International Journal of Research in Social Sciences
Vol. 13 Issue 01, January 2023,
ISSN: 2249-2496 Impact Factor: 7.081
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Bayesian networks and causal discovery: what lessons for the synthetic indicator of the quality of education systems in OECD countries?

François-Marie Gerard[i*]
Bernard Hugonnier[ii**]
Sacha Varin[iii***]

Abstract

Educational research is largely based on observational studies. The possibility of demonstrating causal relationships in such studies is under debate. However, several methods of causal analysis for such data have been developed over the past twenty years. The present research aims to identify causal relationships between the six criteria defining the summary indicator of the quality of education systems (ISQ) and its final score in 2018. For this purpose, causal Bayesian networks are used and, more specifically, directed acyclic graphs that allow the identification of causalities.

Keywords:

Bayesian networks
Synthetic indicator of the quality of education systems "ISQ"
Directed acyclic graph
Causality

Copyright © 2022 International Journals of Multidisciplinary Research Academy. *All rights reserved.*

Author correspondence:

Sacha Varin
Professor of Mathematics and Statistics, Education Professor
College Villamont, Lausanne, Switzerland
Email: varinsacha@yahoo.fr

1. Introduction

According to Talbot (2012), the simplest and safest way to show causal inferences requires randomized experiments, widely used in medicine for example. Experimental units are then randomly divided into two groups that unknowingly receive either the treatment to be tested or another treatment or placebo (so-called double-blind research). The difference in the effects in the two groups can thus prove the effectiveness of the treatment.

What happens when a randomized study is impossible? Should we abandon the idea of looking for causal inferences altogether? The answer given by Judea Pearl (2000) is an emphatic "no". But this will come at a price: identification is contingent on the modeling assumptions to be discussed in this paper.

The approach proposed by Pearl consists of drawing a directed acyclic graph (DAG). This is a graph where the variables under examination are connected by arrows that indicate the direction and importance of causalities (Pearl, 1995, 2000, 2003, 2009).

Using this method, we seek to highlight causalities within the most recent summary indicator of the quality of education systems in OECD countries (ISQ), which is the one established from the 2018 PISA data. This indicator is composed of six criteria: effectiveness, efficiency, equity, parent engagement, student engagement and teacher engagement. Each of these criteria is defined by a score, with the average of the set constituting the final ISQ score. The final ISQ score is calculated diachronically (following the sequence of

[i*] Scientific Advisor, BIEFOR, Belgium
[ii**] Education professor, Paris Catholic University, Former deputy director of the education directorate of the OECD, France
[iii***] Professor of Mathematics and Statistics, Department of Mathematics, College Villamont, Switzerland

the PISA studies that provide most of the primary data). We do not go into further detail here on the six criteria and the ISQ; interested readers can read about them in our previous writings: Gerard, Hugonnier and Varin, 2017, 2018, forthcoming.

The two main objectives of our research are:

1. on the one hand - thanks to the directed acyclic graph - to identify, measure, understand and interpret all the causal relations among these different criteria;
2. on the other hand, to measure the strength of the six functional relationships between each of the criteria and the final ISQ 2018 score, the latter being just the average of the scores obtained by each criterion.

The results will help determine the confidence that can be placed in the strength and direction of each of the DAG arrows.
In this article, we first present the theoretical framework related to the search for causality; we then present the methodological tools used to highlight the different direct causal effects. Finally, we present the main results and discuss them before drawing conclusions.

2. Theoretical framework

We would first like to clarify that two main tasks can be distinguished in the field of causality: causal discovery and causal inference. In the latter, we know, thanks to experimental data, that such a causal link exists between such variables. The causal relations are therefore defined at the beginning of the analysis. Our work, on the other hand, is clearly in the domain of causal discovery, i.e., from a set of observational data, to try to deduce all the causal relationships. Causal discovery does not assume any a priori relationship between the variables involved. It is the discovery process that allows relationships to be inferred directly from the variables.
In this section, following a review of the literature concerning the modeling and analysis of causal networks since the beginning of the 2000s, we approach the notion of Bayesian networks through a presentation of graph theory and the method, derived from causal Bayesian networks, called directed acyclic graph (DAG). This method allows us to identify causalities in observational data (Maathuis, Kalisch, Bühlmann, 2009). Finally, we briefly describe the notion of causal Bayesian networks.

2.1 Literature review of causal network modeling and analysis

Causal inference and the various theories associated with it have been widely discussed and summarized in numerous works (Pearl, 2000, 2003, 2009; Spirtes et al., 2000; Morgan and Winship, 2014; Imbens and Rubin, 2015; Vanderweele, 2015; Peters and Janzing, 2017; Hernàn and Robins, 2018). Causal network modeling and analysis emerged in the last decades of the 20th century, and their use is currently flourishing.
Probabilistic graphical models, and more specifically Bayesian networks, were originally developed in the 1980s by Judea Pearl, who made a major contribution to the theory of structural models for identifying and estimating causal effects from observational data. He also developed the DAG which is part of the causal Bayesian networks.

2.2 Bayesian networks

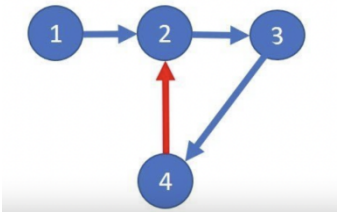
Bayesian networks are probabilistic graphical models that represent random variables and their conditional dependencies as well as probability tables that allow knowledge to be acquired, developed and exploited (Chickering, Geiger and Heckerman, 1995; Chickering and Heckerman, 1996; Naïm, Wuillemin, Leray, Pourret and Becker, 2007).
The networks allow the representation of probabilities and the efficient calculation of probabilities useful for decision making. The particular interest of Bayesian networks is to simultaneously consider the *a priori* knowledge of experts on the subject and the information contained in the data.
Given the great flexibility of Bayesian networks, they have been used in many disciplines: finance, economics, medicine, robotics, civil engineering, geology, genetics, criminology, ecology, industry, etc. (Ben Hassen, Masmoudi and Rebai, 2008; Lauritzen, 1996; Naïm, Pourret and Marcot, 2008).

2.3 Brief presentation of graph theory

Graphs are a way of thinking that allows us to model a wide variety of problems by reducing them to the study of vertices and edges. Vertices are often represented by variables, and edges can be directed. For DAG, an arrow indicates the direction. Alternatively the edges can be undirected, for example in Markov fields (Pearl, 1988).
More precisely, a graph G = (V,E) consists of a set of graphical representations of nodes (V) and the edges (E) that connect them. The nodes represent random variables V = (V1,...,Vp) and the edges are the links between the variables.

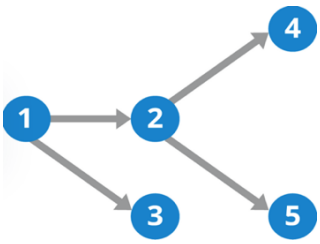
Our study focuses on the directed acyclic graph (DAG). It is distinguished by the presence of unidirectional directed arrows and no loops. Figures 1 and 2 illustrate the difference between cyclic and acyclic directed graphs.

Figure 1: Directed cyclic graph (DCG): presence of a loop between 2, 3 and 4



Source: https://dev.to/jjb/part-16-detecting-graph-cycles-with-depth-first-search-4nh3

Figure 2: Directed acyclic graph (DAG): presence of directed arrows with only one direction and no loop



Source: https://hazelcast.com/glossary/directed-acyclic-graph/

2.4 Causal Bayesian networks

We distinguish two kinds of networks: those where the arrows do not necessarily have to be interpreted in terms of causality; and those where the arrows of a directed graph represent direct cause and effect relationships among all the variables. Causal Bayesian networks are in fact an extension of classical Bayesian networks where any relationship between variables corresponds to a causal relationship (Pearl, 2000). A causal Bayesian network is a Bayesian network, but with an additional main property (Pearl, 2000) that each set of parent → child arrows (Pa(Vi) → Vi) no longer represents just a probabilistic dependence, but a causal relationship. In a causal Bayesian network, each conditional probability table (CPT) represents a stochastic process in which the values of Vi are chosen based on the values of Pa(Vi) but not vice versa.

This property makes causal Bayesian networks graphical probabilistic models whose structure is even more readable by experts in the subject being studied. A causal Bayesian network can thus be used to test causal hypotheses.

3. Research Methodology

In this section, we focus on the very important notions of strength and direction of arrows by addressing the construction of the DAG through the hill climbing algorithm and the bootstrap technique using the R software (R Core Team, 2022) and more specifically the packages bnlearn (Scutari, 2017) and pcalg (Kalisch, Mächler, Colombo, Maathuis, Bühlmann, 2012; Maathuis, Colombo, Kalisch, Bühlmann, 2010). We also discuss the assumptions and conditions required for the DAG to be interpreted in terms of causality.

3.1 Measuring the strength and direction of arrows

Measuring the confidence of arrows in a Bayesian network such as the DAG graph is a major problem in causal inference. Friedman, Goldszmidt and Wyner (1999) have introduced a way to quantify this confidence level. This involves generating several DAG by non-parametric bootstrapping and estimating the frequency of occurrence of arrows; a bootstrap consists of an extensive replication of the data using the resampling technique.

We use the boot.strength() function from the bnlearn package (Scutari, 2017). This function estimates the strength of each arrow based on its empirical frequency in a set of DAG constructed from bootstrap samples. It calculates the probability of each arrow (*modulo* its direction) and the probabilities of the directions of each provided that it is present in the DAG.

In other words, the strength and direction of the relationship are measured and defined by the frequency of appearance of the arrows in the DAG constructed by bootstrap. The objective is to find out in what

proportions the presence and direction of an arrow between two variables appear in the 100,000 DAG constructed by bootstrapping.

3.2 The hill climbing algorithm

Before presenting the algorithm that allowed us to perform these operations, it is important to explain the reason why we chose the hill climbing algorithm rather than another one. Compared to other algorithms, hill climbing is the one that minimizes the BIC score (Bayesian information criterion, see in 3.2.1) during a cross-validation analysis using the hold-out method.

The hill climbing algorithm (Gámez, Mateo, & Puerta, 2011), also called greedy search, tries to maximize a network score reflecting its goodness of fit with the available data. This algorithm focuses on the construction of a DAG in a global way, the score being calculated on the whole structure. Thus this score estimates the quality of a network in its entirety based on the observations. Each score must maximize the probability $P(G|D)$ of the graph G given the observations.

According to Bayes' formula: $P(G|D) = \frac{P(D|G)\times P(G)}{P(D)}$, the probability $P(D|G)$ represents the marginal likelihood of the data given the model, which is the maximand.

A key issue for the hill climbing algorithm is that it may find a local maximum instead of the global maximum. Accordingly the algorithm must be run several times (iterated hill climbing), and the graph with the best score is retained. The score we use for the algorithm is the BIC. To build the DAG, the hill climbing will therefore maximize the marginal likelihood based on the BIC score.

3.2.1 The BIC score

There are several widely-used criteria for goodness-of-fit: the AIC, the BIC, the BD, etc. The BIC, which takes into account *a priori* information, is defined as $BIC = log(P(D|G)) - \frac{d}{2}log(n)$. The first term is the probability of having data D, given the graph G.

The second term enforces parsimony: it penalizes overparameterized models. In the formula "n" represents the number of sample observations, and "d" is the number of parameters associated with the network (i.e., the number of variables in the network). The objective of the hill-climbing algorithm is to maximize the BIC score. We chose the BIC because it can produce valid results with small samples. Specifically, a sample size between 20 and 30 is more than sufficient (Murphy, 2007).

3.3 Construction of the DAG using the bootstrap technique

Given the relatively small size of our sample, the DAG is constructed using the bootstrap technique with 100,000 replications. This number being very high, we obtain a very accurate estimate of the empirical frequencies of the strength/presence and direction of the arrows. The DAG is obtained by keeping only the arrows that are present in at least 70% of the bootstrap DAG. While statistical theory does not provide a specific cut-off criterion, researchers typically confirm causal relationships for proportions above 70%-80% in the case of arrows and above 50% for the arrows' directions. This is a subjective choice. To retain the presence of an arrow, we set a minimum threshold of 70%.

3.4 Conditions to be met

In order to interpret arrows as causalities, several conditions must be satisfied.

The first condition - this is the main assumption made by causal Bayesian networks - is called the causal Markov condition. A DAG must verify this condition. Two variables that are not directly causally related are independent conditionally on their set of common Markovian parents. In other words, each variable is independent of its non-descendants conditionally on its parents. This assumption allows us to distinguish between correlation and causation. The causal Markov condition proposes that all the statistical knowledge necessary for the modeling of the current process is contained in the present. This assumption, which limits the number of variables in the model, can almost always be satisfied in practice. It is indeed sufficient to define "the present" correctly. For example, since we are basing our study on 2018 data, we will say that the present is the year 2018.

The second condition is the assumption of faithfulness between the graph and the probability distribution underlying our "P" data. This is the existence of a Bayesian network that is the P-map of the independence model associated with the probability distribution "P". In other words, a model is faithful if it does not miss any conditional independence.

The third condition is causal sufficiency. Under this condition, all potentially causal variables are included in the analysis. The set of variables in the DAG is sufficient to represent all conditional independence relationships that could be extracted from the data. Can we claim that there are no unmeasured confounding variables? An analysis using the Fast Causal Inference (FCI) algorithm (Spirtes et al., 1999) reveals that we do not have enough information to answer this question. This means that we may or may not have

confounding variables. It would be surprising if there were no unmeasured confounding variables in such a complex topic as the quality of an educational system, which is why the DAG presented here is a potential solution, but there may be another DAG that contain unmeasured confounding variables.
All three conditions must be met to interpret the DAG in terms of causality. It is important to note that these three conditions are not easily tested or verified: the data cannot tell us if these three hypotheses are appropriate. Whether they are satisfactory is a matter of specific knowledge and judgment. Only experienced researchers who are very knowledgeable about the topic and the variables of the DAG can determine whether the three conditions are satisfied. The slogan of N. Cartwright (1994) is very evocative in this respect. “No causes in, no causes out.” If we are to be able to estimate a direct causal effect, we need a general qualitative understanding of the causal structure in which that effect is embedded.
It is also important to discuss another equally important condition before we can interpret the DAG causally: the variables must not be functionally dependent; otherwise, they may create a bias in the DAG. Now our data are functionally dependent since the variable "ISQ 2018 final score" is the average of the other six variables. Therefore we removed the variable "ISQ 2018 final score" from the analysis, after which the analysis identified three arrows. Subsequently we restored the variable "ISQ 2018 final score" to the model as functionally dependent on the other six variables.
In addition, measuring the strength and direction of the functional dependencies is important to our research, as they lead to a better understanding of the confidence one can have in each criterion that makes up our synthetic indicator (ISQ). Because the strengths of the six variables that make up the mean variable "ISQ 2018 final score" are not equal, our data are not highly collinear (VIF < 2).
To distinguish probabilistic from functional dependencies in the DAG, we present solid arrows for probabilistic dependencies and halftone arrows for functional dependencies.
Given the limited size of the dataset used to construct the DAG, we have opted for bagging/bootstrap approaches to obtain more robust results.
Finally it bears repeating that the hill climbing algorithm can get stuck at a local maximum. Indeed Bayesian networks often contain many local optima; accordingly we reran the algorithm 100,000 times. But since a global maximum cannot be guaranteed absolutely, our DAG must be interpreted with due caution; and its use as a tool will have to be verified, if possible, experimentally.

4. Results and discussions

In this section, we present all the results from the DAG (Figure 3), including the functional dependencies (shown in halftone).

4.1 Presence/strength and direction of arrows

As indicated in the methodology, we have kept only those arrows that have a strength/presence of at least 70%, symbolizing this by the thickness of the arrow.

Table 1 shows the two proportions (presence and direction) for the nine arrows of the DAG. This table shows our confidence in the presence of the edges and their direction.

Table 1: Presence and direction of arrows

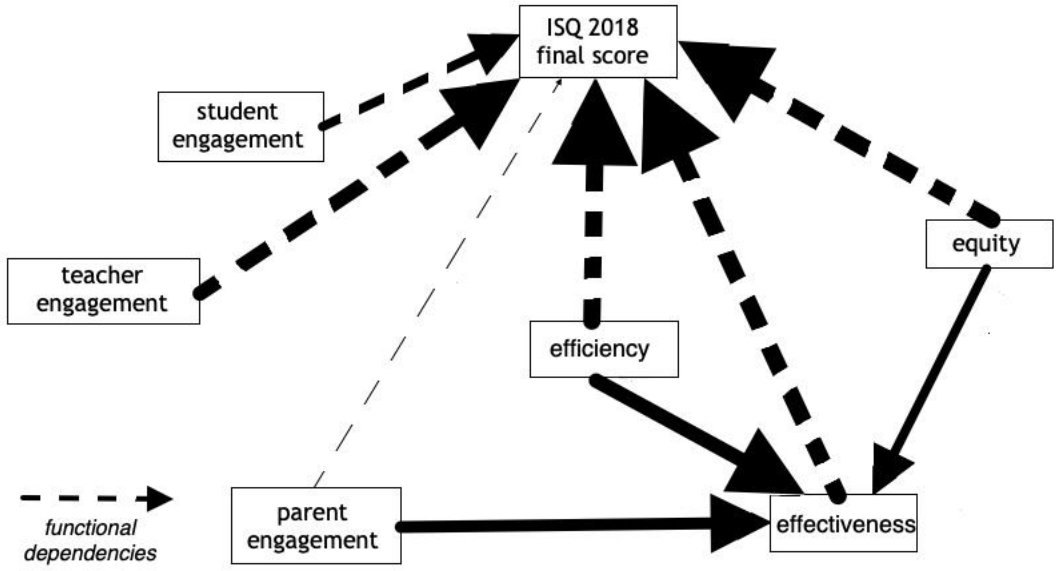
From	To	Presence	Direction
student engagement	ISQ 2018 final score	83%	91%
teacher engagement	ISQ 2018 final score	95%	76%
parent engagement	ISQ 2018 final score	70%	91%
effectiveness	ISQ 2018 final score	98%	51%
efficiency	ISQ 2018 final score	99%	65%
Equity	ISQ 2018 final score	100%	83%
parent engagement	effectiveness	92%	92%
Equity	effectiveness	85%	88%
efficiency	effectiveness	96%	58%
effectiveness	*efficiency*	*96%*	*47%*

The relationships between effectiveness and efficiency are peculiar as shown in the last two rows of the table. The probability of presence is strong and identical: 96%. On the other hand, the probability of direction is 58% in the direction "Efficiency Effectiveness" while it is 47% in the direction "Effectiveness Efficiency". This second direction is a functional dependency: the calculation of efficiency depends on the effectiveness which is related to the level of the means implemented.

4.2 Main results from the DAG

The diagram (Figure 3) and Table 1 show the following results:

Figure 3: Directed Acyclic Graph (DAG), with functional dependencies



- Of all the functional dependencies, equity - defined as the ability of a school system to compensate for the impact of social background on school performance - seems to have the most important role for all countries, although this importance must certainly vary from country to country. This functional relationship is the only one that is 100% present to explain the ISQ 2018 final score. Equity also has a strong causal relationship to explain effectiveness. This could mean that the more equitable an education system is, the more effective it is. In other words, giving special attention to students who have less vocabulary than others (which hinders their understanding of the course), or who have attention or concentration difficulties, or who have not mastered school codes, would, depending on the country, contribute to increasing the effectiveness of education systems and their quality.

- Parent engagement has a special status: its functional dependence on the ISQ 2018 final score is relatively low. As we pointed out in a previous publication (Gerard, Hugonnier and Varin, forthcoming), the introduction of this factor has little effect on the ISQ 2018 final score, i.e. the overall quality level of education systems in OECD countries as a whole. The explanatory hypothesis we put forward was linked to the age of the students represented in the PISA studies: 15 years. At this age, students seem to be less attentive and sensitive to their parents' advice. Some work (Fan and Chen, 2001; Hoover-Dempsey, Battiato, Walker, Reed, DeJong and Jones, 2001; Organisation for Economic Co-operation and Development [OECD], 2019) also suggests that high parental involvement with young people of this age may have a negative effect on their outcomes, in contrast to what seems to be the case with children at the beginning of school learning.

- On the other hand, the DAG indicates an important causal effect of parental engagement on effectiveness. It is even the strongest relationship of the probabilistic dependencies. This could be explained by the fact that, for most parents, what matters most is their children's grades at school (which is partly reflected in the effectiveness criterion), especially since this is almost the only information they have about their children's performance (unless they make an appointment with the principal teacher). This finding can only invite us to continue to consider the "parental engagement" dimension in our future analyses, even if its impact on the ISQ 2018 final score is relatively weak.

- Effectiveness - that is, the ability of a school system to enable students to perform well in school - is a factor that affects the ISQ 2018 final score not only directly, but also because of causal relationships exerted on it by parental engagement, efficiency, and equity as probabilistic dependencies. In this respect, effectiveness is a singular factor.

- It is no coincidence that this factor is often considered the main characteristic of the performance, or even the quality, of an education system: it would make no sense to claim that an education system is of good

quality if it does not achieve its objectives in terms of student learning. Our research on DAG further establishes that the importance of effectiveness is explained by other factors: equity, parental engagement, and efficiency.

- As noted in the presentation of Table 1, the functional dependence between effectiveness and efficiency is singular; it is strongly present in the different computational iterations (96%). In 58% of the iterations, the relationship indicates that it is efficiency that influences effectiveness, while the opposite direction is present in 47%. This finding is surprising, because by construction, efficiency depends on effectiveness (which is therefore a functional dependency): efficiency is in fact the relationship between effectiveness and the means used to achieve it. By definition, with equal means, the greater the effectiveness, the higher the efficiency. However, the results of the DAG favor the opposite causal hypothesis: the higher the efficiency, the higher the effectiveness. This influence of efficiency on effectiveness may occur since when efficiency is low, there is strong political pressure to take action to increase efficiency.

- The fact that all six criteria have a direct and high functional dependence on the ISQ 2018 final score makes sense. Nevertheless, this finding is important because it shows how the use of causal Bayesian networks, and especially the DAG network, can identify causal relationships not discernible by conventional statistical techniques. In another paper we have shown, using Kendall's tau coefficient, that the presence or absence of a single criterion does not significantly modify the ISQ 2018 final score of countries (Gerard, Hugonnier and Varin, forthcoming). Thanks to the DAG, we can qualify this finding: the functional dependencies between each criterion and the level of quality of education systems can be quantified. The more stable (or reliable) functional dependencies concern equity and student engagement, which is a very important result for educational policy.

5. Conclusion

Through this research and the use of the DAG as a methodological tool, we have uncovered causal links that were previously impossible to identify. Awareness of these results would enable political decision-makers to act positively on the quality of education systems. Indeed, equity and, to a lesser extent, parental engagement and efficiency are three effective levers on which the political world could act concretely to try to increase the quality of education systems.

However, as already mentioned, the analysis made here concerns all OECD countries. Thus, establishing that working to increase equity, parental engagement and efficiency can have a significant impact on the quality of education systems in OECD countries does not mean that this is true for each country. In all our work (Gerard, Hugonnier and Varin, 2017, 2018, forthcoming), we have reiterated that what matters is that each country, through the ISQ 2018 final score, can analyze its individual situation and make decisions that are appropriate for it. In this regard, in these other articles, the results, criterion by criterion, for each country are shown. It is on this basis that countries can decide whether to take measures that concern them.

The DAG approach also makes it possible to indicate that certain factors exert - overall - stronger influence than others. This helps to understand the overall dynamics but should not lead to the same policy recommendations for all countries.

Without question, the analytical tools used in this article highlight their great usefulness in finally being able to discern the causal links among several variables. This opens great prospects for the social sciences in general and more particularly for research on the synthetic indicator measuring the quality of education systems in OECD countries (ISQ). Indeed, moving from partial correlations to causal relationships is an opportunity for researchers to better understand social, economic, and scientific phenomena, although Bayesian networks must meet strict conditions before they can be used as causal Bayesian networks.

Acknowledgement

The authors thank Professor Eric Blankmeyer (Texas State University) for careful reading.

Conflict of interest

The authors declare no conflict of interest.

References

Ben Hassen, H., Masmoudi, A. and Rebai, A. (2008). Causal inference in bio-molecular pathways using a bayesian network approach and an implicit method. Journal of Theoretical Biology, 253(4) :717 - 724.

Cartwright, N. (1994). Nature's Capacities and Their Measurement. Oxford University Press.

Chickering, D., Geiger, D. and Heckerman, D. (1995). Learning bayesian networks: Search methods and experimental results. Proceedings of Fifth Conference on Artificial Intelligence and Statistics, 112-128.

Chickering, D. and Heckerman, D. (1996). Efficient Approximation for the Marginal Likelihood of Incomplete Data given a Bayesian Network. UAI'96, 158-168. Morgan Kaufmann.

Fan, X., Chen, M. (2001). Parental Involvement and Students' Academic Achievement: A Meta-Analysis.Educational Psychology Review 13, 1-22.

Friedman, N., Goldszmidt, M. & Wyner, A. (1999). Data analysis with bayesian networks: A bootstrap approach. UAI'99: Proceedings of the 15th annual conference on uncertainty in artificial intelligence, 196-205.

Gámez, J.A., Mateo, J.L. & Puerta, J.M. (2011). Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood. Data Min Knowl Disc 22, 106-148.

Gerard, F.-M., Hugonnier, B. & Varin, S. (2017). La qualité des systèmes éducatifs des pays de l'OCDE enfin mesurée, in B. Hugonnier & G. Serrano (Eds.). Réconcilier la République et son école. Paris: Éditions du Cerf, 61-73.

Gerard, F.-M., Hugonnier, B. & Varin, S. (2018). Mesure de la qualité des systèmes éducatifs des pays de l'OCDE, in ADMEE-Europe, L'évaluation en éducation et en formation face aux transformations des sociétés contemporaines, Proceedings of the colloquium, Esch-sur-Alzette: Université de Luxembourg, 131-143.

Gerard, F.-M., Hugonnier, B. & Varin, S. (forthcoming). Indicateur synthétique de la qualité des systèmes éducatifs des pays de l'OCDE: comparaison des résultats 2015 et 2018.

Hernán, M.A., Robins, J.M. (2018). Causal Inference. Chapman and Hall/CRC.

Hoover-Dempsey, K. V., Battiato, A. C., Walker, J. M. T., Reed, R. P., DeJong, J. M., & Jones, K. P. (2001). Parental involvement in homework. Educational Psychologist, 36(3), 195-209.

Imbens, G.W., Rubin, D.B. (2015). Causal Inference for Statistics, Social, and Biomedical Sciences. Cambridge university press.

Kalisch, M., Mächler, M., Colombo, D., Maathuis, MH., Bühlmann, P. (2012). Causal Inference Using Graphical Models with the R Package pcalg. Journal of Statistical Software, 47(11), 1-26.

Lauritzen, S. (1996). Graphical Models. Oxford University Press, Oxford.

Maathuis, MH., Colombo, D., Kalisch, M., Bühlmann, P. (2010). Predicting Causal Effects in Large-Scale Systems from Observational Data. Nature Methods, 7, 261-278.

Maathuis, MH., Kalisch, M., Bühlmann, P. (2009). Estimating High-Dimensional Intervention Effects from Observational Data. The Annals of Statistics, 37, 3133-3164.

Morgan, S.L., Winship, C. (2014). Counterfactuals and Causal Inference. Cambridge university press, 2nd edition.

Murphy, K. (2007). How to use the Bayes Net Toolbox. Available at: https://web.archive.org/web/20140626142205/http://bnt.googlecode.com/svn/trunk/docs/usage.html

Naïm, P., Wuillemin, P-H., Leray, P., Pourret, O. and Becker, A. (2007). Réseaux bayésiens. Eyrolles, Paris, 3rd edition.

Naïm, P., Pourret, O and Marcot, B. (2008). Dirichlet process gaussian mixture models :Choice of the base distribution. Bayesian Networks: A Practical Guide to Applications, Wiley.

Organization for Economic Cooperation and Development (2019). PISA 2018 results (Volume I): Students' knowledge and skills. OECD Publishing. https://doi.org/10.1787/ec30bc50-fr

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann.

Pearl, J. (1995). Causal diagrams for empirical research. Biometrika, 82(4):669-688.

Pearl, J. (2000). Causality: models, reasoning, and inference. Cambridge university press.

Pearl, J. (2003). Statistics and causal inference: A review. Test, 12(2):281-345.

Pearl, J. (2009). Causality: Models, Reasoning, and Inference. Cambridge University Press, 2nd edition.

Peters, J., Janzing, D. (2017). Elements of Causal Inference - Foundations and Learning Algorithms. MIT Press.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Scutari, M. (2017). Bayesian Network Constraint-Based Structure Learning Algorithms: Parallel and Optimized Implementations in the bnlearn R Package. Journal of Statistical Software, 77(2), 1-20. doi:10.18637/jss.v077.i02.

Spirtes, P., Meek, C. and Richardson, T. (1999). An algorithm for causal inference in the presence of latent variables and selection bias. In Computation, Causation, and Discovery, 211-252. AAAI Press, Menlo Park, CA

Spirtes, P., Glymour, C., Scheines, R. (2000). Causation, Prediction, and Search. Adaptive Computation and Machine Learning, 2nd edition. MIT Press, Cambridge.

Talbot, D. (2012). Introduction to a graphical approach to causal inference. Association des statisticiennes et statisticiens du Québec. Available at: https://www.association-assq.qc.ca/2012/02/16/introduction-a-une-approche-graphique-dinference-causale/

Vanderweele, T.J. (2015). Explanation in Casual Inference: Methods for Mediation and Interaction. Oxford University Press.

Websites for figures 1 and 2

https://dev.to/jjb/part-16-detecting-graph-cycles-with-depth-first-search-4nh3
https://hazelcast.com/glossary/directed-acyclic-graph/