A Bayesian Network for School Performance

Massoud Moussavi[*]
Causal Links, LLC

Noel McGinn[†]
Causal Links, LLC

Abstract

This paper describes a Bayesian Network model to diagnose the causes of low effectiveness of certain schools. Our aim is to build tools to assist policymakers in education to think through a policy, evaluate various scenarios, and choose among competing policy options. These tools would help decision makers to make their tacit knowledge more explicit, and assimilate and systematize information from other sources. The model we describe has two potential uses: the explanation of learning outcomes in terms of conditions and processes within schools that are difficult to observe directly; and the estimation of the probability that a given intervention will affect those conditions and processes and hence learning outcomes. We believe that models of this kind can be effective aids in making decisions, and in learning from them.

1 INTRODUCTION

In the last two decades, a growing body of research has focused on understanding and developing causal as well as diagnostic models and their applications to policy analysis. (Pearl 2000; Hausmann, Klinger, and Wagner 2008). Our approach to development policy analysis is to build on these advances, develop theoretical models, and link them to a specific body of development knowledge. Specifically, we aim to develop a diagnostic model of school effectiveness or quality. That is, we focus on identifying major factors that cause an education system to be ineffective. The underlying framework for our approach is to build a model that represents the teaching and learning process as it takes place in schools. The work requires synthesizing results of empirical research and field experience, consistent with advances in learning and organizational theory. Bayesian modeling approach is particularly suited for integrating and synthesizing information from a variety of sources.

Using a Bayesian network approach model as our inference engine, we have built a software tool, called Policymakers Workbench. Through this tool, policymakers can run unlimited iterations of various policy scenarios and connect existing knowledge to effective action. Policymakers Workbench has two main components: a "theoretical engine" and a computer user interface. The theoretical engine or the model is built upon a set of underlying relations or functions, such as "if attendance goes up, achievement goes up, but only if there is time on task". This model is elaborated as a Bayesian belief network that specifies dependencies among the variables, which are based on the results of previous research, direct experience, and the testimony of experts and stakeholders in the particular domain or field. The computer interface enables users to evaluate the effect of changes in specific variables; perform diagnostics; and analyze the impact of various interventions or policies. A brief description of the user interface is given in Appendix 1.

While much of the discussion in this paper is focused on education, the framework and the tool that we discuss is general and could be employed in policy analysis in other development areas, such as health, the environment, economic growth, migration, and so forth.

[*] moussavi@causallinks.com
[†] mcginn@causallinks.com

2 CONSTRUCTION OF THE MODEL FOR SCHOOL PERFORMANCE

2.1 BACKGROUND AND PREVIOUS WORK

Relatively little work has been published on the use of Bayesian networks in primary and secondary education. The most common application has been to individualization of instruction (Chung et al., 2007; Piombo, Batatia, & Ayache, 2003; Tapia, Lopez, Galan, & Rubio, 2008). Some researchers have focused on models designed to predict which schools are most likely to implement certain innovations in health education (Bosworth, Gingiss, Potthoff, & Roberts-Gray, 1999; Gingiss, Roberts-Gray, & Boerm, 2006; Roberts-Gray, Gingiss, & Boerm, 2007). To our knowledge, our model is the only model that deals with the school performance.

We reviewed the research on determinants of student achievement, school effectiveness and learning outcomes (Coe & Fitz-Gibbon, 1998; Dimmock, 1993; Fuller, 1987; Glewwe & Kremer, 2005; Haertel, Walberg, & Weinstein, 1983; Marzano, 2003; Reynolds & Teddlie, 1999; Scheerens, 2000; Wyatt, 1996). We looked also at conceptual models of how schools produce learning (Heneveld, 1994; McIlrath & Huitt, 1995; Van der Werf, Creemers, De Jong, & Klaver, 2000).

2.2 SCHOOL PERFORMANCE MODEL

We defined the amount of learning of curriculum content attributed to schools as a function of how much time is spent by students on learning that content and the rate at which students learn. Originally proposed by Carroll (1963), this conceptualization makes it possible to posit causal relationships between student and teacher characteristics and behavior, students' family experiences, school and community contextual variables, and learning outcomes.

School effectiveness refers to the achievement of the system's objectives, for example learning of specified contents, skills and values. Our focus, therefore, is on factors that affect the amount of time students spend on learning curriculum content. Schools are organized to provide opportunities for learning, principally through teaching but also through self-instructional methods. Students can also learn the curriculum outside schools, through teaching provided by others and by self-instruction.

A broad and partial conceptualization of our model is shown in Figure 1[1]. As seen in this partial model the effect of input variables (e.g., *student attendance, teacher attendance, class size, teachers' academic education*, etc.) on the learning outcome (which is measured by the output variable expected score) is mediated by a number of hidden or latent variables (*teacher motivation, teacher knowledge, curriculum coverage, quality of teaching/learning*, etc.)

Our complete model for school performance has about 60 linked variables, with built-in probabilities based on the research literature[2]. Various data sets provide data for about only one third of the variables in the model[3]. In our model, the majority of the nodes without parents are input variables for which typically data is regularly collected by national governments or international agencies. Thus, for these variables prior probabilities are assigned on the basis of existing data for schools in a particular country. We developed the structure of the model – how variables influence each other–based on our review of research literature and our personal experience. Once we established the variables and the relationships among them, we specified the conditional probabilities according to experts' opinion, the literature review, and in a few cases based on automated learning from data. But as there is very limited data available on the hidden variables defined in the model, we could not rely on automated learning of all the conditional probabilities from data.

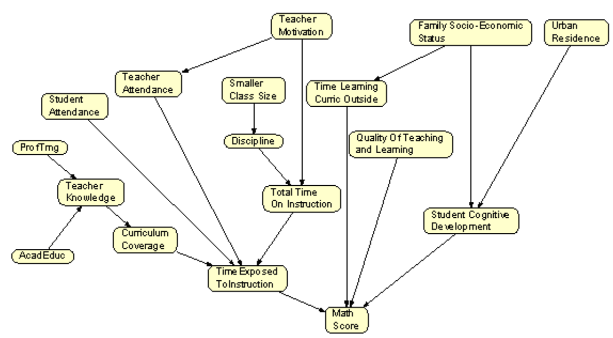

Figure 1: The broad and partial outline of the model

In order to simplify the acquisition of these probabilities from the experts, all the variables have been

[1]Not all the variables or links are shown in this partial model. For example, some variables such as Quality of Teaching and Learning that are shown without parents here are influenced by many other variables in the complete model.

[2]We have built slightly different versions of the model given different countries' context and the available data.

[3]As this model has been developed for the developing countries, we have used data from various development agencies such as UNESCO, USAID, and The World Bank.

defined as having binary values. Rather than assigning a specific probability value to a link, we chose to qualitatively assign nominal values of *very low, low, medium, high, very high.* Internally, these qualitative values were defined as various constants so that we could easily modify them in the system for testing the sensitivity of the results to small changes. Generally we settled on 0.1, 0.3, 0.5, 0.75, and 0.9 for these qualitative values respectively. We have performed numerous analyses regarding the sensitivity to changes in these numbers. In general, little or no sensitivity was observed. We have not yet tested the sensitivity of the model to domain size (i.e., specifying more than two values for each variable). There is a fairly significant number of studies concerning sensitivities to imprecision in Bayesian networks. Some early research on diagnostic models has shown that relatively large models are not sensitive to an increase in domain size. (Pradhan, et al.,1996). However, more recent research has highlighted sensitivity to imprecision. (Druzdzel and Onisko, 2008) and (Coupe and van der Gaag, 2002).

We used the software package Netica (www.norsys.com) to construct our model.

2.3 DATA SETS USED

Data for our study were provided by the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), a joint venture of the International Institute for Educational Planning of UNESCO and 15 ministries of education in the region (International Institute for Educational Planning, 2006).[4] The SACMEQ data were sufficient to construct variables representing 14 variables in our model. The data were not sufficient, however, to provide values for any of the hidden variables.

The analyses reported in this study were based on data from Botswana. The Botswana sample includes complete data from 3322 grade 6 pupils, 170 teachers and principals (or their representatives) in 170 schools. Most school samples included 20 pupils; some 78 cases were lost because of missing data or records. The final sample is equally balanced between boys and girls. In schools with more than one grade 6 class, students were sampled from each class, and one teacher was randomly selected. Responses were coded and data cleaned by the central Consortium. Data files are available through IIEP or SACMEQ.

[4] An earlier version of the model used Research Triangle Institute's data assessment tool called Snapshot for School Management Effectiveness (SSME). That version was tested in a pilot application in primary schools in Peru and Jamaica under a USAID funded project.

2.4 VALIDATION

The standard procedure for validation of a regression equation is estimation of the statistical significance of the regression coefficient. We constructed a regression equation to predict individual reading and mathematics test scores. We obtained an adjusted R square of .23 for the equation predicting English scores, and an adjust R square of .18 for the equation predicting Math scores. The R square is usually interpreted as the amount of variance in the dependent variable that is explained by the regression equation. These R squares are equivalent in size to those found in other studies of external or "input" variables explaining achievement test scores (Scheerens, 2005).

We then used our model to predict English and Math scores. It turned out that our model's prediction was as good as the regression analysis. This could be interpreted to indicate that our Bayesian network model is at least as valid as the regression model in explaining variance in student achievement scores.

Our main interest, however, is not to use the model to predict test scores but rather to determine the causes of low effectiveness (often measured by low test scores) in schools. School effectiveness is usually understood in terms of learning in school, but could also include curriculum learning outside school (e.g., through homework). In fact, most of the input variables for which data is available contribute learning outside of school. Furthermore, the effects of variables like teacher education and training are mediated by other intervening variables such as teachers' knowledge and skill.

3 USING THE NETWORK TO IDENTIFY CAUSES OF LOW EFFECTIVENESS

Using the model in a diagnostic fashion, we can identify the probability estimates for variables that strongly affect the learning outcome (i.e. test scores). Some of the most important variables turn out to be the following: *TimeExposedToInstruction*, which captures the influence of teacher and student attendance; *Curriculum Coverage* which reflects the combined impact of time on instruction and teacher's attention to the curriculum; *QualityOfTeaching and Learning* which reflects the influence of student health and attitudes as well as of variety in teaching practices; *Time Learning Outside School* which represents the impact of family and community on school learning beyond the classroom; and *StudentCognitiveDev* which measures the development stage of the student which is affected by a variety of factors including his previous schooling, family status, etc.

The results presented above hold no surprises for good teachers and students of the teaching-learning process. What they demonstrate is that in order to improve learning in school various factors must be considered. As a result interventions that would be most effective might be quite different from interventions at the input level (e.g., textbooks, library, etc.) that are commonly decided as national policies.

4 CONCLUSION

Many of the features of the teaching and learning process are difficult to observe and not measured in studies of school operation and student learning. Most current policy analysis relies on data that describes only some of the material and human resource inputs to the school and characteristics of students. These factors interact in unspecified ways in the complex process of instruction and learning, and are insufficient to explain most of the variation in measures of learning outcomes. This complexity is seen in our analysis of different data sets. Schools achieve relative equal levels of effectiveness (average student test scores) with widely differing levels of inputs and combinations of instructional practices. Our model reflects at least some of the complexity of teaching and learning.

At the same time, the lack of data describing the intermediate variables of the network limits our confidence in statements about that complexity. These variables have been measured in small sample, experimental studies, but not in sample surveys such as SACMEQ and used in policy analysis. Networks of the kind reported in this paper would be much more convincing if they included direct measures of key process variables. For our purposes, the more important variables to observe would be coverage of curriculum and student attention. Asking teachers to indicate subjects covered during the year could assess coverage; attention would require direct observation. Our analysis of the various datasets suggests that little would be gained by observing specific instructional practices.

A model of the kind we have presented can be used in many ways. First, it can be used to distinguish between schools that are truly ineffective because no learning is taking place, and schools whose low average test scores are attributable more to external factors than to school performance as test scores depend not only on what students acquire in school, but also what they have learned elsewhere. Second, the model can be used to suggest different strategies for improving learning, some that change inputs and instructional practices in schools, others that change the school's relationship with families and the community. Third, in cases where no reliable or standard test scores are available, the model can be used in a predictive fashion to determine identify "failing schools. Finally, the model can be used as a very practical tool for educating policymakers and school administrators.

As for future work on the model, many challenging issues remain. Perhaps the most important is to develop dynamic models. Education by nature is a long term process and requires models that can take advantage of time series data. In this regard, of course, collection of data for the same schools over a number of years and the quality of data sets are key challenges. Another interesting and challenging work would be to learn more from data. Currently, in our model we have mostly relied on experts' opinion and existing research on building the structure of the model and estimating the conditional probabilities. Given more comprehensive datasets which include observations about some of the key intervening variables in our model (e.g., coverage of curriculum), we can perform a more reasonable automated learning. In fact, we feel that raising the awareness on the importance of such variables has been a major contribution of our work. In our discussions with the development agencies, they have indicated that in their future data collection efforts they will include variables such as curriculum coverage.

5 Acknowledgement

The authors wish to thank the anonymous reviewers for their helpful comments. They are also thankful to Ryan Dellolio for helping with the typesetting of the paper.

References

Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J.-D. (2007). Modeling diagnostic assessments with Bayesian networks. Journal of Educational Measurement, 44(4), 341-359.

Berger, J. O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. Journal of the American Statistical Assocation, 95(452), 1269-1276.

Bosworth, K., Gingiss, P. M., Potthoff, S., & Roberts-Gray, C. (1999). A Bayesian model to predict the success of the implementation of health and education innovations in school-centered programs. Education and Program Planning, 22, 1-11.

Carroll, J. (1963). A model of school learning. Teachers College Record, 64, 723-733.

Chung, G. K. W. K., Delacruz, G., C., & Bewley, W. L. (2006). Performance assessment models and tools for complex tasks. CSE Technical Report 682 . Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Chung, G. K. W. K., Delacruz, g. C., Dionne, G. B., Baker, E. L., Lee, J., & Osmundson, E. (2007, April 10, 2007). Towards individualized instruction with technology-enabled tools and methods. Paper presented at the Annual Meeting of the American Education Research Association, Chicago.

Coe, R., & Fitz-Gibbon, C. T. (1998). School effectiveness research: criticism and recommendations. Oxford Review of Education, 24(4), 421-438.

Coupe, V.M.H. and Van der Gaag, L.C. 2002. Properties of sensitivity analysis of Bayesian belief networks. Annals of Mathematics and Artificial Intelligence 36:323-356.

Dimmock, C. (Ed.). (1993). School-Based Management and School Effectiveness. London: Routledge.

Druzdzel, M. and Onisko, A. (2008). Are Bayesian Networks Sensitive to Precision of Their Parameters? In S.T. Wierzchon, M. Klopotek, and M. Michalewicz (eds.), Intelligent Information Systems XVI, Proceedings of the International IIS'08 Conference, pages 35-44, Academic Publishing House EXIT, Warsaw, Poland, June 2008.

Fuller, B. (1987). What school factors raise achievement in the Third World? Review of Educational Research, 57(3), 255-292.

Gingiss, P. M., Roberts-Gray, C., & Boerm, M. (2006). Bridge-It: A system for predicting implementation fidelity for school-based tobacco programs. Prevention Science, 7(2), 197-207.

Glewwe, P., & Kremer, M. (2005). Schools, Teachers, and Education Outcomes in Developing Countries. In E. A. Hanushek & F. Welch (Eds.), Handbook on the Economics of Education (Vol. 2,). Amsterdam: Wlsevier.

Haertel, G. D., Walberg, H. J., & Weinstein, T. (1983). Psychological models of educational performance: A theoretical synthesis of constructs. Review of Educational Research, 53(1), 75-91.

Hausmann, R., Klinger, B., & Wagner, R. (2008). Doing Growth Diagnostics in Practice: A 'Mindbook'. Center for International Development at Harvard University, Working Paper No. 177.

Heneveld, W. (1994). Planning and Monitoring the Quality of Primary Education in Sub-Saharan Africa: AFTHR Technical Note No. 14 . Washington, DC: World Bank.

International Institute for Educational Planning. (2006). SACMEQ–from project to development model. IIEP Newsletter, XXIV(1), 1-5.

Lempert, R. J. (2002). A new decision sciences for complex systems. Proceedings of the National Academy of Sciences of the United States of America, 99(2, Supplement 3), 7309-7313.

Marzano, R. J. (2003). What Works in Schools: Translating research into action. Baltimore, MD: Association for Supervision and Curriculum Development.

McIlrath, D., & Huitt, W. G. (1995). The teaching-learning process: a discussion of models . Valdosta, GA: Valdosta State University.

Piombo, C., Batatia, H., & Ayache, A. (2003). A framework for adapting instruction to cognitive learning styles. Paper presented at the 3rd IEEE International Conference on Advanced Learning Technologies.

Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge, UK.

Pradhan, P., Henrion, G., Provan, G., del Favero, B., & Huang, K. 1996. The sensitivity of belief networks to imprecise probabilities: An experimental investigation. Artificial Intelligence, 85(1-2) 363:397.

Reynolds, D., & Teddlie, C. (1999). The International Handbook of School Effectiveness Research. London: Routledge.

Roberts-Gray, C., Gingiss, P. M., & Boerm, M. (2007). Evaluating school capacity to implement new programs. Evaluation and Program Planning, 30(3), 247-257.

Scheerens, J. (2000). Improving School Effectiveness.

(Vol. No. 68). Paris: UNESCO International Institute for Educational Planning.

Scheerens, J. (2005). Review and Meta-Analyses of School and Teaching Effectiveness . Twente: University of Twente, Department of Educational Organizaton and Management.

Tapia, F. J., Lopez, C. A., Galan, M. J., & Rubio, E. (2008). Bayesian model for optimizing adaptive E-learning process. International Journal of Emerging Technologies in Learning, 3(3), 38-52.

Van der Werf, G., Creemers, B., De Jong, R., & Klaver, E. (2000). Evaluation of school improvement through an educational effectiveness model: The case of Indonesia's PEQIP Model. Comparative Education Review, 44(3), 329-355.

Wooff, D. A., & Schneider, J. M. (2006). A Bayesian Belief Network for quality assessment: Application to employment officer support. Journal of Intellectual Disability Research, 50(2), 109-126.

Wyatt, T. (1996). School effectiveness research: Dead end, damp squib or smouldering fuse? Issues in Educational Research, 6(1), 79-112.

Appendix 1

The User Interface of Policymakers' Workbench

The user interface of Policymakers' Workbench is developed in Java. This program interacts with the quality of schooling model described in this paper.

The user interface enables users to analyze the impact of various interventions or policies and also to examine the trade-offs among different factors, taken individually and in various combinations. Figures 2 and 3 on the following page provide sample screen shots of the system's user interface. Figure 2 shows how the user can decide to look at historical cases or perform a diagnostic. If the option "Perform Diagnostics" is chosen, the user is provided with the screen shown in Figure 3. On this screen, the user can change the value of any variable and observe its impact on all the other variables. As can be seen in the top section of Figure 3, in the educational application the input variables are classified in four general categories: **School Organization** (Principal, Textbooks, Library, etc.), **Family** (Family Socioeconomic Status, Family Involvement, etc.), **Student** (Academic History, Health, Nutrition, etc.), and **Teachers** (Academic Education, Variety in Methods, Training, etc.) The bottom section of the screen in Figure 3, displays the computed or inferred variables (e.g., Total Amount of Learning, Expected Reading Score, etc.) Here again variables are grouped in four categories.

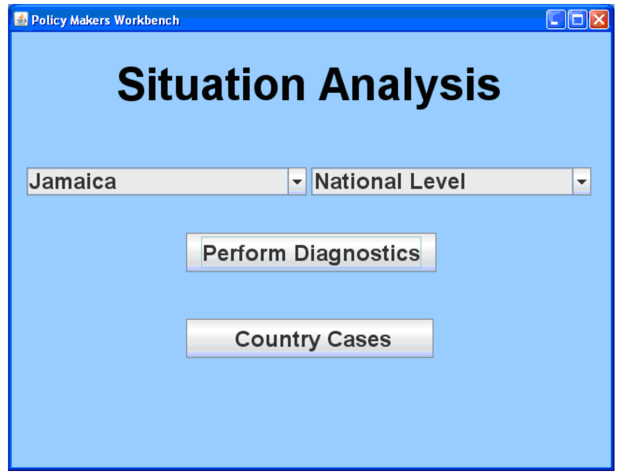

Figure 2: Situation assessment screen

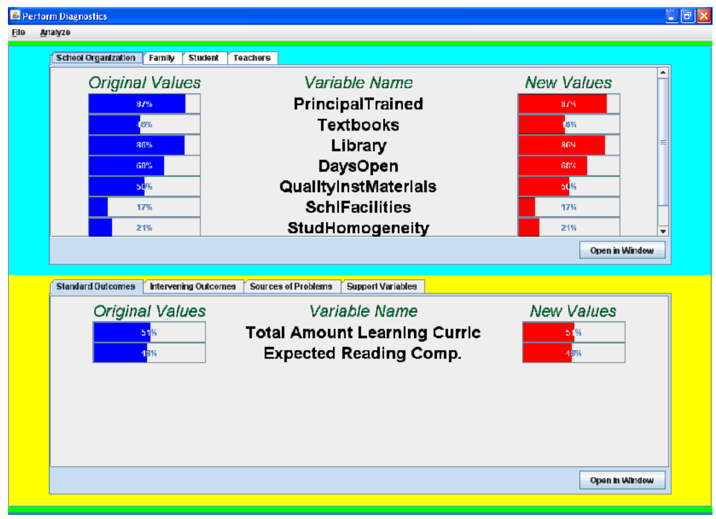

Figure 3: The input and output variables