MOX-Report No. 14/2025

Stratified Multilevel Graphical Models: Examining Gender Dynamics in Education

Nicolussi, F.; Masci, C.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it https://mox.polimi.it

Stratified Multilevel Graphical Models: Examining Gender Dynamics in Education

Federica Nicolussi*, Chiara Masci†

February 28, 2025

Abstract

This study proposes a methodological approach to investigate gender disparities in education, particularly focusing on the schooling phase and its influence on career trajectories. The research applies multilevel linear models to analyse student performance concerning various factors, with a specific emphasis on gender-specific outcomes. The study aims to identify and test context-specific independencies that may impact educational disparities between genders. The methodology includes the introduction of supplementary parameters in multilevel models to capture and examine these independencies. Furthermore, the research proposes encoding these novel relationships in graphical models, specifically stratified chain graph models, to visualize and generalize the complex dependencies among covariates, random effects, and gender influences on educational outcomes.

Keyword:Context-specific independence, Multilevel models, Graphical models, Gender, Education.

1 Introduction

A central aim of Horizon Europe is to formulate effective strategies for integrating the gender perspective, which primarily involves delving deeper into research and innovation into the gender dimension [26]. Gender disparity manifests in various aspects of contemporary society, often reflecting deeply rooted structural inequalities. Some of the most important areas where it is evident include economic inequality, healthcare access, political representation, social and cultural norms, legal and institutional barriers, and education. The educational journey can be pinpointed as the initial stage where a gender disparity emerges, subsequently influencing career trajectories. Stereotypes regarding "appropriate" fields of study often contribute to the underrepresentation of women in Science,

*MOX - Department of Mathematics, Politecnico di Milano; federica.nicolussi@polimi.it

†Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano; chiara.masci@unimi.it

Technology, Engineering, and Mathematics (STEM). Gender studies have traditionally traced back these gender differences to disparities in educational outcomes [10]. While girls tend to outperform boys in reading, the gap in mathematics is structurally in favour of boys in most European countries [5]. Meanwhile, establishing a fair educational system is recognized as a necessary step in the political agendas of global economies [9].

The educational achievements of students result from a cumulative process in which various inputs are transformed into outputs. Key inputs and determinants of these outputs include socio-economic background, personal and psychological characteristics, school environment, students' habits and much more. These factors interact dynamically, influencing one another and evolving throughout a student's academic journey. Many educational and economical studies have analyzed the gender gap in education in terms of performances, but most of them assume that the determinants of educational achievement have the same impact across gender. While boys and girls share some common determinants, few studies show that significant gendered differences persist. For instance, studies in [1, 7] show that socioeconomic status (SES) and parental involvement significantly affect educational performance but operate differently by gender. Girls often benefit more from supportive parental involvement, especially when mothers are employed or educated, suggesting a stronger influence of maternal role modeling. Boys, however, appear more influenced by higher parental pressures within affluent families, and they may struggle more without such support, particularly in low-SES households. In terms of student perceptions and attitudes, the work in [27] shows that boys typically report greater self-efficacy in subjects like mathematics but are more likely to disengage academically when faced with challenges. Girls tend to underestimate their abilities in STEM, which can be compounded by higher levels of anxiety and stereotype threats in these areas. Girls' advantage in reading literacy sometimes translates into better performance in math items with high reading demands, illustrating an interplay between skills. Lastly, also school and classroom environment results to have gender-specific effects. Boys tend to benefit from active teaching methods and strong peer engagement in STEM. Conversely, girls often thrive when female teachers provide role models and construct gender-sensitive curricula, especially in traditionally male-dominated fields like STEM [14, 4].

These evidences highlight the importance of modelling student achievements' determinants with a special focus on the boys' and girls' heterogeneity, by assuming different educational production functions [2].
In this perspective, this research endeavours to elucidate the factors pertinent to the schooling phase that mostly affect students achievements, examining the ones that exert gender-specific effects on the outcome. We leverage data from the OECD Programme for International Student Assessment (PISA) database (`www.oecd.org/en/about/programmes/pisa.html`) of fifteen-years old students attending Italian schools. The aim of the paper is twofold. The former is to model the educational process by investigating the factors that influence its evolution, with a gender perspective. The latter is to test and visualize in an effective way the gender-specific de-

pendencies between students characteristics and achievements. To achieve this, we follow a structured two-steps pipeline. First, we identify key areas of student characteristics, categorizing them based on an assumed hierarchy of influence. This hierarchy distinguishes between *primary* characteristics - those inherent and immutable - and *secondary* characteristics, which may be shaped or influenced by the primary ones. At the highest level of this hierarchy is our main target variable: students' PISA mathematics scores. To explore the relationships among these characteristics, we build regression and classification models that, by including interaction terms, examine how the association among variables vary by gender. *Gender-specific* relationships are quantified using these interaction terms. When modelling the student performance, we take into account the nested nature of students within schools by means of a multilevel model [25], to disentangle the effects exerted by the two different data levels. Second, in order to properly capture the complexity of the phenomenon and, in particular, the gender-specific dependencies, we employ graphical models based on chain graphs. These models, originally theorized in studies employing the Lauritzen-Wermuth and Frydenberg (LWF) chain graph framework [17], [11], have since demonstrated their suitability for analyzing hierarchical data [12]. They leverage the graphical representation to illustrate various dependencies among covariates and random effects. This methodology offers powerful tools for investigating the relationship between covariates themselves and the effect of covariates on the target variable. The inclusion of the interaction term - with gender, in our specific case - into the models allows to identify and test the so-called *Context-Specific* Independence (CSI) [3], where the context is represented by gender.

The CSIs can be seen as particular conditional independencies that exist only for certain modalities of the conditioning variable. Under this setting, it is interesting to see where a certain mode of a variable blocks the relationship between two other factors. If so, we can argue that this particular characteristic plays a crucial role in determining independence. In our case study, identifying the CSIs related to gender provides a deeper understanding of the subtle, indirect ways gender differences manifest within the educational production function. These differences often go beyond direct measures of academic achievement, revealing nuanced pathways through which various factors influence the learning processes of boys and girls differently. By analyzing these associations, we can uncover how specific influences - such as socio-economic conditions, school environments, psychological traits, and cultural resources - interact with gender to shape educational behaviors and outcomes.

The proposed approach allows us to differentiate factors that are uniquely associated with boys or girls, highlighting variables that may require gender-specific attention, helping us identify factors that exhibit contrasting effects across genders. Such insights are crucial for designing targeted interventions aimed at reducing disparities and support equitable learning opportunities for all students. In the context of graphical models, the representation of CSIs, or asymmetrical independence relations in general, has been dealt with in different models. The representational aspect of these relationships is just as important as the mathematical formalization

of the problem, as it requires finding a way to add new information to the graph in a manner that ensures clarity, effectiveness, and ease of communication. Different ways have been proposed in the literature. In [3], the authors suggest to introduce supplementary conditional nodes for capturing CSIs in Bayesian Network. An alternative visualization, for undirected graphs, was proposed in [15], where the starting graph is enriched by split trees taking into account the different contexts (i.e. oriented arcs representing the levels of the variables are introduced). Other representations can be find in [16] with the introduction of nodes representing the categories. In [19] the CSIs are captured and learnt by means of the staged tree which is a suitable tool for representing asymmetrical relationships. In this work, we take advantage of the stratified graphs that were proposed in [23]. These objects add a label to certain edges reporting the conditions according to which the edge disappears. In [22], the stratified graph is extended to chain graphs in order to capture the CSIs in the graphical models proposed by [28]. In this work, we propose a new graphical model that exploits the stratified graph to represent the CSIs in multilevel block recursive regression models, namely stratified multilevel LWF chain graph models. The final model will therefore be able to consider variables belonging to different blocks whose system of relationships is more complex than the classical regression model. It will be able to capture the unobservable effect due to subjects belonging to different groups. Finally, it will be able to represent both the direct effect of gender on the other variables and indirectly by studying the changing relationships when gender changes.
The paper is structured as follows: in Section 2 we present the OECD-PISA dataset we employ in the analysis; Section 3 describes the methodology and Section 4 reports the results; in Section 5 we draw our conclusions.

2 OECD-PISA Data

OECD PISA assesses student performance, on a triennial basis, in science, mathematics, reading, collaborative problem solving and financial literacy, since 2000. The ability of 15-year-olds to use their skills to meet real-life challenges are measured in these subjects. For our analysis, we observe 2018 data. The survey [24] covers the entire Italian country, analysing 6455 students, attending 512 schools, scattered throughout the country. The sample includes students attending both general and vocational schools, but, given the different educational programs offered by the two types of schools, we opted to analyse students data only from general schools[1]. After deleting students with missing values in the variables of interest, our sample consists of 3465 students attending 238 general schools. The main target variable is the maths PISA test score (*math_PISA_score*) obtained by the student, that takes values on a scale [0,100]. Together with it, students, parents and school principals have to fill out tailored questionnaires that create a huge and rich database. The variables that we include in our analysis regard different areas. In particular, we con-

[1] A supplementary analysis can be performed on students attending vocational schools.

sider ($\mathbf{X}_1$) demographic characteristics (e.g., gender, socioeconomic index, cultural possession at home, parents' educational level, immigrant status and school grade), ($\mathbf{X}_2$) students' habits (study time, use of internet and video games), ($\mathbf{X}_3$) a series of indicators built by PISA summarizing the perception that the student has of the school he/she is attending and the support received by teachers, and ($\mathbf{X}_4$) a component inherent to the psychological characteristics of the student and to his/her motivation at school. Table 1 reports the list and description of the variables used in the analysis, while Table 2 reports their descriptive statistics, for the entire sample and separately for boys and girls.

	Variable	**Description**	**Type**
	school_id	Anonymous school ID	factor
	math_PISA_score	Maths PISA student score	num
$\mathbf{X}_1$	gender	Student gender (0 = male, 1 = female)	binary
	immig	Immigrant status (0 = native Italian, 1 = 1^{st} or 2^{nd} generation immigrant)	binary
	misced	Student mother's level of education on a scale [0-6] (0 = primary education not completed, 6 = postgraduate)	int
	cultural_possession	Std. ind. of cultural possessions at home	num
	home_educ_resources	Std. ind. of home educational resources	num
	ESCS	Std. ind. of economic, social and cultural status	num
	late	Indicator for late-enrolled student	binary
	early	Indicator for late-enrolled student	binary
$\mathbf{X}_2$	video_games	Indicator of video-game devices user	binary
	internet	Indicator of frequent internet user for fun	binary
	mmins	Learning time (minutes per week) - mathematics	num
	tmins	Learning time (minutes per week) - total	num
$\mathbf{X}_3$	sc_DISCRIM	Std. ind. of perceived discriminatory school climate	num
	sc_BELONG	Std. ind. of the sense of belonging to school	num
	sc_PERCOMP	Std. ind. of perceived competitiveness in the school	num
	sc_PERCOOP	Std. ind. of the perceived cooperation climate in the school	num
	TEACHSUP	Std. ind. of the perceived teacher support	num
$\mathbf{X}_4$	EMOSUPS	Std. ind. of perceived parents' emotional support	num
	COGFLEX	Std. ind. of student's cognitive flexibility	num
	GFOFAIL	Std. ind. of the student fear of failure	num
	COMPETE	Std. ind. of student competitiveness	num
	EUDMO	Std. ind. of student 'Eudaemonia': meaning in life	num

Table 1: List and explanation of the student-level variables used in the analysis. Legend: Std. ind. = Standardized indicator built by OECD-PISA.

3 Method

3.1 Multilevel model with context-specific constraints

Let us consider a target $n \times 1$ continuous variable $\mathbf{Y}$ and a set of covariates collected in the $n \times p$ matrix $\boldsymbol{X}$, that can be either continuous or categorical. The set of possible values assumed by the generic X_k is $\mathbb{R}$,

	Variable	% / Median (IQR) All	% / Median (IQR) Girls	% / Median (IQR) Boys
	math_PISA_score	0.46 (-0.15, 1.03)	0.31 (-0.28, 0.87)	0.73 (0.15, 1.29)
$\mathbf{X}_1$	gender	1: 62%; 0:38%	-	-
	immig	1: 6.2%; 0: 93.8%	1: 6%; 0: 94%	1: 6.3%; 0: 93.7%
	misced	4 (4, 6)	4 (4, 6)	4 (4, 6)
	cultural_possession	0.46 (0.05, 0.84)	0.45 (0.05, 0.84)	0.52 (0.05, 0.84)
	home_educ_resources	1.18 (-0.14, 1.18)	1.18 (-0.14, 1.18)	1.18 (-0.14, 1.18)
	ESCS	0.09 (-0.54, 0.83)	0.01 (-0.61, 0.74)	0.25 (-0.40, 0.92)
	late	1: 7.3%; 0: 92.7%	1: 6.2%; 0: 95.8%	1: 9%; 0: 91%
	early	1: 7.7%; 0: 92.3%	1: 7.9%; 0: 92.1%	1: 7%; 0: 93%
$\mathbf{X}_2$	video_games	1: 60%; 0: 40%	1: 48%; 0: 52%	1: 78%; 0: 22%
	internet	1: 77%; 0: 23%	1: 74%; 0: 26%	1: 81%; 0: 19%
	mmins	200 (165, 275)	180 (165, 250)	240 (180, 300)
	tmins	1,620 (1,540, 1,800)	1,620 (1,540, 1,800)	1,620 (1,485, 1,750)
$\mathbf{X}_3$	sc_DISCRIM	-0.53 (-1.15, 0.21)	-0.66 (-1.15, 0.20)	-0.42 (-1.15, 0.48)
	sc_BELONG	0.05 (-0.44, 0.68)	0.03 (-0.45, 0.68)	0.16 (-0.34, 0.68)
	sc_PERCOMP	-0.23 (-0.61, 0.55)	-0.23 (-0.77, 0.29)	-0.23 (-0.61, 0.69)
	sc_PERCOOP	-0.30 (-0.94, 0.60)	-0.30 (-0.94, 0.60)	-0.28 (-0.94, 0.60)
	TEACHSUP	-0.03 (-0.63, 0.70)	-0.09 (-0.71, 0.70)	0.00 (-0.60, 0.70)
$\mathbf{X}_4$	EMOSUPS	0.22 (-0.66, 1.03)	0.22 (-0.66, 1.03)	0.21 (-0.66, 1.03)
	COGFLEX	-0.41 (-0.83, 0.22)	-0.46 (-0.87, 0.18)	-0.25 (-0.83, 0.22)
	GFOFAIL	0.05 (-0.61, 0.46)	0.11 (-0.38, 0.82)	-0.20 (-0.69, 0.46)
	COMPETE	0.20 (-0.57, 0.79)	-0.10 (-0.57, 0.57)	0.20 (-0.27, 1.18)
	EUDMO	-0.27 (-0.98, 0.26)	-0.31 (-0.98, 0.26)	-0.17 (-0.86, 0.26)

Table 2: Descriptive statistics of student-level variables. Binary variables are described as percentages, while numerical variables as median and IQR. Descriptive statistics are reported for the entire sample and separately for girls and boys. All OECD-PISA numerical indicators are standardized.

if X_k is continuous, and $\mathcal{I}_k = \{1, 2, \ldots, I_k\}$, if X_k is categorical. A linear model is defined as:

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta}_1 + \boldsymbol{\epsilon} \qquad (1)$$

where $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1)$ is the $(p+1)-$dimensional vector of parameters and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \Sigma)$ is the error vector.

In such a model, CSIs can be identified by including interaction terms. In particular, the CSIs that we aim to identify are a generalization of conditional independence relationships in the following form. Let $\boldsymbol{X}$ be a vector of random variables, let A, B, and C be three subsets of variables' indices, and $\boldsymbol{X}_C$ be a subset of categorical variables taking values in $\mathcal{I}_C = \times_{k \in C} \mathcal{I}_k$, where $\times$ denotes the Cartesian product between all vectors $\mathcal{I}_k$ with $k \in C$. Then, a CSI is defined as:

$$\boldsymbol{X}_A \perp \boldsymbol{X}_B | \boldsymbol{X}_C = \boldsymbol{i}_C, \qquad \boldsymbol{i}_C \in \mathcal{K}_C \subset \mathcal{I}_{\mathcal{C}}, \qquad (2)$$

where $\mathcal{K}_C$ is the subset of values taken by $\boldsymbol{X}_C$ for which conditional independence in the formula holds. For each value of $\boldsymbol{X}_C$ in $\mathcal{I}_C \backslash \mathcal{K}_C$ the independence in formula 2 is no longer verified[2]. This concept can

[2]The definition of context-specific independence can also be extended to continuous conditioning variables, see [23], but it is beyond the scope of this paper.

be extended in the contest of regression and classification models where a continuous or categorical variable of interest Y is regressed against a set of P predictors $\mathbf{X}$ and we are interested in estimating the association between Y and $\mathbf{X}$, conditionally on a dichotomous - without loss of generality - variable $\tilde{X}$. In this case, the CSI for each $p = 1, \ldots, P$, can be expressed as:

$$Y \perp X_p | \boldsymbol{X}_{-p}, \quad \text{when } \tilde{X} = 0, \qquad Y \not\perp X_p | \boldsymbol{X}_{-p}, \quad \text{when } \tilde{X} = 1, \tag{3}$$

or vice-versa, where $\boldsymbol{X}_{-p}$ contains all predictors P but the $p-$ th and $\tilde{X}$ is the conditioning variable for which we want to investigate the CSI relationships, which plays the role of $\boldsymbol{X}_C$ in formula 2. In order to estimate these CSIs, model in Eq. 1 can be modified as[3]:

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta}_1 + (\mathbf{X}_{-\tilde{X}} \times \tilde{X})\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}. \tag{4}$$

where we assume that $\tilde{X}$ is one of the variables contained in $\mathbf{X}$ and $\mathbf{X}_{-\tilde{X}}$ is the covariates matrix without the column referring to $\tilde{X}$. By introducing the interaction term between each variable in $(\mathbf{X} \backslash \tilde{X})$ and $\tilde{X}$, we are able to investigate the association between each variable in $(\mathbf{X} \backslash \tilde{X})$ and Y, conditionally on $\tilde{X}$.

When data have a hierarchical structure, the procedure can be extended as follows. Let us consider a target continuous variable Y_{ij} for the $i - th$ subject (first level unit) of the $j - th$ cluster (second level unit), with $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$ and a set of covariates $\boldsymbol{X}_{ij} = (X_{1,ij}, \ldots, X_{k,ij}, \ldots, X_{p,ij})$ described as above. The multilevel linear model [25] is defined as:

$$Y_{ij} = \beta_0 + b_{0j} + (\boldsymbol{\beta}_1 + \boldsymbol{b}_{1j})\boldsymbol{X}_{ij} + \epsilon_{ij} \tag{5}$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$ are the first level residuals and $\boldsymbol{b}_j = (b_{0j}, \boldsymbol{b}_{1j}) \sim N(\mathbf{0}, \Omega_{(p+1)\times(p+1)})$ are the second level residuals, for each cluster j. Further, the correlation between the residuals and the covariates is assumed to be null. The CSIs for the conditioning variable $\tilde{X}$ can be derived from the following formula:

$$Y_{ij} = \beta_0 + b_{0j} + (\boldsymbol{\beta}_1 + \boldsymbol{b}_{1j})\boldsymbol{X}_{ij} + (\boldsymbol{\beta}_2 + \boldsymbol{b}_{2j})\tilde{X}_{ij}\boldsymbol{X}_{-\tilde{X},ij} + \epsilon_{ij}. \tag{6}$$

The term $(\boldsymbol{\beta}_2 + \boldsymbol{b}_{2j})\tilde{X}_{ij}\boldsymbol{X}_{-\tilde{X},ij}$ models the interaction between $\tilde{X}$ and the other covariates, investigating both the heterogeneity at individual level (through the fixed-effects $\boldsymbol{\beta}_2$) and at second level (through the random effects $\boldsymbol{b}_{2j}$). On one hand, for each $k = 1, \ldots, p$, having $\beta_{1k} + b_{1kj} = 0$ suggests that, when $\tilde{X}_{ij} = 0$, the variable $X_{k,ij}$ has no effect on the target variable Y_{ij}. On the other hand, having $\beta_{1k} + b_{1kj} + \beta_{2k} + b_{2kj} = 0$

[3]The generalization to the case of a categorical response Y is straightforward.

suggests that the previous independence holds when $\tilde{X}_{ij} = 1$.

Definition 3.1 Given the multilevel model in formula 6, the CSIs in formula 3 correspond to the following constraints:

- $Y_{ij} \perp X_{kij} | \tilde{X}_{1ij} = 0$ corresponds to $\beta_{1k} + b_{1kj} = 0$
- $\beta_{1k} + b_{1kj} + \beta_{2k} + b_{2kj} = 0$ corresponds to $Y_{ij} \perp X_{k,ij} | \tilde{X}_{ij} = 1$

Therefore, by means of hypothesis tests on these parameters, we are able to identify the CSIs, investigating both the average CSIs in the population (identified by the parameters $\boldsymbol{\beta}$) and cluster-specific CSIs (identified by the parameters $\boldsymbol{b}$).

3.2 Chain Graphical Models

Basic notation on LWF chain graph model Graphical models are tools that exploit graphs to give an effective representation of how a set of variables interacts. For the future understanding of the work, it is necessary to spend a few lines recalling definitions from graph theory. The notation used echoes the work of [11].

A *graph*, G, is a mathematical object composed of two sets, one of vertices V and edges E. The edges can be undirected, i.e. if $e_{ij} = (v_i, v_j) \in E$ then $e_{ji} = (v_j, v_i) \in E$ and represented by a segment, or they can be directed, i.e. if $e_{ij} \in E$, then $e_{ji} \notin E$ and represented by an arrow.

Two vertices v_i and v_j are *adjacent*, $v_i \sim v_j$, if $e_{ij}, e_{ji} \in E$. If $v_i \to v_j$, v_j is a *parent* of v_i and v_i is a child of v_j. A *path* is a sequence of consecutive edges, while a *directed path* must respect the direction of the arrow.

A graph is entirely described by the nature of its arcs, so if all the arcs are undirected, the graph will be *undirected*, if all the arcs are directed you will have an *directed* graph. The latter becomes a directed acyclic graph if cycles -directed paths beginning and ending in the same node- are forbidden.

In an undirected graph, a set of nodes C *separates* two disjoint subsets of vertices A and B if each path between one node in A to one node in B passes from some nodes in C.

Let A be a subset of the nodes in V, the *induced sub-graph* $G_A = \{A, E_A, \}$ where $E_A = E \cap (A \times A)$.

A *chain graph*, $G = (V, E)$, is a graph where the set of vertices V is clustered in different components, namely chain blocks, $\mathcal{T} = \{T_k\}$, and the set of edges, E, contains undirected arcs if the linked vertices belong to the same chain block, directed arcs otherwise. In addition, the cycles and semi-cycles are forbidden. The chain blocks form a partially ordered set where a block descends from another only if there is a directed path between the two components. of a set A adds to the boundary the set itself, i.e. $cl(A) = bd(A) \cup A$. The *ancestral*, $an(A)$ set of A is obtained by removing all the chain blocks descending from A.

Finally, the definition of *moral graph*, G^m, for a chain graph extends the one provided for the directed acyclic graph, [17], i.e., it is obtained by applying two steps: all the directed edges are replaced by undirected edges and additional edges are added between unlinked nodes that are parents

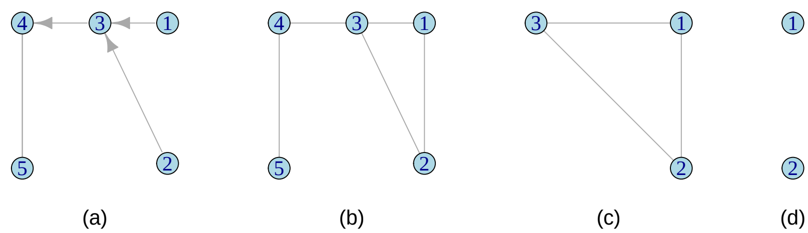

Figure 1: (a) Chain Graph G; (c) moral graph $(G_{an(123)})^m$; (c) moral ancestral graph $(G_{an(123)})^m$; (d) moral ancestral graph $(G_{an(12)})^m$

of the same node in the original chain graph G.

Example 1
Figure 1 (a) shows a chain graph where the blocks are formed by $(1, 2)$; (3) and $(4, 5)$ respectively. The moral graph associated with the entire graph is shown in Figure 1 (b), where the arc between nodes 1 and 2 is added because they both point to the same node. Ancestral sets are obtained by recursively removing the terminal components. Thus we get $(1, 2, 3)$ and $(1, 2)$. The correspondent moralized induced subgraphs are depicted in Figure 1 (c) and (d).
Let $\boldsymbol{X} = (X_1, \ldots, X_p)$ be a random vector obeying a probability distribution $\mathcal{P}$ of both categorical and quantitative variables. A graphical model able to describe the joint distribution of $\boldsymbol{X}$ as a recursive multivariate regression model is presented by the chain graph model proposed by [17] and [11], hereafter LWF chain graph model. In this model, the chain blocks represent the regression structures. Specifically, oriented arcs define the relationship between dependent variable, represented by the node to which the arc points, and covariates, represented by all nodes from which the arc starts. This allows us to describe through a single graph, complex systems where some variables play only the role of targets (final component of the graph), and some variables play only the role of covariates (initial components of the graph). All remaining components can be in turn either targets or covariates, depending on the subset of variables considered. Thus, if $A, B \subseteq V$ are two disjoint sets of vertices corresponding to two components and the oriented arcs start from the nodes in A and point to the nodes in B, then $\boldsymbol{X}_B$ represents a set of covariates for $\boldsymbol{X}_A$.
As detailed in [17], the variables depicted in the graph can be either qualitative or quantitative. In the graph, the different nature of the variable is represented by a different color of the vertex contour, black for qualitative variables and grey for quantitative variables.
The global LWF Markov property, also known as the block concentration Markov property, states the rules to express separation sets in a graph as independencies in the probability distribution $\mathcal{P}$.

Definition 3.2 Let $G = \{V, E\}$ be a chain graph. The joint probability $\mathcal{P}$ of a random vector $\boldsymbol{X}$ is said *global* $G-$Markovian if $\boldsymbol{X}_A \perp \boldsymbol{X}_B | \boldsymbol{X}_C$ whenever C separates A and B in $(G_{an(A \cup B \cup C)})^m$, where $(G_{an(A \cup B \cup C)})^m$ is the moralized subgraph induced by the smallest ancestral set containg $A \cup B \cup C$.
Example 2
Consider a random vector $\boldsymbol{X} = \{X_1, \ldots, X_5\}$ represented by the graph in Figure 1. $\boldsymbol{X}$ is globally G-Markovian if its probability function $\mathcal{P}$ satisfies the list of independencies obtained by applying the definition 3.2. Thus, from the moral graph in (b), we find that 3 separates the sets $(4, 5)$ and $(1, 2)$, thus $\boldsymbol{X}_{45} \perp \boldsymbol{X}_{12} | X_3$. Further, (4) separates (5) and $(3, 4, 5)$ which implies $X_5 \perp \boldsymbol{X}_{123} | X_4$. The moral graph in (c) is complete thus there are no separator sets. Finally, from the last moral graph in (d), we obtain the marginal independence $X_1 \perp X_2$.

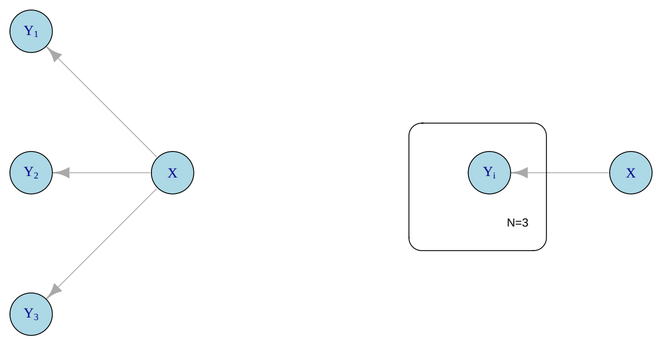

Figure 2: (a) Graph with nodes representing individual variables Y_1, Y_2 and Y_3; (b) the correspondent plate notation.

Multilevel chain graph models Chain graphical models adapted to describe the relationships of a mixed model expressed in formula 1, were proposed in [12] and [13]. Here, the authors adapt the LWF chain graphical models to consider the dependence among data within each cluster. With this aim, they propose adding one extra node for each random effect. These additional nodes represent unobserved variables, thus we use a squared shape for the node, instead of the classical circle, to depict the hidden nature of the variable.To emphasize independence among observations conditioned on random effects, it was essential to introduce individual nodes for the response variable, representing the target variable for each statistical unit.
In this work, we take advantage of the plate notation to represent the individual nodes, for each observation in the cluster $j-$th and the random effect associated with each cluster. The plate notation consists of replacing a series of repeated variables with a single indexed node enclosed in a rectangle specifying the number of variables that the single node represents. An example is shown in Figure 2 where on the left are represented N conditionally independent random variables given the variable b, while on the right, the same relationship is synthesized with the plate representing a single node with the index N indicating the number of synthesized variables.

Example 3 The graph in Figure 2 (a) represents the variable Y observed on 3 subjects and a variable X. By conditioning concerning the variable X, the Y_i are mutually independent. This situation is summarised in graph (b) where the generic variable Y_i is shown inside a plate specifying the maximum value assumed by the index i.

Stratified LWF chain graph model. We are now extending the multilevel LWF chain graph model to also incorporate CSIs, as specified in the model outlined in formula 4 through the parameter constraints described in definition 3.1. In this context, the only variable with an undirected effect captured through CSIs is the dichotomous variable $\tilde{X}$.

To represent this new type of independence, we require a mathematical structure that accounts for the presence or absence of an arc based on specific values taken from a subset of the parent nodes. In [22], a stratified graph was proposed to encapsulate the Markov property discussed in [28]. In this work, we propose utilizing a stratified graph for this different graphical model.

The fundamental idea behind these graphs is to add labeled edges, referred to as strata, to the traditional edges, which can be either directed or undirected. The label on these edges indicates the conditions under which the arc is deleted from the graph. Each of these representations illustrates the graph in a distinct context.

By considering the CSIs in formula 3, the class of labeled arcs, L, contains bi-variate objects of the form $L_{ij}(\tilde{x}) = (e_{ij}, \tilde{x})$, where the edge that is present in the labeled arc, $e(L_{ij}(\tilde{x}))$, belongs to the set of edges, i.e. $e(L_{ij}(\tilde{x})) = e_{ij} \in E$, and $\tilde{x} \in \{0, 1\}$. Then, we can define the stratified graph as follows.

Definion 3.3 A stratified CG, $SG = \{V, E, L\}$, is a chain graph with nodes V and edges E, and with a list of labeled arcs reported in L. An example of SG is provided in Figure 3 where the class of labeled arcs is composed only of one element $L_{Y,X_1}(\tilde{x} = 1)$. Thus, when the variable $\tilde{X} = 1$ the graph changes and the arc between Y_{ij} and X_1 vanishes.
A Stratified LWF (S-LWF) chain graph model takes advantage of the SG to represent the joint probability distribution $\mathcal{P}$ of a random vector. This new configuration is able to represent a model as depicted in formula 6 and its generalization.
The SG in Figure 3 represents the pure response variable Y_{ij} in double plate, one for the individuals indexed with $i = 1, \ldots, n_j$ and one for the clusters indexed with $j = 1, \ldots, J$. The last plate also contains the random intercept b_{0j}, in a squared node to highlight its hidden nature. In addition, the remaining nodes represent two pure covariates, $\tilde{X}$ and X_1. The color of the contour of the nodes suggests the nature of the variables, *black* for the qualitative and *gray* for the quantitative ones.

To clarify the relationship underlying the S-LWF chain graph model, we extract a *full* chain graph and a series of *reduced* chain graphs to cover any situation. The *full* chain graph is obtained by replacing any labeled arc with the corresponding unlabeled arc and represents the classical conditional independencies. In constrast, for any strata, it is possible to build a *reduced* graph, obtained by deleting the corresponding labeled arc and replacing the other labeled arcs with the corresponding unlabeled arcs. These *reduced* graphs represent the CSIs.
Given a stratified chain graph, $SG = \{V, E, L\}$, *full* chain graph is $G^F = \{V, E\}$. Given a stratified chain graph, $SG = \{V, E, L\}$, for any $\tilde{x}$ such that there is at least one edge e_{ij} with $L_{ij}(\tilde{x}) \neq \emptyset$, the *reduced* chain

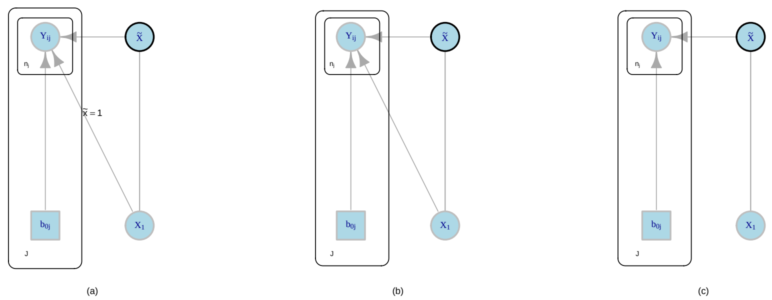

Figure 3: (a) Stratified graph SG with $L = \{(Y_{ij}, X_1); \tilde{x} = 1\}$, individual nodes Y_{ij} and b_{0j} are placed in 2 plates, the first representing individuals $i = 1, \dots, n_j$, and the second representing clusters $j = 1, \dots, J$; (b) the correspondent full graph G^{full}; (c) the reduced graph $SG^R(L_{YX_1}(\tilde{x} = 1))$.

graph, $G^R(\tilde{x}) = \{V, E^R_{\tilde{x}}\}$, where the element of $E^R_{\tilde{x}}$ in position (i, j), $e^R_{ij} = e_{ij} I_{[L_{ij}(\tilde{x}) = \emptyset]}$. Thus, the generic element e^R_{ij} of $E^R_{\tilde{x}}$ is null if $e(L_{ij}(\tilde{x})) \neq 0$, and equal to the correspondent element on the matrix E otherwise.

The Markov property for a S-LWF chain graph model is derived by applying the LWF MP in formula 3.2 to the corresponding *full* and *reduced* chain graphs,the resulting list of independencies hold for the general context (**G1**) and for each specific context (**G2**).

Definition 3.4 Given a stratified chain graph, $SG = \{V, E, L\}$, and a random vector $\boldsymbol{X}$ with joint probability distribution P, we say that P is *global* Markovian w.r.t. SG if:

G1) $X_A \perp X_B | X_C$ whenever A and B are separated in $(G^{full}_{an(A \cup B \cup C)})^m$ and

G2) $X_A \perp X_B | \tilde{X} = \tilde{x}$, whenever $\tilde{X}$ separates A and B in $(G^R_{an(A \cup B \cup \tilde{X})}(\tilde{x}))^m$ for any $\tilde{x}$ such that there is at least one edge e_{ij} with $L_{ij}(\tilde{x}) \neq \emptyset$.

Example 4
Figure 3 (b) reports the full graph of (a). The structure of the two graphs is identical, the only distinction being the absence of the labels in the latter. From this figure, we can identify the general independences that hold for each context. In this case, we get that the response variables Y_{ij} of all subjects $i = 1, \dots, n_j$ within the same group j are mutually independent given the random effect b_{0j} and the covariates. Further, the random effect is independent of the covariates $\tilde{X}$ and X_1. The graph in Figure 3 (c) reports the reduced graph in the context of $\tilde{X} = 1$. In this reduced graph, the missing edge suggests that $Y \perp X_1 | \tilde{X} = 1$. The

following theorem establishes the connection between the graph structure and the constraints on the parameters in model 4.

Theorem 1 Let $SG = \{V, E, L\}$ be a stratified chain graph, and let $\boldsymbol{X}$ be a random vector with joint strictly positive probability distribution $\mathcal{P}$. Then the following statements are equivalent:

C1 $\boldsymbol{X}$ is globally SG-MP according to the definition 3.2;

C2 The parameters in the (multilevel) regression model as in definition 3.1 holds 0 for all separator sets

Proof. **From C1 to C2**: Consider the MPs in definition 3.2. If $\boldsymbol{X}$ is *globally* SG-Markovian, then, according to **G1**, $\boldsymbol{X}$ is also G-Markovian w.r.t. the full graph $G^{full} = \{V, E\}$. Independencies identified in this manner constrain the parameters of a block regression to zero as described in [17] and [12]. Further, according to **G2**, the conditional distribution of $\boldsymbol{X}\backslash\tilde{X}|\tilde{X} = \tilde{x}$ is also G-Markovian w.r.t. any reduced graph $G^R = \{V, E\}(\tilde{x})$, for $\tilde{x} \in \{0, 1\}$. This result leads to constraints on parameters that are not available on the classic linear regression model because it should involve the variable $\tilde{X}$ and the other variables included in $\boldsymbol{X}_A$ and $\boldsymbol{X}_B$ according to **G2**. These additional parameters are the ones added to the model 4 and denoted with β_2 and the possible random slope b_2.
From C2 to C3: As outlined in [17] and [12], we implement a block regression procedure that expresses one dependent variable as a function of the variables represented by nodes in the same component, as well as those in components that have at least one directed arc pointing to the aforementioned component. This approach allows us to model the dependence between two variables represented by nodes in the same component and in different components, using specific parameters. Based on the results in definition 3.1, certain constraints on these parameters lead to CSI statements that correspond to those depicted in the SG model.

□

3.3 Learning procedure

To investigate the CSIs of interest, we construct a set of models that relies on specific assumptions about the educational production functions, particularly emphasizing the sequential nature of student characteristics. The assumed structure is illustrated in Figure 4. Specifically, we make the following assumptions:

- **Variables in $\mathbf{X}_1$**
  - These represent the initial characteristics of students, which are immutable and cannot be predicted by any other variables.
- **Variables in $\mathbf{X}_4$**
  - These are predictable based on other variables within $\mathbf{X}_4$ and variables in $\mathbf{X}_1$
  - **Model for $\mathbf{X}_4$**: For each variable in $\mathbf{X}_4$, we build a regression or classification model where predictors include variables in $\mathbf{X}_1$ and the remaining variables within $\mathbf{X}_4$.

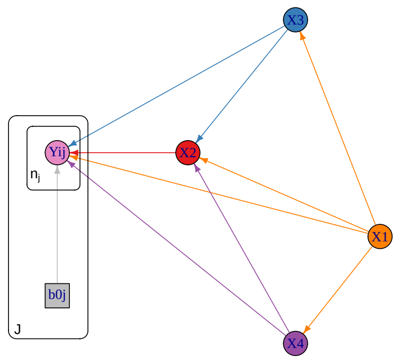

Figure 4: Directed acyclic graph representing the assumption on the graph structure.

- **Variables in $\mathbf{X}_3$**
  - These are predictable based on other variables within $\mathbf{X}_3$ and variables in $\mathbf{X}_1$
  - **Model for $\mathbf{X}_3$**: For each variable in $\mathbf{X}_3$, we build a regression or classification model where predictors include variables in $\mathbf{X}_1$ and the remaining variables within $\mathbf{X}_3$.
- **Variables in $\mathbf{X}_2$**
  - These are predictable based on other variables within $\mathbf{X}_2$ and variables in $\mathbf{X}_1$, $\mathbf{X}_3$ and $\mathbf{X}_4$
  - **Model for $\mathbf{X}_2$**: For each variable in $\mathbf{X}_2$, we build a regression or classification model where predictors include variables in $\mathbf{X}_1$, $\mathbf{X}_3$, $\mathbf{X}_4$ and the remaining variables within $\mathbf{X}_2$.
- **Y** (`math_pisa_score`)
  - This is predictable based on variables in $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$
  - **Model for** Y: We construct a multilevel regression model where predictors include variables in $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$, along with a random intercept[4] to account for group-level (i.e., school-level) effects[5].

[4] More complex models that include random slopes could also be considered. Here, based on the results of a preliminary exploratory analysis, which did not indicate heterogeneity in gender effects across schools, we decided to include only a random intercept in the model.

[5] We specify that, in models for responses in $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$, we do not consider the nested structure of students within schools because the involved student characteristics might be the result of an over-time process, during which students attended different schools.

For each model involving response variables in the sets $Y, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$, we adhere to the following procedure:

- the variable `gender` (from X_1) is included as a single covariate and as an interaction term with all other predictors;
- a backward stepwise procedure is implemented to identify the significant variables and interaction terms;
- following the stepwise selection, the final model is constructed with a specific distinction between the coefficients (β) for males and females for covariates where the interaction with gender is found to be significant. If, for a given covariate, the coefficient β is significant for females but not for males (or vice-versa), this indicates the presence of a CSI.

4 Results

Results of the four final sets of models built by following the structure in Figure 4 are reported in Table 4 and represented in the graphs in Figure 5. Table 4 reports the coefficients relative to CSIs (identified by CSI_label equal to F or M), as well as the coefficients that exert both significant but different effects on the response across genders (identified by CSI_label equal to A). Figure 5 reports a graph for each set of models. Each graph shows a directed edge when a covariate is predictive of the response, independently of gender; an undirected edge when the two variables are mutually predictive, again independently of gender; a bold edge (either directed or undirected) marked by 'F' or 'M' when we observe a CSI[6].

[6]Edges marked by 'F' indicates that the covariate is predictive for males, but not for females and edges marked by 'M' indicates that the covariate is predictive for females, but not for males.

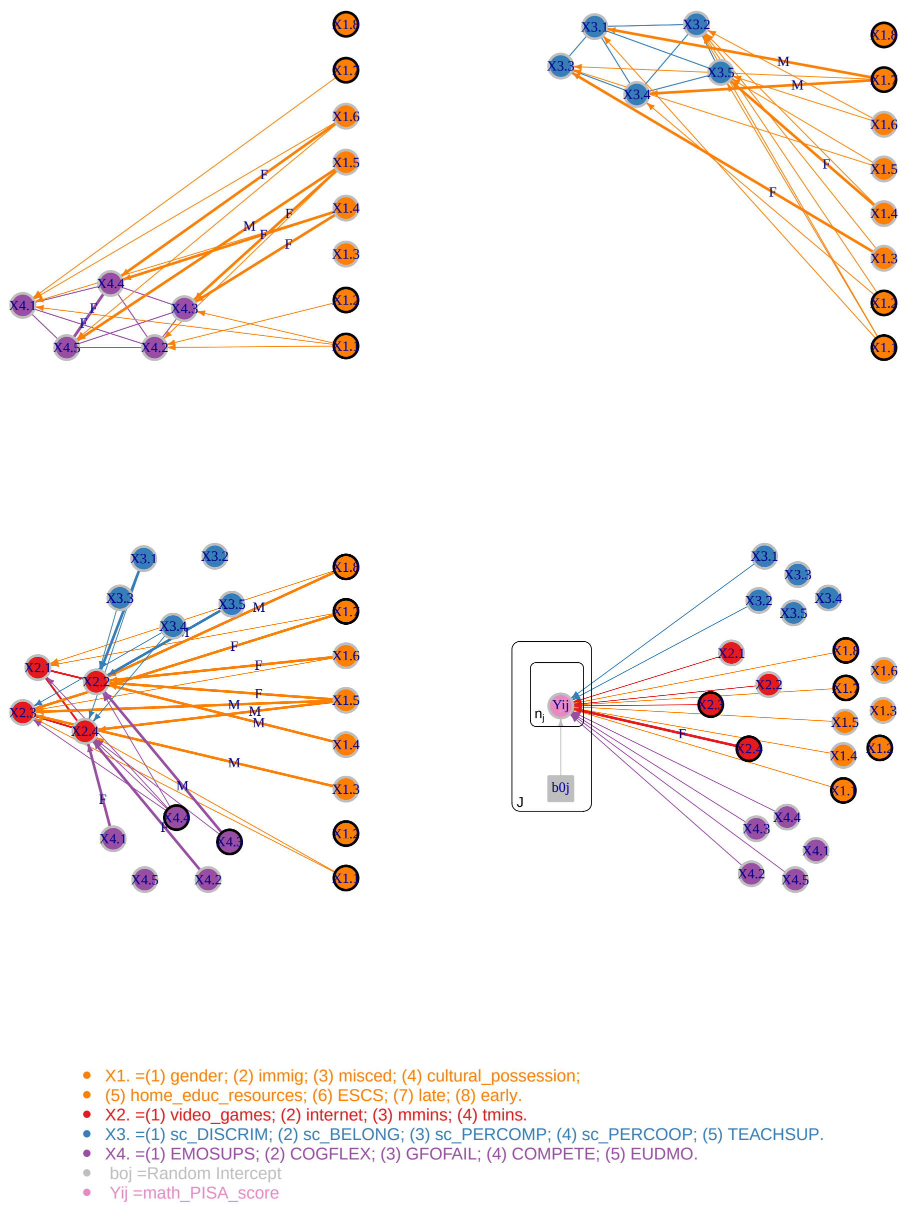

Figure 5: SG corresponding to the four block regressions. **Top left**: target variables belong to $\boldsymbol{X}_4$ and predictors belong to $\boldsymbol{X}_1$. **Top right**: target variables belong to $\boldsymbol{X}_3$ and predictors belong to $\boldsymbol{X}_1$. **Bottom left**: target variables belong to $\boldsymbol{X}_2$ and predictors belong to $\boldsymbol{X}_1$, $\boldsymbol{X}_3$, and $\boldsymbol{X}_4$. **Bottom right**: target variable is Y and predictors belong to $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, and $\boldsymbol{X}_4$.

Contest-specific independencies

Given our focus on the CSIs related to gender, we begin by examining the coefficients that are significant exclusively for one of the two genders. It is noteworthy that most CSIs do not directly appear in the statistical model that predicts student achievement as measured by PISA scores. Instead, these influences are more prominent in the models describing the underlying dynamics that precede and shape this final outcome. These intermediary dynamics reveal intricate patterns of gender-based differences in how certain factors influence students' habits, behaviours, and perceptions.

In particular, the majority of gender-based disparities emerge in the variables that define students' study habits (variables in $\mathbf{X}_2$). For instance, boys' study time in mathematics (`mmins`) is associated with cultural possession, access to home educational resources, and experiences of perceived discrimination at school. These factors, however, do not significantly influence girls' study time. Conversely, girls' study time in mathematics is influenced by a different set of factors: the socio-economic index (SES), early school enrollment, fear of failure, and the perceived support of their teachers (these variables do not significantly affect boys in this context). The use of video games (`video_games`) also exhibits gender-specific predictors. For boys, being a late-enrolled student reduces the likelihood of engaging in video games. In contrast, this factor does not appear significant for girls. On the other hand, for girls, cultural possession and the level of their mother's education play pivotal roles. Cultural possession has a positive association with video game usage among girls, while the mother's education level shows a negative relationship. These associations are not observed in boys. Gender differences are also evident in the determinants of internet usage (`internet`). For girls, internet use is positively associated with cultural possession, highlighting the role of enriched cultural resources in shaping digital habits. For boys, however, such association is not observed. Instead, boys' internet use is positively influenced by cognitive flexibility and their perception of emotional support, suggesting a more psychologically driven dynamic.

In terms of students' perception of the environment, some gender-based differences are observed. Late-enrolled girls, for instance, tend to report higher levels of perceived discrimination (`sc_DISCRIM`) and a lower sense of cooperation (`sc_PERCOOP`) within their schools. Boys, on the other hand, experience an increase in perceived school competitiveness (`sc_PERCOMP`) as their mother's education level rises. Additionally, boys' perception of teacher support (`TEACHSUP`) decreases as their socio-economic index increases, indicating a complex interplay between socio-economic status and the perceived quality of educational relationships.

Significant gender differences are also evident in the associations that involve students' psychological characteristics. For boys, fear of failure (`GFOFAIL`) is influenced by cultural possession and socio-economic status, but these factors are not significant predictors for girls. Competitiveness (`COMPETE`) among boys is positively associated with cultural possession and *eudaemonia* (a measure of well-being), while it decreases with socio-economic status. For girls, an intriguing pattern emerges where *eudaemonia* (`EUDMO`) decreases as cultural possession increases, suggesting that these resources may shape girls' well-being differently than boys'.

Finally, when examining the direct association of CSIs with PISA achievement (`math_PISA_score`), gender-specific patterns persist. For boys, internet usage positively correlates with higher PISA scores. However, this association is not observed for girls, suggesting that the way digital habits influence academic performance varies across genders.

Dependent Variable	Covariate	$\hat{\beta}_M$	se($\hat{\beta}_M$)	$pvalue_M$	$\hat{\beta}_F$	se($\hat{\beta}_F$)	$pvalue_F$	CSI_lable
tmins	mmins	0.3789	0.1113	7e-04	0.6746	0.0973	0	A
mmins	ESCS	-0.0117	3.4606	0.9973	14.0928	2.7278	0	M
mmins	cultural_possession	-9.9636	3.0313	0.001	-2.2286	2.4911	0.371	F
mmins	home_educ_resources	5.0078	2.2508	0.0262	0.2286	1.7926	0.8985	F
mmins	early1	2.6638	6.9351	0.7009	-24.5863	5.2937	0	M
mmins	GFOFAIL	-1.9992	1.979	0.3125	4.4128	1.5811	0.0053	M
mmins	sc_DISCRIM	-4.8573	1.8977	0.0105	-1.1157	1.7427	0.5221	F
mmins	TEACHSUP	0.2629	1.8897	0.8894	5.8007	1.4931	1e-04	M
video_games	misced	0.0453	0.0475	0.3402	-0.0649	0.032	0.0422	M
video_games	cultural_possession	-0.0538	0.1004	0.5919	0.1488	0.0709	0.0359	M
video_games	late1	-0.4478	0.2202	0.042	0.1972	0.1896	0.2983	F
video_games	internet1	0.8062	0.1586	0	0.4597	0.1019	0	A
internet	cultural_possession	-0.0614	0.1056	0.561	0.2085	0.078	0.0075	M
internet	video_games1	0.8114	0.1588	0	0.465	0.102	0	A
internet	EMOSUPS	0.1617	0.0783	0.0388	0.0079	0.0537	0.8835	F
internet	COGFLEX	0.2602	0.0913	0.0044	-0.0426	0.0635	0.5026	F
sc_DISCRIM	late1	-0.0108	0.0833	0.897	0.3597	0.0776	0	M
sc_DISCRIM	sc_PERCOMP	0.1724	0.0251	0	0.0878	0.019	0	A
sc_BELONG	ESCS	0.1588	0.0329	0	0.0799	0.0282	0.0046	A
sc_BELONG	sc_PERCOOP	0.1917	0.0244	0	0.269	0.0192	0	A
sc_PERCOMP	misced	0.0493	0.0183	0.0072	0.0135	0.0147	0.3588	F
sc_PERCOMP	sc_PERCOOP	0.1124	0.0277	0	-0.0641	0.0217	0.0031	A
sc_PERCOOP	late1	-7e-04	0.0919	0.9943	-0.2	0.0833	0.0164	M
sc_PERCOOP	sc_BELONG	0.2136	0.0281	0	0.3143	0.0225	0	A
sc_PERCOOP	sc_PERCOMP	0.1102	0.0265	0	-0.0489	0.0199	0.0142	A
TEACHSUP	ESCS	-0.0984	0.033	0.0029	-0.0359	0.0261	0.1689	F
TEACHSUP	sc_DISCRIM	-0.0802	0.027	0.0029	-0.1711	0.0247	0	A
COGFLEX	cultural_possession	0.1526	0.0319	0	0.228	0.0262	0	A
COGFLEX	GFOFAIL	-0.1506	0.0236	0	-0.0712	0.0193	2e-04	A
COGFLEX	EUDMO	0.1211	0.0234	0	0.1863	0.0191	0	A
GFOFAIL	ESCS	-0.0722	0.0328	0.028	-0.0082	0.0258	0.7501	F
GFOFAIL	cultural_possession	0.1426	0.0398	3e-04	0.0231	0.0336	0.491	F
GFOFAIL	COGFLEX	-0.1845	0.0297	0	-0.0902	0.0242	2e-04	A
COMPETE	ESCS	-0.1257	0.0437	0.0041	0.032	0.0355	0.367	F
COMPETE	home_educ_resources	0.0995	0.0316	0.0017	0.0284	0.0255	.2657	F
COMPETE	EUDMO	0.1877	0.0274	0	0.0383	0.0227	0.0914	F
EUDMO	cultural_possession	0.034	0.0354	0.337	-0.0642	0.0296	0.0299	M
EUDMO	COGFLEX	0.153	0.0294	0	0.226	0.0236	0	A
EUDMO	COMPETE	0.1675	0.0252	0	0.0353	0.0204	0.0845	F
math_PISA_score	internet	0.2016	0.0501	0.0001	0.0377	0.0349	0.2796	F

Table 3: Results of the CSIs identified in the four sets of models. For each CSI, we report the coefficient β with its standard error and pvalue, for both boys and girls. Column CSI_label indicate the gender for which the two variables of interest result to be conditionally independent.

Further gender-specific differences

Apart from the CSIs, other coefficients are also significant in describing education dynamics for both genders, albeit with differing effects. These are labeled with CSI_label = A in Table 4. Evidences show that the association between the time spent to study mathematics and the time spent to study in general is nearly twice as strong in females compared to males. The association between the use of video games and the internet usage is much stronger in males than in females. Males who perceive higher levels of competition perceive also higher levels of discrimination in the school, with an effect that is greater than in females. The positive effect of socio-economic status on the sense of belonging to school is nearly twice as strong in males compared to females. The positive association between the perceived cooperation and the sense of belonging to school is significantly higher in females than in males. The only variable that exerts a significant opposite effect between boys and girls is the perceived cooperation when used to predict the perception of competition. In particular, the perceived level of cooperation on average increases the perception of competition in males but reduces it in females. The positive effect of the sense of belonging on the perception of cooperation is stronger in females than in males. The negative effect of teacher support on perceived discrimination is more pronounced in females than in males. For females, the effect of cultural possession on cognitive flexibility is on average stronger than for males. For both genders, the association between cognitive flexibility and fear of failure is negative and significant, but it is more pronounced in males. This suggests that the cognitive flexibility has a stronger impact on decreasing the fear of failure in males than in females. The positive effect of *eudaemonia* on the cognitive flexibility is stronger in females than in males. Lastly, the positive effect of the cognitive flexibility on *eudaemonia* is significantly higher in females than in males.

Again, most of the gender-based differences are observed in the models describing the underlying dynamics that precede and shape the final student achievement. No gender-based differences emerge in the association between student-level characteristics and math PISA score. However, also school-related factors may encompass unobserved gender-based disparities. Indeed, the percentage of variability explained (PVRE) in Model 4 is 26.18%, indicating significant heterogeneity in the determinants of student achievement across schools. This heterogeneity could potentially obscure additional gender differences that may manifest in specific schools or geographical regions and that could be further explored.

5 Conclusion

This study introduces a methodological framework for examining gender disparities in education, leveraging stratified chain graphical models and multilevel statistical analyses. By focusing on the interplay between contextual, sociocultural, psychological factors and gender, the research highlights the nuanced ways in which factors such as socio-economic status, cultural possession, school environment, and psychological traits shape students' educational outcomes differently for boys and girls.

Key findings reveal that while some determinants of educational achievement are shared across genders, significant differences persist in how these factors interact to determine students' habits and perceptions. For example, boys' study habits are more influenced by resources such as cultural possessions and experiences of discrimination, whereas girls' study time correlates strongly with socio-economic status, teacher support, and fear of failure. Similarly, gendered patterns emerge in the use of technology, where boys' internet use aligns with cognitive flexibility and emotional support, while girls' usage is tied to cultural resources.

These insights underscore the importance of gender-specific approaches in education policy and practice. By identifying CSIs unique to each gender, we can design targeted interventions aimed at addressing disparities and fostering equitable opportunities. For instance, programs that enhance teacher support or alleviate fear of failure may be particularly beneficial for girls, while initiatives that improve boys' sense of belonging and address perceived discrimination could better support their educational journey.

The innovative use of stratified chain graphical models in this work enhances our ability to model interdependencies across variables and contexts, offering a comprehensive understanding of the dynamics at play. This methodological contribution not only provides a clearer visualization of complex dependencies but also sets a precedent for future research on educational inequality. Further research could explore the applicability of these methods in diverse educational settings, e.g. investigating the dynamics within vocational schools, or in more complex multilevel contexts, e.g. investigating the dynamics related to gender across schools or geographical areas, broadening the scope of their impact.

In conclusion, this study contributes valuable insights into the mechanisms underpinning gender disparities in education and provides actionable guidance for promoting gender equity.

Acknowledgement: The present research is part of the activities of "Dipartimento di Eccellenza 2023–2027", MUR, Italy, Dipartimento di Matematica, Politecnico di Milano. Chiara Masci acknowledges support from the Italian Ministry of University and Research (MUR) under the Department of Excellence 2023-2027 grant agreement "Centre of Excellence in Economics and Data Science" (CEEDS).

Data Availability Statement

The PISA2018 raw data are available form the OECD site `https://www.oecd.org/en/data/datasets/pisa-2018-database.html`. From the original dataset, only the country Italy and the school type general (`iscedo=gen`) were considered.

Declarations

Conflict of interest/Competing interests: None

References

[1] G. Bertocchi and M. Bozzano, *Gender gaps in education*, Springer, 2020.

[2] A. Bertoletti et al., "The determinants of mathematics achievement: a gender perspective using multilevel random forest," *Economies*, vol. 11, no. 2, p. 32, 2023.

[3] C. Boutilier et al., "Context-specific independence in Bayesian networks," *CoRR abs/1302.3562*, 2013.

[4] D. Checchi and L. Flabbi, "Intergenerational mobility and schooling decisions in Germany and Italy: The impact of secondary school tracks," IZA discussion paper, 2007.

[5] D. Contini, M. L. Di Tommaso, and S. Mendolia, "The gender gap in mathematics achievement: Evidence from Italian data," *Economics of Education Review*, vol. 58, pp. 32–42, 2017.

[6] D. R. Cox and N. Wermuth, *Multivariate dependencies: Models, analysis and interpretation*, Chapman and Hall/CRC, 2014.

[7] A. G. de San Román and S. De La Rica, "Gender gaps in PISA test scores: The impact of social norms and the mother's transmission of role attitudes," *Estudios de Economía Aplicada*, vol. 34, no. 1, pp. 79–108, 2016.

[8] M. Drton, "Discrete chain graph models," 2009.

[9] L. Dunnzlaff et al., "Chapter 5 Equality of Opportunity and Redistribution in Europe," in *Inequality of Opportunity: Theory and Measurement*, Emerald Group Publishing Limited, pp. 99–129, 2011.

[10] D. K. Evans, M. Akmal, and P. Jakiela, "Gender gaps in education: The long view," *IZA Journal of Development and Migration*, vol. 12, no. 1, 2020.

[11] M. Frydenberg, "The chain graph Markov property," *Scandinavian journal of statistics*, pp. 333–353, 1990.

[12] A. Gottard and C. Rampichini, "Chain graphs for multilevel models," *Statistics & Probability Letters*, vol. 77, no. 3, pp. 312–318, 2007.

[13] A. Gottard, L. Grilli, and C. Rampichini, "A chain graph multilevel model for the analysis of graduates' employment," in *Effectiveness of University Education in Italy: Employability, Competences, Human Capital*, Physica-Verlag, Heidelberg, pp. 169–182, 2007.

[14] M. I. Honig and L. R. Rainey, "Supporting principal supervisors: what really matters?," *Journal of Educational Administration*, vol. 57, no. 5, pp. 445–462, 2019.

[15] S. Højsgaard, "Statistical inference in context specific interaction models for contingency tables. *Scandinavian journal of statistics*, 31(1), 143–158, 2004.

[16] L. La Rocca, and A. Roverato, . Discrete Graphical Models and Their Parameterization. *In Handbook of Graphical Models* pp. 191–216. CRC Press, 2018.

[17] S. L. Lauritzen and N. Wermuth, "Graphical models for associations between variables, some of which are qualitative and some quantitative," *The Annals of Statistics*, pp. 31–57, 1989.

[18] S. L. Lauritzen, "Causal inference from graphical models," *Monographs on Statistics and Applied Probability*, vol. 87, pp. 63–108, 2001.

[19] M. Leonelli, and G. Varando. Context-specific causal discovery for categorical data using staged trees. *In International conference on artificial intelligence and statistics*, pp. 8871-. 8888,2023.

[20] C. Masci, G. Johnes, and T. Agasisti, "Student and school performance across countries: A machine learning approach," *European Journal of Operational Research*, vol. 269, no. 3, pp. 1072–1085, 2018.

[21] F. Nicolussi and M. Cazzaro, "Context-specific independencies in hierarchical multinomial marginal models," *Statistical Methods & Applications*, vol. 29, no. 4, pp. 767–786, 2020.

[22] F. Nicolussi and M. Cazzaro, "Context-specific independencies in stratified chain regression graphical models," 2021.

[23] H. Nyman et al., "Stratified graphical models-context-specific independence in graphical models, *Bayesian Analalysis*, 9(4): 883–908, 2014.

[24] OECD-PISA: www.oecd.org/en/about/programmes/pisa.html

[25] J. C. Pinheiro and D. M. Bates, "Linear mixed-effects models: Basic concepts and examples," in *Mixed-effects models in S and S-Plus*, Springer, pp. 3–56, 2000.

[26] Schiebinger, L., and Klinge, I. (2020). "Gendered innovations 2: how inclusive analysis contributes to research and innovation". Luxembourg: Publications Office of the European Union.

[27] J. Sikora and D. G. W. Pitt, "Does advanced mathematics help students enter university more than basic mathematics? Gender and returns to year 12 mathematics in Australia," *Mathematics Education Research Journal*, vol. 31, pp. 197–218, 2019.

[28] N. Wermuth and D. R. Cox, "Joint response graphs and separation induced by triangular systems," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 66, no. 3, pp. 687–717, 2004.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

13/2025 Scimone, R.; Menafoglio, A.; Secchi, P.
Estimating Non-Stationarity in Spatial Processes: an approach based on Random Domain Decomposition

12/2025 Alessandro Andrea Zecchi, Claudio Sanavio, Simona Perotto e Sauro Succi
Improved amplitude amplification strategies for the quantum simulation of classical transport problems

11/2025 Tonini, A.; Dede', L.
Enhanched uncertainty quantification variational autoencoders for the solution of Bayesian inverse problems

10/2025 Botti, M.; Mascotto, L.; Vacca, G.; Visinoni, M.
Stability and interpolation estimates of Hellinger-Reissner virtual element spaces

09/2025 Quarteroni, A.; Gervasio, P.; Regazzoni, F.
Combining physics-based and data-driven models: advancing the frontiers of research with Scientific Machine Learning

08/2025 Botti, M.; Fumagalli, I.; Mazzieri, I.
Polytopal discontinuous Galerkin methods for low-frequency poroelasticity coupled to unsteady Stokes flow

07/2025 Patanè, G.; Nicolussi, F.; Krauth, A.; Gauglitz, G.; Colosimo, B. M.; Dede', L.; Menafoglio, A.
Functional-Ordinal Canonical Correlation Analysis With Application to Data from Optical Sensors

06/2025 Torzoni, M.; Manzoni, A.; Mariani, A.
Enhancing Bayesian model updating in structural health monitoring via learnable mappings

04/2025 Andrini, D.; Magri, M.; Ciarletta, P.
Nonlinear morphoelastic theory of biological shallow shells with initial stress

05/2025 Buchwald, S.; Ciaramella, G.; Verani, M.
Greedy reconstruction algorithms for function approximation