

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 (???????); development of executive functions (?), increased college attendance, higher salaries, lower possibility of premature parenthood (?), among others. Similarly, the literature has shown most of the negative impacts resulting from the presence of teacher shortages¹ (?????) or ineffective teaching practices (?).

However, while the evidence have a solid methodological support, ? 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 (???); teaching hours (?); years of experience or educational degree (?????); among others.

Given the lack of consistency of the effects that arises from using such proxies, ? 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 (??).

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 (?). 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 (?).

But because of the uncertainty surrounding such specific requirements, these conditions are not easy to define. Nevertheless, while the specifics are hard to determine, ? 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 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 (????????????????); 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 ? 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 com-

¹? 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

²? 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 ?.

pared to other instruments, valid⁴ and reliable⁵ standardized tests⁶ stand out not only for its cost-effectiveness and a much simpler implementation (?), 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 (?). As established by ?, 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 (?).

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 (?). 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 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

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

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 (?)

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.

AFDELING

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

