariates.

Table 17 lists the estimated ATEs under the causal model without posttreatment covariates on the transition probabilities from age 34 to age 51; the results reveal that none of the ATEs are significant, except the coefficient referring to low-ability at the transition from the first to the second latent state. Attending low-ability classes leads to a lower probability of belonging to the 2nd labeled “only voting".

Table 13 Descriptive statistics showing the weighted means and proportions of the pretreatment covariates (collected at age 7 in the second sweep in 1967, except for social class, collected at birth) according to treatment levels

Treatment				
Covariate	Low ability	Average ability	High ability	No stream
BSAG score	9.993	8.601	7.542	8.701
Math score	4.481	5.117	5.633	5.151
Reading score	21.131	23.342	25.221	23.420
Parents' social class				
I – Top	3.282	2.731	2.878	2.721
II	16.001	16.234	17.805	15.871
III	50.681	51.622	51.782	52.031
IV	15.273	14.982	13.671	15.341
V – Bottom	13.860	14.453	14.771	14.059
Father working class				
Managerial/Technical	14.283	14.147	16.672	14.571
Partly skilled	19.033	17.011	13.812	18.481
Professional	4.834	4.935	7.567	5.128
Skilled	54.733	57.113	56.143	55.133
Unskilled	7.140	6.812	5.821	6.701
Parents' home ownership				
Owner	42.714	42.446	44.937	42.123
Rented	57.294	57.560	55.072	57.883
Mother > compulsory education				
No	75.162	75.226	71.571	75.133
Yes	24.840	24.771	28.431	24.870
Father > compulsory education				
No	78.716	77.374	71.001	76.822
Yes	21.244	22.635	29.004	23.181
Female	48.201	50.364	45.131	48.505

Table 14 Estimates of the intercepts of the logit models (based on the difference between logits) for the transition probabilities of the latent process (with significance indicated at the †10%, *5%, and **1% level)

Row	Intercept 2	Intercept 3
1	1.492**	− 34.897
2	− 1.677**	− 23.729
3	13.314	− 10.922

Table 15 Relative frequencies of individuals assigned to each state at every time occasion

Latent state (h)			
Year	No engage. (1)	Only voting (2)	Participating (3)
1991	0.382	0.508	0.110
2000	0.395	0.500	0.105
2009	0.429	0.471	0.100

Table 16 ATEs of streaming under the causal model with treatment and weights at age 33 on civic engagement measured for the second and third latent states with respect to the first latent state (with significance indicated at the †10%, *5%, and **1% level)

Treatment	Logit 2	Logit 3
Low-ability	– 0.233**	– 1.048**
Average-ability	– 0.036	– 0.316**
High-ability	0.425**	0.433**

Table 17 ATEs of streaming under the causal model with treatment and weights at ages 42 and 51 on civic engagement measured on the transition probabilities based on the differences between logits (with significance indicated at the †10%, *5%, and **1% level)

Treatment	Diff. logit 2	Diff. logit 3
Low-ability	– 0.226**	3.886
Average-ability	0.093	0.133
High-ability	0.038	– 0.101

Acknowledgements The authors are grateful to the Centre for Longitudinal Studies (CLS), at the UCL Social Research Institute, for the use of data from the British National Child Development Study and to the UK Data Service for making the data available. However, neither the CLS nor the UK Data Service bear any responsibility for the analysis or interpretation of these data. F. Bartolucci and F. Pennoni acknowledge the financial support from the grant "*Hidden Markov Models for Early Warning Systems*" of Ministero dell'Università e della Ricerca (PRIN 2022TZEXKF) funded by European Union - Next Generation EU.

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement.

Declarations

Conflict of interest There are no conflicts of interest to declare.

References

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Chapman and Hall/CRC Press, Boca Raton, FL.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2014). Latent Markov models: A review of a general framework for the analysis of longitudinal data with covariates (with discussion). *TEST, 23*, 433–465.

Bartolucci, F., Pandolfi, S., & Pennoni, F. (2017). LMest: An R package for latent Markov models for longitudinal categorical data. *Journal of Statistical Software, 81*, 1–38.

Bartolucci, F., Pandolfi, S., & Pennoni, F. (2022). Discrete latent variable models. *Annual Review of Statistics and its Application, 9*, 425–452.

Bartolucci, F., Pennoni, F., & Vittadini, G. (2016). Causal latent Markov model for the comparison of multiple treatments in observational longitudinal studies. *Journal of Educational and Behavioral Statistics, 41*, 146–179.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics, 41*, 164–171.

Bojinov, I. I., Pillai, N. S., & Rubin, D. B. (2020). Diagnosing missing always at random in multivariate data. *Biometrika, 107*, 246–253.

Bowling, A., Pikhartova, J., & Dodgeon, B. (2016). Is mid-life social participation associated with cognitive function at age 50? Results from the British National Child Development Study (NCDS). *BMC Psychology, 4*, 1–15.

Cela, J. (2017). Evaluation of promotional campaign effects with self-selection of participation-propensity score application. *Albanian Journal of Mathematics, 11*, 35–71.

Celeux, G., Frühwirth-Schnatter, S., & Robert, C. P. (2019). *Model selection for mixture models - perspectives and strategies*. In: Frühwirth-Schnatter S, Celeux G, Robert CP (eds) Handbook of Mixture Analysis. Chapmann and Hall/CRC, New York, pp.118–154

Dardanoni, V., Modica, S., & Peracchi, F. (2011). Regression with imputed covariates: A generalized missing-indicator approach. *Journal of Econometrics, 162*, 362–368.

Dearden, L., Machin, S., & Reed, H. (1997). Intergenerational mobility in Britain. *The Economic Journal, 107*, 47–66.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B, 39*, 1–38.

Dockx, J., De Fraine, B., & Vandecandelaere, M. (2019). Tracks as frames of reference for academic self-concept. *Journal of School Psychology, 72*, 67–90.

Epple, D., & Romano, R. E. (2011). *Peer effects in education: A survey of the theory and evidence*. In: Benhabib J, Bisin A, O. Jackson M, (eds) Handbook of Social Economics. Elsevier, Amsterdam, pp. 1053-1163.

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The big-fish-little-pond effect on academic self-concept: A meta-analysis. *Frontiers in Psychology, 9*, 1–11.

Favaro, D., Sciulli, D., & Bartolucci, F. (2020). Primary-school class composition and the development of social capital. *Socio-Economic Planning Sciences, 72*, 1–26.

Flanagan, C., & Levine, P. (2010). Civic engagement and the transition to adulthood. *The Future of Children, 20*, 159–179.

Francis, B., Craig, N., Hodgen, J., Taylor, B., Tereshchenko, A., Connolly, P., & Archer, L. (2020). The impact of tracking by attainment on pupil self-confidence over time: Demonstrating the accumulative impact of self-fulfilling prophecy. *British Journal of Sociology of Education, 41*, 626–642.

Francis, B., Taylor, B., & Tereshchenko, A. (2019). *Reassessing 'ability' grouping: Improving practice for equity and attainment*. Routledge.

Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review, 57*, 812–828.

Hawkes, D., & Plewis, I. (2006). Modelling non-response in the national child development study. *Journal of the Royal Statistical Society, Series A, 169*, 479–491.

Holbein, J. B. (2017). Childhood skill development and adult political participation. *American Political Science Review, 111*, 572–583.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945–960.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika, 87*, 706–710.

Johnston, O., & Wildy, H. (2016). The effects of streaming in the secondary school on learning outcomes for Australian students–A review of the international literature. *Australian Journal of Education, 60*, 42–59.

Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics, 33*, 251–272.

Kiessling, L., & Norris, J. (2023). The long-run effects of peers on mental health. *The Economic Journal, 133*, 281–322.

Leite, W. (2016). *Practical Propensity Score Methods Using R*. Sage Publications, Thousand Oaks, California

Little, R. J. A., & Rubin, D. B. (2020). *Statistical Analysis with Missing Data*. Wiley.

Lu, G., & Copas, J. B. (2004). Missing at random, likelihood ignorability and model completeness. *The Annals of Statistics, 32*, 754–765.

Marsh, H. W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education, 28*, 165–181.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*, 280.

Marsh, H. W., Chessor, D., Craven, R., & Roche, L. (1995). The effects of gifted and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal, 32*, 285–319.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133–163.
Maruotti, A., & Punzo, A. (2021). Initialization of hidden Markov and semi-hidden Markov: A critical evaluation of several strategies. *International Statistical Review, 89*, 447–480.
Mulkey, L. M., Catsambis, S., Steelman, L. C., & Crain, R. L. (2005). The long-term effects of ability grouping in mathematics: A national investigation. *Social Psychology of Education, 8*, 137–177.
Palacios, D., Dijkstra, J. K., Villalobos, C., Treviño, E., Berger, C., Huisman, M., & Veenstra, R. (2019). Classroom ability composition and the role of academic performance and school misconduct in the formation of academic and friendship networks. *Journal of School Psychology, 74*, 58–73.
Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-group effects: Instructional, social, or institutional? *Sociology of Education* 27–46.
Parry, J., Brookfield, K., & Bolton, V. (2021). "The long arm of the household'': Gendered struggles in combining paid work with social and civil participation over the lifecourse. *Gender, Work and Organization, 28*, 361–378.
Pennoni, F., Paas, L. J., & Bartolucci, F. (2023). A causal hidden Markov model for assessing effects of multiple direct mail campaigns. *TEST,* 1–29.
Plowden, B. B. (1967). *Children and their primary schools: A report of the Central Advisory Council for Education*. England HM Stationery Office.
Putnam, R. D. (2000). *Bowling alone: The Collapse and Revival of American Community*. Simon and Schuster, New York.
R Core, T. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/
Robins, J. M. (2003). General methodological considerations. *Journal of Econometrics, 112*, 89–106.
Robins, J. M., Hernán, M. Á., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*, 550–560.
Robins, J. M., & Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association, 90*, 122–129.
Rogers, C. R. (1961). *On Becoming a Person: A Therapist's View of Psychotherapy*. Constable, London.
Rosenbaum, P. (2020). Modern algorithms for matching in observational studies. *Annual Review of Statistics and its Application, 7*, 143–176.
Rosenbaum, P., & Rubin, D. (2023). Propensity scores in the design of observational studies for causal effects. *Biometrika, 110*, 1–13.
Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A, 147*, 656–666.
Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association, 82*, 387–394.
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*, 688–701.
Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592.
Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling and decisions. *Journal of the American Statistical Association, 100*, 322–331.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.
Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*, 1–21.
Terrin, E., & Triventi, M. (2022). The effect of school tracking on student achievement and inequality: A meta-analysis. *Review of Educational Research, 93*, 236–274.
Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research, 37*, 483–504.
Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory, 13*, 260–269.
Welch, L. R. (2003). Hidden Markov models and the Baum–Welch algorithm. *IEEE Information Theory Society Newsletter, 53*, 1–13.
Wilkinson, I. A., Hattie, J. A., Parr, J. M., Townsend, M. A., Fung, I., Ussher, C., & Robinson, T. (2000). *Influence of peer effects on learning outcomes: A review of the literature*. Ministry of Education.