

Hierarchical latent variable models:

Method, extensions, advantages, and educational policy implications.

Jose Manuel RIVERA ESPEJO

Supervisor: Prof. Geert Molenbegrhs
Affiliation (optional)

Co-supervisor: Prof. Wim Van den
Noortgate *(optional)*
Affiliation (optional)

Mentor: *(optional)*
Affiliation (optional)

Proposal presented in
fulfillment of the requirements
for the degree of Master of Science
in Statistics and Data Science:
Social, Behavioral and Educational Sciences

Academic year 2020-2021

The short and long term benefits of effective teaching practices can be observed throughout the literature: improvements in student achievements (Rockoff, 2004; Rivkin et al., 2005; Duflo et al., 2009; Hanushek and Rivkin, 2012; Muralidharan and Sundararaman, 2013; Chetty et al., 2014a; Araujo et al., 2016); development of executive functions (Araujo et al., 2016), increased college attendance, higher salaries, lower possibility of premature parenthood (Chetty et al., 2014b), among others. Similarly, the literature has shown most of the negative impacts resulting from the presence of teacher shortages¹ (Duflo et al., 2009; Muralidharan and Sundararaman, 2013; Chetty et al., 2015; Ayala, 2017; Marotta, 2019) or ineffective teaching practices (Hanushek and Rivkin, 2012).

However, while the evidence have a solid methodological support, Hanushek and Rivkin (2006) have indicated that some of the proxy variables used, are not consistently related to either teacher effectiveness or quality of instruction, examples of such are: out of field teaching² (Ingersoll, 1998; Dee and Cohodes, 2008; Bertoni, Elacqua, Marotta, Martinez, Méndez, Montalva, Olsen, Santos and Soares, 2020); teaching hours (Bruns et al., 2015); years of experience or educational degree (Rockoff, 2004; Rivkin et al., 2005; Clotfelter et al., 2006, 2007; Hanushek and Rivkin, 2012); among others.

Given the lack of consistency of the effects that arises from using such proxies, Hanushek and Rivkin (2012) have pointed out that the analysis of teacher effectiveness has largely turned away, from attempts to identify the teacher’s specific characteristics, to focus its attention into measuring the direct relationship between them and the student outcomes³. For that reason, considerable uncertainty is still present in the literature, regarding exactly which aspects of teachers are key for the student’s learning and whether those qualities can be measured (Rockoff, 2004; Clotfelter et al., 2006).

However, because the evidence still largely supports the perception that teachers are the main driver behind the student’s learning processes, one of the main points in the agenda of any educational authority should be the design of an assessment system that can attract, select, develop, and retain the most effective ones (Elacqua et al., 2018). But first, the authority would have to define the Educational Performance Standards (EPS) that best agrees with the country’s context. With the EPS establishment, the authorities can set clear expectations about what a ”good” teacher should know and know to do (Cruz-Aguayo et al., 2020).

But because of the uncertainty surrounding such specific requirements, these conditions are not easy to define. Nevertheless, while the specifics are hard to determine, Cruz-Aguayo et al. (2020) has hinted that most of them can be largely grouped into two: (i) to have the disciplinary knowledge and pedagogical practices adequate to the classroom characteristics, context and teaching level, and (ii) to display such knowledge and practices

¹Bertoni, Elacqua, Marotta, Martinez, Méndez, Montalva, Olsen, Santos and Soares (2020) defined it as the context in which the teacher’s supply, i.e. the number of available teachers in the system, is less than its demand. The authors further elaborate that one of the causes of these shortages is related to the applicants’ lower quality or due to their faulty initial training, implying that the shortage can also be conceived as the lack of good quality teachers. In this sense, the evidence of such shortage has been more prevalent, but not decisive, with temporary teachers, as they are usually associated with inferior attributes, compared to their contracted counterparts

²Medeiros et al. (2018) defines it as teachers teaching a subject in which they are not specialized or do not have the appropriate certificate.

³The method is known as value-added analysis, and it is based on the perspective that a good teacher is one who consistently gets higher achievement from students after other determinants of such are controlled for. For a more detailed explanation of the method refer to Scherrer (2011).

in the classroom, using the appropriate material and technological resources available.

As one can infer from the previous general conditions, and the slew evidence, knowledge is a relevant observable factor that it is consistently associated with teacher effectiveness and growth in student's achievement (Santibañez, 2006; Clotfelter et al., 2006, 2007; Hanushek and Rivkin, 2006; Marshall, 2009; Rockoff et al., 2011; Kane et al., 2010; Kane and Staiger, 2012; Ome, 2012; Metzler and Woessmann, 2012; Kane et al., 2013; Araujo et al., 2016; Bietenbeck et al., 2018; Estrada, 2019); and in that sense, its measurement should be of interest for any educational authority.

The measurement of knowledge has a myriad of available tools, and while Bertoni, Elacqua, Méndez, Montalva, Munevar, Olsen and Román (2020) had advocated for the use of multiple instruments, it is important to keep in mind that any educational department are bounded by budgetary constraints. In this setting, and compared to other instruments, valid⁴ and reliable⁵ standardized tests⁶ stand out not only for its cost-effectiveness and a much simpler implementation (Cruz-Aguayo et al., 2020), but also because, they are one of tools with less subjective scoring processes and interpretations.

However, as no instrument is perfect, the teacher's subject knowledge scores will likely reflect measurement error (Metzler and Woessmann, 2012). As established by Angrist and Krueger (1999), measurement error in the explanatory variable could bias the estimated coefficients. This last result implies, that evidence based on test scores could be an attenuated reflection of the true effects. On the other hand, the use of one composite value, i.e. the score, does not allow to test which specific factors -if any- leads to better or worse teacher performance, making also difficult to know which teachers should be hired or what should be done to train them (Hanushek and Rivkin, 2012).

But beyond the use of test results as explanatory variables in modeling processes, there is one more pressing argument on why the issue of measurement error should be addressed: approximately 60% of the Caribbean and Latin American countries use standardized test scores as part of or as a main teacher selection tool (Cruz-Aguayo et al., 2020). In this setting, devoting effort to assess the issues related to measurements errors, could help the educational authorities to understand if the scores thresholds used for the selection are appropriately set, and ultimately, to know the what kind of teachers are being integrated into the public teaching staff.

In summary, teachers are one of the main drivers behind the student achievements. However, some of the evidence supporting this claim has been based on proxy variables that are not consistently related to the quality of instruction, or methods that are not concerned with the outline of the teaching factors responsible for the student's learning. Nevertheless, while the literature still reflects considerable uncertainty on what are the "ingredients for a good teacher", a good amount of evidence has supported the disciplinary and pedagogical knowledge as relevant components of the teacher effectiveness. Finally, the literature has shown that valid and reliable standardized tests are among the best tools to assess such factors, but also have emphasized that such scores could reflect the

⁴the extend to which a measurement tool is well-founded and accurately corresponds to the real measure (Kelley, 1927)

⁵the overall consistency of a measure under consistent conditions.

⁶Assessment instrument in which the implementation, questions, scoring processes, and interpretations are consistent with a predetermined or typified way. The instrument is usually composed of questions or items that fulfill three conditions: (i) they are polytomous, i.e. they have multiple choices, (ii) the choice categories are nominal, i.e. do not present any specific order, and (iii) there is only one "correct" category or answer (Rivera, 2019)

teacher's abilities with considerable noise.

In that sense, this research plans to fill three main literature gaps. First, the researcher will use hierarchical latent variable models to obtain a noise-free score for the competencies of teachers. Second, the method would also help to obtain a dynamic multidimensional depiction of their disciplinary abilities. And lastly, the researcher will tests the real implications of the method in a data composed of repeated large standardized educational assessments from Peru.

Concerning the first two objectives of the research, the author expects to appraise: (i) if hierarchical latent variable models can provide a general framework that could serve multiple psychometric purposes; and (ii) what are the advantages or disadvantages of using such models.

For the last objective, the author expects to shed some lights about key policy decisions related to those large evaluation processes. To mention a few: (i) do the instruments guarantee a fair assessment of minority groups with different abilities?; (ii) are we screening the most knowledgeable teachers?; (iii) what are the general characteristics of the career applicants?; (iv) what differentiate a contract teacher from a temporary one?; (v) what is the main evolution of the disciplinary knowledge of the teachers?, and, is there any identifiable divergence from such pattern?; (vi) does initial training or socioeconomic status proxy variables explain different levels of disciplinary knowledge?; (vii) what specific factor of the disciplinary knowledge is consistently related to classroom observation scores.

In this sense, the researcher believes the present thesis contributes to the literature in two aspects: (i) in a the theoretical and methodological sense, as the research is focused on offering a exhaustive description and analysis of the models; and (ii) in a more practical sense, as it helps to provide evidence on some of key policy decisions that Latin America countries are currently facing.

References

- Angrist, J. and Krueger, A. (1999). Chapter 23 empirical strategies in labor economics, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 3, Elsevier, pp. 1277 – 1366.
URL: <http://www.sciencedirect.com/science/article/pii/S1573446399030047>
- Araujo, M., Carneiro, P., Cruz-Aguayo, Y. and Schady, N. (2016). Teacher quality and learning outcomes in kindergarten, *The Quarterly Journal of Economics* **131**(3): 1415–1453.
URL: <https://publications.iadb.org/publications/english/document/Teacher-Quality-and-Learning-Outcomes-in-Kindergarten.pdf>
- Ayala, M. (2017). *Efecto de los docentes provisionales sobre desempeño escolar - evidencia para la educación secundaria oficial en colombia*, Master’s thesis, Universidad de los Andes.
URL: <http://biblioteca.uniandes.edu.co/acepto201699.php?id=11802.pdf>
- Bertoni, E., Elacqua, G., Marotta, L., Martinez, M., Méndez, C., Montalva, V., Olsen, A., Santos, H. and Soares, S. (2020). Escasez de docentes en latinoamérica: ¿cómo se puede medir y que políticas están implementando los países para resolverlo?, *Technical report*, Banco Interamericano de Desarrollo.
- Bertoni, E., Elacqua, G., Méndez, C., Montalva, V., Munevar, I., Olsen, A. and Román, A. (2020). Concurso docentes en latinoamérica: Claves para mejorar calidad, eficiencia y equidad en educación, *Technical report*, Banco Interamericano de Desarrollo.
- Bietenbeck, J., Piopiunik, M. and Wiederhold, S. (2018). Africa’s skill tragedy: Does teachers’ lack of knowledge lead to low student performance?, *Comparative Education Review* **53**(3): 553–578.
URL: <http://jhr.uwpress.org/content/53/3/553.abstract>
- Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research*, Methodology in the Social Sciences, The Guilford Press.
- Bruns, B., Luque, J., De Gregorio, S., Evans, D., Fernández, M., Moreno, M., Rodriguez, J. Toral, G. and Yarrow, N. (2015). Great teachers: How to raise student learning in latin america and the caribbean, *Technical report*, World Bank Group.
- Chetty, R., Friedman, J. and Rockoff, J. (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates, *American Economic Review* **104**(9): 2593–2632.
URL: <https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2593>
- Chetty, R., Friedman, J. and Rockoff, J. (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood, *American Economic Review* **104**(9): 2633–2679.
URL: <https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2593>

- Chetty, R., Friedman, J. and Rockoff, J. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from kenyan primary schools, *Journal of Public Economics* **123**: 92–110.
URL: <http://www.sciencedirect.com/science/article/pii/S0047272714002412>
- Clotfelter, C., Ladd, H. and Vigdor, J. (2006). Teacher-student matching and the assessment of teacher effectiveness, *Working Paper 11936*, National Bureau of Economic Research.
URL: <http://www.nber.org/papers/w11936>
- Clotfelter, C., Ladd, H. and Vigdor, J. (2007). How and why do teacher credentials matter for student achievement?, *Working Paper 12828*, National Bureau of Economic Research.
URL: <http://www.nber.org/papers/w12828>
- Cruz-Aguayo, Y., Hincapié, D. and Rodríguez, C. (2020). Profesores a prueba: claves para una evaluación docente exitosa, *Technical report*, Banco Interamericano de Desarrollo.
- Dee, T. and Cohodes, S. (2008). Out-of-field teachers and student achievement: Evidence from matched-pairs comparisons, *Public Finance Review* **36**(1): 7–32.
URL: <https://doi.org/10.1177/1091142106289330>
- Duflo, E., Dupas, P. and Kremer, M. (2009). Additional resources versus organizational changes in education: Experimental evidence from kenya.
- Elacqua, G., Hincapié, D., Vegas, E. and Alfonso, M. (2018). Profesión: profesor en américa latina ¿por qué se perdió el prestigio docente y cómo recuperarlo?, *Technical report*, Banco Interamericano de Desarrollo.
- Estrada, R. (2019). Rules versus discretion in public service: Teacher hiring in mexico, *Journal of Labor Economics* **37**(2): 545–579.
URL: <https://doi.org/10.1086/700192>
- Hanushek, E. and Rivkin, S. (2006). Chapter 18 teacher quality, in E. Hanushek and F. Welch (eds), *Handbook of the Economics of Education*, Vol. 2, Elsevier, pp. 1051 – 1078.
URL: <http://www.sciencedirect.com/science/article/pii/S1574069206020186>
- Hanushek, E. and Rivkin, S. (2012). The distribution of teacher quality and implications for policy, *Annual Review of Economics* **4**(1): 131–157.
URL: <https://doi.org/10.1146/annurev-economics-080511-111001>
- Ingersoll, R. (1998). The problem of out-of-field teaching.
URL: https://repository.upenn.edu/gse_pubs/137
- Kamata, A. and Bauer, D. (2008). A note on the relation between factor analytic and item response theory models, *Structural Equation Modeling: A Multidisciplinary Journal* **15**(1): 136–153.
URL: <https://doi.org/10.1080/10705510701758406>

Kane, T., McCaffrey, D., Miller, T. and Staiger, D. (2013). Have we identified effective teachers? validating measures of effective teaching using random assignment, *Research paper*, Bill Melinda Gates Foundation.

URL: <https://files.eric.ed.gov/fulltext/ED540959.pdf>

Kane, T. and Staiger, D. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains, *Research paper*, Bill Melinda Gates Foundation.

URL: https://k12education.gatesfoundation.org/download/?Num=2678filename=MET_Gathering_Fee

Kane, T., Taylor, E., Tyler, J. and Wooten, A. (2010). Identifying effective classroom practices using student achievement data, *Working Paper 15803*, National Bureau of Economic Research.

URL: <http://www.nber.org/papers/w15803>

Kelley, T. (1927). *Interpretation of educational measurements*, Measurement and adjustment series, World Book Co.

Marotta, L. (2019). Teachers' contractual ties and student achievement: The effect of temporary and multiple-school teachers in brazil, *Comparative Education Review* **63**(3): 356–376.

Marshall, J. (2009). School quality and learning gains in rural guatemala, *Economics of Education Review* **28**(2): 207–216.

URL: <http://www.sciencedirect.com/science/article/pii/S0272775708000745>

Medeiros, M., Gómez, C., Sánchez, M. and Orrego, V. (2018). Idoneidad disciplinar de los profesores y mercado de horas docentes en chile, *Calidad en la Educación* (48): 50–95.

URL: <https://doi.org/10.31619/caledu.n48.479>

Metzler, J. and Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation, *Journal of Development Economics* **99**(2): 486–496.

URL: <https://ideas.repec.org/a/eee/deveco/v99y2012i2p486-496.html>

Muralidharan, K. and Sundararaman, V. (2013). Contract teachers: Experimental evidence from india, *Working Paper 19440*, National Bureau of Economic Research.

URL: <http://www.nber.org/papers/w19440>

Muthén, B., Kao, C. and Burstein, L. (1991). Instructional sensitivity in mathematics achievement test items: Applications of a new irt-based detection technique, *Journal of Educational Measurement* **28**(1): 1–22.

URL: <https://files.eric.ed.gov/fulltext/ED338678.pdf>

Ome, A. (2012). The effects of meritocracy for teachers in colombia, *Research report*, Fedesarrollo.

URL: <https://ideas.repec.org/p/col/000124/010260.html>

Rivera, J. (2019). *El modelo de respuesta nominal: Aplicación a datos educacionales*, Master's thesis, Pontificia Universidad Católica del Peru.

URL: <http://hdl.handle.net/20.500.12404/14600>

- Rivkin, S., Hanushek, E. and Kain, J. (2005). Teachers, schools, and academic achievement, *Econometrica* **73**(2): 417–458.
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2005.00584.x>
- Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data, *The American Economic Review* **94**(2): 247–252.
URL: <http://www.jstor.org/stable/3592891>
- Rockoff, J., Jacob, B., Kane, T. and Staiger, D. (2011). Can you recognize an effective teacher when you recruit one?, *Education Finance and Policy* **6**(1): 43–74.
URL: https://doi.org/10.1162/EDFP_a00022
- Santibañez, L. (2006). Why we should care if teachers get a's: Teacher test scores and student achievement in mexico, *Economics of Education Review* **25**(5): 510–520.
URL: <http://www.sciencedirect.com/science/article/pii/S0272775705000804>
- Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea, *NASSP Bulletin* **95**(2): 122–140.
URL: <https://doi.org/10.1177/0192636511410052>
- Sutcher, L., Darling-Hammond, L., and Carver-Thomas, D. (2016). A coming crisis in teaching? teacher supply, demand, and shortages in the u.s., *Technical report*, Learning Policy Institute.
- Takane, Y. and de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables, *Psychometrika* **52**(3): 393–408.
URL: <https://doi.org/10.1007/BF02294363>

AFDELING

Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be

