Inequality in Learning Outcomes: Unveiling Educational Deprivation Through Complex Network Analysis

Harvey Sánchez-Restrepo [1[0000-0001-7311-8622]] and Jorge Louçã [2[0000-0003-4766-5627]]

[1] University of Lisbon, Cidade Universitária, 1649-004, Lisboa, Portugal
Harvey_Restrepo@iscte-iul.pt
[2] Information Sciences, Technologies and Architecture Research Center. ISCTE-IUL,1649-026, Lisbon, Portugal
jorge.l@iscte-iul.pt

Abstract. Understanding which factors are determinant to guarantee the human right to education entails the study of a large number of non-linear relationships among multiple agents and their impact on the properties of the entire system. Complex network analysis of large-scale assessment results provides a set of unique advantages over classical tools for facing the challenge of measuring inequality gaps in learning outcomes and recognizing those factors associated with educational deprivation, combining the richness of qualitative analysis with quantitative inferences.

This study establishes two milestones in educational research using a census high-quality data from a Latin American country. The first one is to provide a direct method to recognize the structure of inequality and the relationship between social determinants as ethnicity, socioeconomic status of students, rurality of the area and type of school funding and educational deprivation. The second one focus in unveil and hierarchize educational and non-educational factors associated with the conditional distribution of learning outcomes. This contribution provides new tools to current theoretical framework for discovering non-trivial relationships in educational phenomena, helping policymakers to address the challenge of ensuring inclusive and equitable education for those historically marginalized population groups.

Keywords: Educational network, Large-scale assessments, Policy informatics.

1 Introduction

With the establishment of the Sustainable Development Goals (SDGs), the 193 countries attached to Unesco promulgated that 'education is a human right' [1] and that the two pillars of quality in education should be learning and equity [1,2], recognizing that all human beings have the right to learn and that the State is obliged to guarantee to all citizens equally [3]. Education is also a source for social mobility: an additional year of quality education can increase a person's income up to 10% [4]. The lack of quality in education worldwide is of such magnitude that Unicef estimates that 250 million children do not have the minimum learning and that more than half of them have been fooled systemically: despite having managed to attend school, many of them fail in developing the minimum learning such as reading, writing or performing basic operations. The causes for this deprivation of learning are multiple but they affect mainly to those belonging to historically marginalized population groups, almost always, the most impoverished [5,6,7]. Further, the modest improvements in learning achievements are usually by the hand of enormous inequalities among students, which has cast doubt on the government actions for improving the quality of education [8].

As a strategy to strengthening public policies, many countries have implemented national Large-Scale Assessments (LSA) to collect valid and reliable data on educational outcomes and Factors (variables) associated with Learning (FAL). The objective of LSA is to have evidence-based information and use quantitative methods to estimate educational changes when varying each factor [9] in many scales of time and across the territory [10], as well to modelling dynamic behavior and asymmetries at school and student level [11]. However, dominant models to study educational phenomena are based on tools that postulate that educational gaps can be explained just through the covariation between learning outcomes and each factor [12]. In this kind of studies, the linear analysis provokes a fragmented view of the system and dismiss very often the interactions between agents, FAL and educational phenomena [13,14], dismissing the structure of the inequality in learning outcomes and its relationship with educational deprivation, which partially explains the lack in explanatory power of those models [15].

To represent the multiple interactions in a system and the emergence of collective properties at different scales from their constituents, several researchers have proposed the use of network theory to model social systems as a result of self-organized processes [17,18]. Given that 'networks are at the heart of complex systems' [16], analyzing statistical and topological properties of the education system through network theory means studying the complexity of the system through its interactions.

Therefore, this research addresses empirical data for estimating educational deprivation in a Latin American country, as well as its relationship with the most relevant social determinants such as Socioeconomic status (SES) of the student and their families, Rurality in the area where the school is located (RA), the Type of school (TS) and Ethnic self-identification of the student (ET), for estimating topological features and order parameters of the network related to out-of-equilibrium states, providing a new kind of information about global and local properties of the structure of educational deprivation and helping to find those key factors driving inequality gaps in learning outcomes.

1.1 Dataset

In this model, a multivariate dataset integrates learning outcomes of every student who has completed the k-12 education process, estimated by the ability's parameter θ^j through a LSA carried out in Ecuador in 2017 using a standardized computer-based test[1] and integrated with a robust dataset with more than 140 variables coming from surveys to student's families and teachers. For building the scores, psychometric parameters were estimated by Item Response Theory through a 2P-Logistic model [19] following equation 1:

$$P(\theta_j) = \frac{e^{[\alpha_i(\theta_j - \beta_i)]}}{1 + e^{[\alpha_i(\theta_j - \beta_i)]}} \tag{1}$$

Con $\theta_j \in (-\infty, \infty), \alpha_i \in (-\infty, \infty)\ y\ \beta_j \in (-\infty, \infty)$.

Raw scores of θ^j were re-scaled to a Learning index ($LI_j \in [4.0,10.0]$), a monotonous transformation, where higher levels of learning are more likely to have higher scores [19]. For measuring relative deprivation, this model uses the sociological proposal that 'needs, thresholds and satisfactions are determined by each society', while absolute deprivation proposes that 'there is an irreducible nucleus of needs that are common to every human being' [20]. All students are classified in levels of achievement (L_k) based on a standard Bookmark process for establishing psychometrical cut points s_i [19]. Students suffering learning deprivation are those that did not meet the minimum learning standards at the end of the compulsory education, denoted by L_0.

1.2 Deprivation learning index

To estimate this index, we use the family of scores $\{LI_j\}\ j = 1, \ldots, N,$ of those students with low level of achievement L_0-class, s_1 is the first cut point —the minimum score to be located at level L_1—. For the L_0-class, absolute deprivation is given by H=$(n(L_0)/\sum n(LI_j)$, where $n(LI_j)$ represents the number of students below the first level of achievement, the intensity $\lambda(LI_j)$ is given by the distance to reach the first level L_1 and the Deprivation learning index (DLI) is given by δ_j=H$\cdot\lambda(LI_j)$, which represents a measure of the collective learning deficit, which considers the magnitude —the number of students with low performance— and intensity —how much below the minimum performance level are located [20].

1.3 Model specification

For analyzing topological properties and pointing out those nodes holding the system out of statistical equilibrium, a three sequential steps model was developed. The first step is focused on disaggregating L_0-class by *SES*, each student is represented by a node and edges, weighted by $\lambda(LI_j)$, are directed to one of the SES-decile nodes

---

[1] Full dataset is available in http://www.evaluacion.gob.ec/evaluaciones/descarga-de-datos/

$\{LI_j(\theta^j \rightarrow L_k^j \rightarrow (SES_d^j)\} \, \forall \, j$, a process which allows to analyze aggregated inequality at school level, as well as In-degree distribution for SES nodes. The second one is an extension for including RA, TS and ET to analyze their effects through the sequence $\{\theta^j \rightarrow L_k^j \rightarrow (SES_d^j) \rightarrow (RA_{C1}^j, TS_{C2}^j, ET_{C3}^j)\} \, \forall \, j$, where C denotes an index for each subcategory of the factors RA, TS and ET. Finally, the third step amplifies and strengthens the analysis through more than one hundred educational and non-educational factors associated with learning achievements trough the sequence $\{\theta^j \rightarrow L_k^j \rightarrow (SES_d^j) \rightarrow (RA_{C1}^j, TS_{C2}^j, ET_{C3}^j) \rightarrow FAL_{Cm}^j\} \, \forall \, j$, for integrating m different educational and non-educational factors. Network analysis was carried out by Gephi 0.9.2 and statistical estimations and plots with R 3.5.0 and Orange 3.3.8.

2 Socioeconomic status and learning deprivation

The first specification estimates the Weighted In-degree distribution of directed edges from social determinants nodes to those representing socioeconomic deciles, given by $\{SES_d\} \; d \in \overline{1,10}$, where each edge represents one student in L_0-class. As inequality implies asymmetries, in conditions of total equity —where socioeconomic factors would not produce differences — we might expect equal distribution of L_0-edges over the network. Therefore, the study of equity can be deepened by analyzing the levels of absolute and relative deprivation experienced by different population groups and their relationship with the SES of the students.

According with LSA estimates, 8 438 of 39 219 students are in L_0-class, a prevalence rate of 0.215, a $LI = 6.32$ and intensity of deprivation λ=0.225, i.e., in average, L_0-student lacks 0.68 standard deviations (SD) of the minimum learning. As shown in Figure 1, distribution of In-degree $P(SES_d)$ is non-uniform and it decreases monotonically as the SES of the group increases.

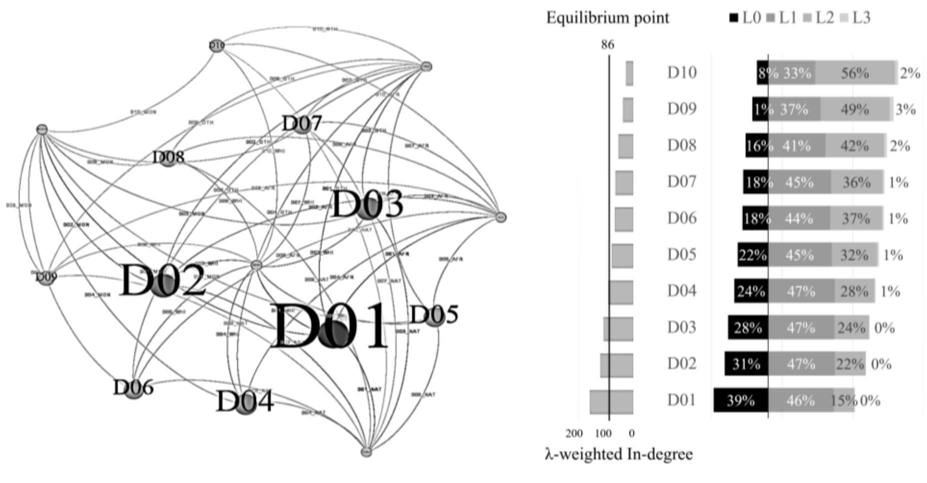

Fig. 1. Network of socioeconomic distribution with λ-weighted learning deprivation and distribution of students among levels of achievement.

For example, $P(SES_1)$=0.39 and $P(SES_{10})$=0.08, this difference of 0.31 means that for each richest-family student who does not learn the minimum, there are 5 poorest-family students in the same situation. As we will see later, this unfortunate situation deepens in rural areas, where the ratio increases to one richest-student for every 7 poorest-students.

Differential centrality of deciles in Figure 1 also shows a non-equilibrium system driven by SES where poorest students dominate the graph: nodes SES_1, SES_2 and SES_3 have stronger connections and the highest prevalence rate (*DPR*=0.3860), Hub parameter (*H*=0.4887), Weighted In-degree (*WID*=158.9420) and PageRank (*PR*=0.4289). On the contrary, the richest students grouped by SES_{10} are relative irrelevant for the network with parameters *DPR*=0.0800, *H*=0.1133, *WID*=27.222 and *PR*=0.0289. As can be seen, in-degree parameter provides an estimation for deprivation rates of each SES_d, showing the size of the gaps among them through detecting SES effects in nodes grouping L_0-students by deciles, pointed out by the negative correlation between LI and SES (*R*= -0.58, *p*<0.001).

To estimate more accurately the cumulative effect of SES on learning outcomes, Figure 2 shows LI (left plot) and intensity of deprivation (right plot), as functions of SES in two dynamical ways: 1) starting with the whole population of students and re-estimating LI and λ while excluding poor-students (black circles), and 2) starting with just poor-students and including richer students (white circles). As can be seen, the biggest circles indicate equal Global Average (GA) for LI (7.61) and λ (0.225), however, for case 1) the effect of removing poor students is that GA goes up immediately after removing SES_1-students, reaching a LI=8.21 (0.65 SD away from GA), as well as λ diminishes 14%. On the contrary, when the estimation process starts with just the poorest students, LI=6.52 (-1.09 SD below from GA) and starts going down after including SES_3 students. In summary, these opposite behaviors show a learning gap of 1.82 SD, equivalent to almost two years of formal schooling between richest and poorest students, and pointing out that, even among those deprived students, the most impoverished get the worst part suffering 14% deeper intensity of deprivation.

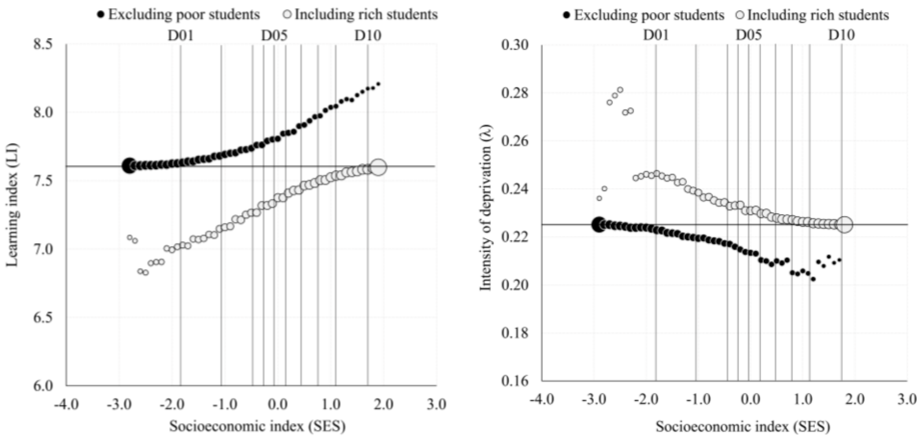

Fig. 2. Cumulative effect of SES in learning outcomes (*left side*) and intensity of deprivation (*right side*).

The previous results provide very clear information about learning inequalities at student level, however, to analyze an aggregated phenomenon, the school level shows local but systemic gaps to better understand the sources of structural deprivation. When studying schools as integrated units, the impact of *SES* becomes even more evident in learning outcomes, in left side of Figure 3, schools are splatted between type of school —sources of funding—, here each school is represented by a circle whose size is proportional to its number of students. The average *SES* of students is located on the horizontal axis and *LI* is on the vertical one. As can be seen, private schools have higher *SES* (0.42) and *LI* (7.96) than the public ones (*SES*= -0.21) and (*LI*=7.44), and there is also a strong positive relationship between *LI* and *SES*.

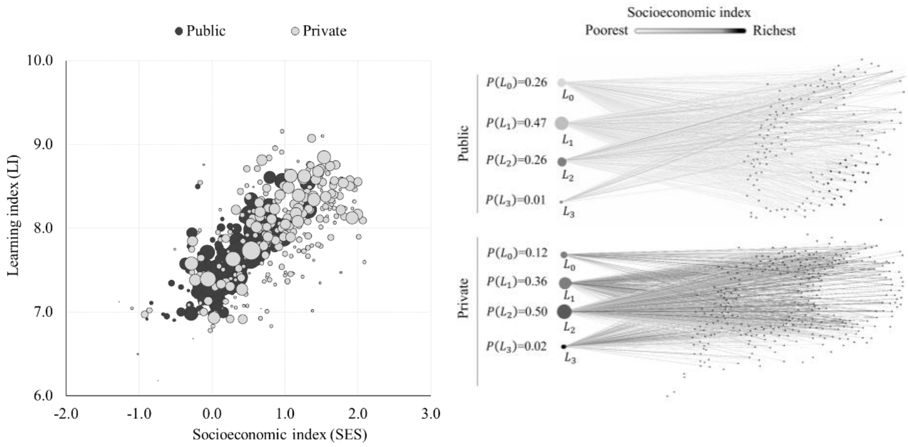

Fig. 3. Relationship between learning and socioeconomic indexes at school level (*left side*), and In-degree distribution of Private and Public networks (*right side*).

Moreover, right side of Figure 3 shows a kind of bipartite data sources [21] for representing the networks of private and public schools, where darker edges refer to higher SES and the size of nodes are masses of probability for in-degree distribution $P(L_i)$ where Private-network has 309 nodes (schools) with 12 823 edges (41 per school) while Public-network has just 167 with 24 671 edges (148 per school), i.e., the private one has 85% more nodes than the public one, but its density is just 52%.

In addition, 70.6% of private schools are linked to L_0 through just 1 553 of their edges (H=0.12 and λ=0.227), while 97.0% of public are linked to the same node through 6 339 edges (H=0.26 and λ=0.223), a rate twice higher than in private sector, showing that SES is a key factor for educational deprivation due mainly to the influence of cultural capital, showing the lack in the capacity of the government to guarantee educational rights.

3 Ethnicity and type of school financing

The rurality of the area where the schools are located is also a factor that impacts on the learning gaps, its effect, combined with the type of school funding, indicates a huge variance among ethnic groups within the socioeconomic deciles. Figure 4 shows a Radviz plot, a non-linear multi-dimensional algorithm [22] and a dendrogram made with a cluster analysis based on $\left(SES_k^i, RA_k^i, TS_k^i, ET_k^i\right)$, this representation shows the hierarchical structure of inequality in learning deprivation for population groups, the points on the circle refer to three indicators attracting the students groups proportionally and the symbol size is proportional to the total deprivation given by δ.

As can be seen, the strong relationship between the SES and the prevalence distributes the groups in a vertical way showing borders between rural (circles) and urban areas (crosses).

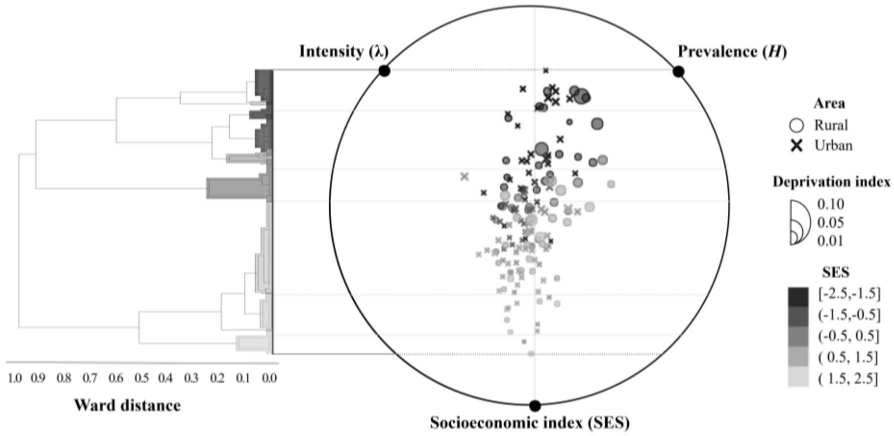

Fig. 4. Deprivation of learning by ethnicity, rurality and socioeconomic status.

Figure 4 shows that gaps between rural and urban areas are quite accentuated, especially among the poorest students in rural areas suffering the highest levels of prevalence and δ. In this sense, among the most deprived students in rural areas, Montubios exhibit the highest prevalence rates, followed by and Indigenous (H=0.64) and Afro-Ecuadorians (H=0.55). The clusters shown by the dendrogram points out that the highest SES students dominates private schools with the lowest levels of deprivation, with the exception of Montubios in rural areas, where no student attending private schools reaches the minimum level and where just 1 of 3 students in public schools is not deprived (H=0.63), showing that social and ethnic classes are splatted into groups of students who have had different learning opportunities inside and outside of the schools, helping to understand how inequality evolves in a structural way [23].

4 Key-factors for public policies

One of the most useful strategies for improving learning and closing gaps is micro-planning, however, to select which needs should be attended first for specific groups is always a great deal. For this reason, once the gaps in population groups are measured, we extend the network for introducing $\{FAL_{Cm}^{j}\}$ to the model and identify which variables are linked with specific population groups and recognizing those that should be intervened first, as well to simulating the variational aspects for establishing the order and control parameters for building more efficient and effective portfolios of policies at local level.

Prevalence rate of L_0-students are related with networks strongly connected and this might be associated with Eigen-Centrality through measuring the influence of each factor for identifying how well connected the *j*-th factor (FAL_j) is and how many links have its connections. In this way, splitting L_0-students in communities becomes in a very valuable tool for developing group-oriented strategies and avoid implementing the same actions for completely different needs. At this point, network simulation offers the advantage of building multiple scenarios across varying *FAL* parameters for recognizing and ranking the most relevant nodes to be attended by policymakers in a hypothetic but very specific situation.

Figure 5 shows a network build through the *ForceAtlas2* algorithm [24] to obtain the layout after a *Modularity process* (with parameter 0.073 at resolution of 0.254) for splitting richest from poorest students to find key factors for educational deprivation in both groups [25]. The Average Weighted Degree of the network is *AWD*=426.9 and its density *D*=5.617. Those large nodes appearing in the network as attractors correspond to two communities: the richest (SES_{10}), with degree of authority $A_{SES_{10}}$=0.158 and just 20 factors —most of them showing very low centrality parameters— and poorest students (SES_1), with A_{SES_1}=0.987 and 57 different factors associated with their deprivation.

For assessing the quality of its connections, PageRank is a well-known algorithm for providing accurate and clear information when looking for the factors dominating deprivation in each community [25,26], which also might order nodes parametrically for selecting those who susceptible to be managed by policy and define 'needs profiles' for focusing actions on those variables susceptible to be managed by policy for specific zones or groups.

For example, a SES_1 profile based on Figure 5 using the 10 most relevant factors might be:

> 'Members of a household who currently receives the Human Development Bond and needs to work for a wage. Their parents have a very low level of education, they do not have a desktop computer neither Internet connection, and also have no books or just a very few. In their school, teachers arrive late to class, are not committed with learning and have low expectations about student's future'.

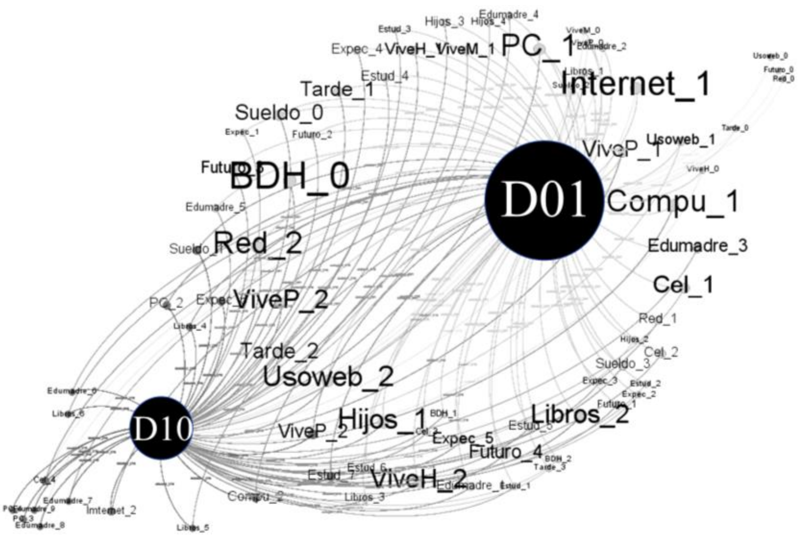

Fig. 5. Factor's network for richest and poorest students' communities. Centrality of nodes in each group points out the dissimilarities in factors provoking deprivation.

As can be seen, this very detailed information is extremely helpful for developing based-evidence policies, to estimate and assign budget and increase the chances of having a successful deployment, the most wanted tool from policymakers.

5 Discussion

This research finds a lot of evidence about the deep lack in equity in a Latin American country and its relationship with social determinants for synthetizing the complex structure of inequality on learning outcomes [27]. Measuring how far is each student of reaching the minimum level to ensure its right to education is crucial because the social deception of accessing education without guarantees of learning, converges at levels of precariousness similar to those found in people outside the scholar system [28]. Likewise, to identify a specific set of factors impacting a group of students that might not be relevant for others is a key issue for developing group-oriented policy [29].

This sequential network model allows to investigate the factors conditioning the exercise of the right to education and opens the possibility of reviewing how the type of school funding promotes inequality in different population groups at country level. Evidence found over deprivation among the socioeconomic groups shows the lack of attention to ethnicity, especially in the urban area. In this sense, when comparing the topologies of the networks of the richest versus the poorest was possible to estimate the gaps in educational outcomes that are associated to other factors like the influence of cultural capital, the context and the families of the students, beyond the schools.

Through this model is also possible to use LSA results for analyzing the interplay of socioeconomic status, rurality, type of school and ethnicity, bringing very useful information about the educational system at meso and macro levels from micro level interactions, offering valuable information for answering five main educational questions:

1. What is the level of deprivation of learning in students at the end of compulsory education?
2. Is learning deprivation associated with schools?
3. How deep is the inequality in the distribution of learning outcomes?
4. What factors are associated with educational deprivation in specific areas of the territory and population groups?
5. Considering limited resources and time, what factors and in what order should they be established in a group-oriented public policy?

Finding answers to these questions using a dataset from a LSA represents a tipping point for integrating thinking tools into current theoretical frameworks to find non-trivial relationships between social conditions and educational deprivation in historically marginalized population groups. In this sense, estimated gaps in deprivation due to rural area and ethnicity, point out the lack of effective inclusion policies, especially of those groups of students who have had different learning opportunities and highlights how inequality is structurally generated in the country.

However, though it is possible to infer the central role played by the schools in determining the more relevant factors for educational deprivation in different SES classes, more research is needed to understand the interactions between the school context and ethnic diversity. Nevertheless, the implications for public policies are: 1) programs aimed at indigenous self-identified students from the poorest quintile should be reviewed in all its pedagogical aspects to guarantee the achievement of meaningful learning; 2) the interculturality programs in the urban area should be broadened so that the ethnic groups can reach the minimum standards at same rates than the other ethnical groups and, 3) it is necessary to better regulate the private schools in the rural area to help close the gap in the absolute deprivation rate.

There is no doubt that the network analysis shows a great vein of scientific development that has not yet been explored to improve knowledge about educational systems, addressing the greatest challenges of educational research, and helping policymakers to develop strategies for an inclusive and equitable education for all.

References

1. OECD (2018). Equity in Education: Breaking Down Barriers to Social Mobility, PISA, OECD Publishing, Paris
2. UN DESA (2018). The Sustainable Development Goals Report 2018, UN, New York
3. Sánchez-Restrepo, H. (2018). Equity: the focal point of educational quality. National Educational Evaluation Policy Gazette in Mexico, Year 4. (10), pp. 42-44.
4. United Nations Children's Fund (Unicef). The investment case for education and equity. DC, Washington: 2015
5. Keeley, B., & Little, C. (2017). The State of the World's Children 2017: Children in a Digital World. UNICEF. 3 United Nations Plaza, New York, NY 10017.
6. Samman, E. (2018). SDG progress: fragility, crisis and leaving no one behind: report.

7. Gray, J., Kruse, S., & Tarter, C. J. (2016). Enabling school structures, collegial trust and academic emphasis: Antecedents of professional learning communities. Educational Management Administration & Leadership, 44(6), 875-891
8. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación – LLECE (2018). Agenda 2018. http://www.unesco.org/new/en/santiago/press-room/newsletters/newsletter-laboratory-for-assesSSent-of-the-quality-of-education-llece/.
9. Motl, T. C., Multon, K. D., & Zhao, F. (2018). Persistence at a tribal university: Factors associated with second year enrollment. Journal of Diversity in Higher Educ, 11(1), 51.
10. Tracey, B., & Florian, K. (Eds.). (2016). Educational Research and Innovation Governing Education in a Complex World. OECD Publishing.
11. Davis, B., & Sumara, D. (2014). Complexity and education: Inquiries into learning, teaching, and research. Routledge.
12. Fox, J. (2015). Applied regression analysis and generalized linear models. Sage Public.
13. Moll, K., Göbel, S. M., Gooch, D., Landerl, K., & Snowling, M. J. (2016). Cognitive risk factors for specific learning disorder: processing speed, temporal processing, and working memory. Journal of learning disabilities, 49(3), 272-281.
14. van der Maas, H., Kan, K., MarSSan, M., & Stevenson, C. (2017). Network models for cognitive development and intelligence. Journal of Intelligence, 5(2), 1–17. doi: 10.3390/jintelligence5020016
15. Gelman, A., & Imbens, G. (2018). Why high-order polynomials should not be used in regression discontinuity designs. Journal of Business & Economic Statistics, 1-10.
16. Barabási, Albert-László, and Márton Pásfai. Network Science. Cambridge: Cambridge University Press, 2016.
17. Johnson, J., Fortune, J., & Bromley, J. (2017). Systems, networks, and policy. In Non-Equilibrium Social Science and Policy (pp. 111-134). Springer, Cham.
18. Grandjean, Martin (2017). "Complex structures and international organizations" [Analisi e visualizzazioni delle reti in storia. L'esempio della cooperazione intellettuale della Società delle Nazioni]. Memoria e Ricerca (2): 371–393.
19. Borsboom, D., & Molenaar, D. (2015). Psychometrics. In J. Wright (Ed.), International encyclopedia of the social & behavioral sciences. Vol.19 pp.418–422.Amsterdam: Elsevier
20. Bracho, T (2017), Índice de déficit en competencias ¿Avanzamos hacia la garantía del derecho a la educación? en Reformas y Políticas Educativas; No. 4, septiembre-diciembre 2017, FCE, México.
21. van Borkulo, C.D., Borsboom, D., Epskamp, S., Blanken, T.F., Boschloo, L., Schoevers, R.A., & Waldorp, L.J. (2014). A new method for constructing networks from binary data. Scientific Reports, 4(5918), 1–10.
22. Leban, G., B. Zupan, et al. (2006). "VizRank: Data Visualization Guided by Machine Learning." Data Mining and Knowledge Discovery 13(2): 119-136.
23. Koller, D., & Friedman, N. (2009). Probabilistic graphical models: principles and techniques. Cambridge, MA, USA: MIT Press.
24. Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2013). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. PLoS One, 9(6), 1-18.
25. Forbush, K., Siew, C., & Vitevitch, M. (2016). Application of network analysis to identify interactive systems of eating disorder psychopathology. Psych Med, 46(12), 2667–2677.
26. Isvoranu, A.M., van Borkulo, C.D., Boyette, L., Wigman, J.T.W., Vinkers, C.H., Borsboom, D., & Group Investigators (2017). A Network Approach to Psychosis: Pathways between Childhood Trauma and Psychotic Symp. Schizophrenia bulletin, 43(1), 187–196.

27. Sánchez-Restrepo, H., & Louçã, J. (2019, September). Topological Properties of Inequality and Deprivation in an Educational System: Unveiling the Key-Drivers Through Complex Network Analysis. In International Conference on Human Systems Engineering and Design: Future Trends and Applications (pp. 469-475). Springer, Cham.
28. Tsatsaroni, A., & Evans, J. (2014). Adult numeracy and the totally pedagogised society: PIAAC and other international surveys in the context of global educational policy on lifelong learning. Educational Studies in Mathematics, 87(2), 167-186.
29. Menefee, T., & Bray, T. M. (2015). Education in the Commonwealth: Quality education for equitable development. Commonwealth Secretariat.