PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

ANALYSIS OF THE QUALITY OF TEACHING PRACTICES IN UNDER-PERFORMING SCHOOLS IN CHILE

MARÍA DEL ROSARIO ESCRIBANO ALISIO

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences

Advisor:
MIGUEL NUSSBAUM V.

Santiago de Chile, January, 2020

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

ANALYSIS OF THE QUALITY OF TEACHING PRACTICES IN UNDER-PERFORMING SCHOOLS IN CHILE

MARÍA DEL ROSARIO ESCRIBANO ALISIO

Members of the Committee:

MIGUEL NUSSBAUM

ERNESTO TREVIÑO

RICARDO PAREDES

RAFAEL CARRASCO

MAYRA MASCAREÑO

JUAN DE DIOS ORTÚZAR

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences

Santiago de Chile, January, 2020

To Martin and Alicia, my life partners.

For my parents and brothers, for teaching me the option of doubt of what is established.

ACKNOWLEDGMENTS

I would like to start by thanking my advisor, Miguel Nussbaum, from his guidance, patience, and support through this process. Miguel did not doubt on inviting me to work on his group, and on my capacities to carry on this challenge.

Ernesto Treviño, as a member of my committee, took a fundamental role in guiding my work into the area of educational inclusion. Without Ernesto´s help, this work wouldn´t have developed as profoundly as it finished. Denisse Gelber, Diego Carrasco, and David Torres, besides teaching me useful tools for my studies, trusted in my work.

I would also like to thank my colleagues and friends throughout these years. Daniel Araneda, Camila Barahona, Ana Cortés, and Damián Gelerstein for their company, discussions, and academic support.

My friends were also very important in this process. Daniela Carrasco, Catalina Infante, and María Fernanda Vera thank you for your emotional support and for sharing survival strategies.

To Mayra Mascareño, Anneke Timmermans and Roel Bosker, for making me feel like at home in my internship in Groningen, and for their insightful academic and methodological guidance.

To my former students in the Liceo Industrial de Temuco. Because I had the privilege of being their teacher, I now have the motivation to dedicate my career to research in education.

Finally, I want to thank in a special way to Martin Caceres, my husband, friend, and academic partner, for his unconditional support, infinite patience, and unlimited love.

GENERAL INDEX

INDEX OF TABLES

Page

INDEX OF FIGURES

Page

RESUMEN

Una preocupación pendiente pero crucial para la mejora de la calidad y la equidad del sistema educativo es lo que sucede dentro de las aulas. Esta tesis se enfoca en describir y analizar la calidad de las prácticas de enseñanza en escuelas de bajo índice de calidad en Chile. En primer lugar, exploraremos la variabilidad de calidad de las prácticas de enseñanza entre y dentro de las escuelas. En segundo lugar, exploraremos la heterogeneidad de las prácticas docentes basados en un enfoque centrado en la persona, con el objetivo de encontrar categorías de docentes y su relación con los factores escolares. Finalmente, investigaremos la heterogeneidad de las oportunidades de aprendizaje dentro del aula a través de mediciones de las interacciones profesor-alumno, con un enfoque en el grupo minoritario de migrantes. El primer estudio muestra que, aunque las escuelas son relativamente homogéneas en sus resultados académicos, la calidad de las prácticas de los maestros dentro de estas escuelas varía enormemente. Este hallazgo es relevante para la política pública, ya que indica que en la mayoría de estas escuelas de bajo rendimiento es posible encontrar maestros con un buen desempeño docente de acuerdo con los estándares de la Agencia de Calidad. Los resultados del segundo estudio muestran cuatro categorías de enseñanza, prácticas de alta calidad (38.47% de la muestra), prácticas de baja calidad con (25, 77%), prácticas orientadas al contenido (18.88%) y prácticas orientadas al clima (16.88%). Este hallazgo muestra que los maestros tienen diferentes patrones en su calidad de prácticas, especialmente aquellos maestros que están en el rango medio de calidad de la enseñanza. Los resultados del tercer estudio muestran que existen diferencias en las oportunidades de aprendizaje proporcionadas por el profesor a diferentes grupos de estudiantes en un aula.

Esta tesis contó con el apoyo de la beca CONICYT PFCHA/BECA DOCTORADO NACIONAL/2016 - 21160063

Palabras Claves: Calidad de prácticas docentes, Escuelas de bajo rendimiento, interacciones de aula, observaciones de aula, métodos mixtos, análisis de clases latentes, redes de interacción profesor-estudiante.

ABSTRACT

A remaining but crucial concern for the improvement of both the quality and equity of the educational system is what happens inside the classrooms. This thesis focuses on describing and analyzing the quality of teaching practices in under-performing schools in Chile. Firstly, we will explore the quality variability of teaching practices between and within schools. Secondly, we will explore the heterogeneity of teaching practices based on a person-centered approach, with the aim of finding categories of teachers and their relationship with school factors. Finally, we will investigate the heterogeneity of learning opportunities within the classroom through measurements of teacher-student interactions, with a focus on the minority group of migrants. The first study shows that, although the schools are relatively homogeneous in their academic results, the quality of the practices of the teachers within these schools varies greatly. This finding is relevant to public policy since it indicates that in most of these low-performing schools, it is possible to find teachers with good teaching performance according to the standards of the Quality Agency.The results of the second study show four teaching categories, high-quality practices (38.47% of the sample), low-quality practices with (25, 77%), content-oriented practices (18.88%), and climate oriented practices (16.88%). This finding shows that teachers have different patterns in their quality of practices, especially those teachers who are in the middle range of teaching quality. The results of the third study show that there are differences in learning opportunities provided by the teacher to different groups of students in a classroom.

This research was supported by the project CONICYT PFCHA/BECA DOCTORADO NACIONAL/2016 – 21160063

Keywords: Quality of Teachers Practices, Underperforming Schools, Classroom Interactions, Classroom Observations, Mixed Methods, Latent Class Analysis, Teacher-students Interaction Networks.

INTRODUCTION

The Chilean school system, according to an OECD report (2016), is the most segregated among the countries participating in the PISA study, with academic results strongly linked to socioeconomic level, which translates into a high level of educational inequality (Valenzuela et al., 2014). This inequality, in practice, implies that schools that serve students with a high vulnerability index generally do not meet the minimum standards to ensure student learning. Some examples include schools where there are no specific measures to overcome the challenges of their particular social context, or there is a lack of infrastructure required for the educational process. This helps to explain why the school effectiveness results in developing countries are different from those of developed countries (Teodorović, 2011, Treviño et al., 2010).

Several studies have consistently shown that what happens in the classroom, specifically the teacher's actions, explains a high percentage of the variance in student achievement (Barber & Mourshed, 2007). Furthermore, variance in student achievement is better explained at the classroom level than at the school level (Muijs and Reynolds, 2011; Bill & Melinda Gates Foundation, 2013; Youn 2016). Considering the segregation of the Chilean system (Valenzuela et al., 2014), the fact that a good teacher can have a significant impact on a student's opportunities (Barber & Mourshed, 2007; Ortega, Malmberg & Sammons, 2017), and that the results of effective teaching practices vary in different contexts (Teodorović, 2011, Treviño et al., 2010), it is important to explore the quality of teaching in under-performing Chilean schools.

In this regard, it is important to consider the conceptualization of the quality of teaching practices, and consequently, how they are measured.

Traditionally, the quality of teaching practices is operationalized as the aggregate effect of increasing student academic achievement (Seidel & Shavelson, 2007). However, with this type of measurement it is not possible to identify which specific practices helped improve student performance (Kane et al., 2011). One option that has been gaining popularity in measuring the quality of teaching practices is the classroom observation model (Pianta et al., 2008; Stallings 1977). Although this model helps to obtain more precise information for the improvement of teaching practices, the way that the concept of teaching quality is operationalized with the information collected through observation remains open to debate (Praetorius and Charalambous, 2018). Additionally, the use of classroom observation as an indicator of teacher quality assumes that the learning opportunities are homogeneous for all students. In other words, this measurement technique does not evaluate whether the quality of teaching practices varies for different students in the same classroom.

Based on the arguments mentioned above, this thesis will focus on describing the quality of teaching practices in under-performing schools in Chile. Firstly, we explore the variation in the quality of teaching practices both between and within schools. Secondly, we explore the heterogeneity of teaching practices based on a person-centered approach, with the aim of identifying categories of teachers and their relationship with school factors. Finally, we investigate the heterogeneity of learning opportunities within the classroom through measurements of teacher-student interactions, with a focus on groups of migrants.

1.1 Policies for improving quality and equity in Chilean schools

In recent years, improving quality and equity in education has been at the center of the public policy debate in Chile. However, although there have been important advances, the goal of delivering quality education to the entire population is still far from being achieved (OECD, 2016). In the 1990s, after the civic-military dictatorship had abandoned educational policy (Cox, 2002), various efforts were made to improve the system. On the one hand, efforts were made to improve infrastructure (Cox, 2002). Subsequently, programs focused on low-performing schools with fewer resources, such as MECE and P900, were designed and implemented (Angell, 2002). Efforts were also made to improve the working conditions of teachers, and school schedules were changed with the implementation of the "Complete School Day" (Martinic and Vergara, 2007). Despite these efforts, national and international assessments in the early 2000s showed insufficient progress (Cox, 2010). In light of this evidence, efforts focused on successive curricular reforms, with the aim of improving student learning in the classroom (Cox, 2010).

Even so, dissatisfaction with the educational system that continued to be highly segregated, and with a profound market orientation, led to strong social movements calling for a change in its organization (Bellei et al., 2018). Due to citizen pressure, the Preferential School Grant was created, which allocates greater resources to schools for each vulnerable, or priority student they serve, and for serving a higher concentration of these students (Treviño et al., 2019). Subsequently, in 2012, a system for quality assurance of education was developed. Under this system, the Agency for Quality Education was created, with the mission of evaluating, ranking, and creating guidelines for schools, while the Education Superintendency was also created, with

the mission of inspecting schools (Treviño, 2018). Subsequently, major reforms were carried out that aimed both to reform the structure of the system and to generate stronger capacities. The Inclusion Law brought an end to for-profit subsidized schools, as well as co-payments by families, and prohibited student selection (20.845, 2015). In addition, a new public education system was developed, which has ended the administration of schools by local municipalities. A Professional Teacher Development system was also established, which improved the working conditions of teachers, and opportunities and regulations have been developed from initial teacher training through to retirement (20.903; 2016).

The latest reforms, which are still in the process of implementation, have focused on structural conditions for greater quality and equity. However, an important debate that is still pending is about what happens in the classroom (Muijs and Reynolds, 2011; Bill & Melinda Gates Foundation, 2013; Youn 2016). For this reason, this thesis focuses on teaching practices in schools with a focus on both equity and quality. The evidence shows that actions are especially critical in more vulnerable schools (Nye et al., 2004; Sanders & Rivers, 1996; Torres, 2018). On the other hand, it is important to note that the context is extremely important since it determines both the results of the research and the success or failure of the policies (Lupton, 2005; Thrupp, M., & Lupton, R. 2006).

1.2 The role of research in public policy

1.2.1 Combining policy and methodology

The educational reforms of recent decades at the international level have tended to be based on accountability systems, which is particularly relevant in Chile, given the privatization of education and the New Public Management structure of educational and social policy (Bellei et al., 2003; Bellei et al., 2015). While these systems include incentives to make decisions that improve resource efficiency and equity, there is also evidence that they may have unwanted effects. Some examples include limiting the curriculum to what can be assessed by standardized tests, establishing economic rationality over pedagogical criteria in school decisions, and selecting students with fewer learning challenges as a way to improve the school's academic performance (Treviño et al., 2018). Thus, as a formula to improve student learning, this type of high-impact accountability reform has not yielded the expected results, since schools tend to look for shortcuts to improve their results rather than profoundly transforming their pedagogy.

In Chile, all students are required to take standardized high-stakes tests, which can lead to monetary benefits for the school or its closure in cases where a trend of poor results is not reversed. For example, currently this decision is based on the performance index calculated by the Agency for Quality Education. This index is mainly based (67%) on academic performance in the standardized SIMCE test. Although these results are corrected by the GSE, it is still a value-added measurement. While value-added measurements are useful in certain contexts, they do not provide information regarding specific practices for improvement (Kane et al., 2011). In Chile, this logic tends to simplify complex processes, focusing exclusively on how schools

perform on these tests, while leaving aside other important pedagogical aspects (Elacqua et al., 2016). Clearly, the underlying methodological approach of public policy instruments in the Chilean educational system is very similar to the accountability approach, which is linked to a top-down model of educational change (Treviño et al.,2016b). This methodology influences school practices, especially classroom practices, but the unit of analysis remains primarily at the school level.

Given that education policy is usually based on different methodological approaches, as highlighted above, addressing educational reform is not only a policy issue, it also has an important methodological aspect in terms of measurement and its consequences.

1.2.2 The school as the change agent

In the quality assurance system, the school is considered as the main agent of change. The Agency for Quality Education evaluates the schools, which has consequences for their financing and they can even be closed if they are poorly evaluated (law 20.248, 2008; law 20.529, 2011). In other countries, such as the US, there is a vision of teacher accountability, but this does not always bring good results since, as mentioned previously, this vision simplifies processes and does not provide enough information for teachers to change their practices.

The focus of efforts on the change agent, in this case the school, is based on a key assumption, which is that schools are homogeneous in terms of their teaching quality. Through quantitative analysis, this research aims to help understand a flaw in public policy development in Chile, focusing on the classroom to gather evidence that presents teachers as change agents.

1.3 What we know and what we still need to learn

1.3.1 What we know: Factors associated with school quality

There is a strain in education literature called school effectiveness, which is dedicated to studying the factors associated with the quality of education. Originally, this was done through input-output type models (Hanushek, 1989). In the 2000s (Schereens, 2000), a more comprehensive model was proposed, which considers the processes in the classroom and the school, while always controlling for contextual factors. Since then, the discussion of factors associated with educational quality has begun with a focus on the related processes. This has resulted in models that not only include the variables of the input-output models, but also variables that attempt to assess the performance of teachers and schools in educational processes.

At the school level, Scheerens (2015) summarizes the results from various studies that include factors associated with the school's academic performance, namely effective leadership, academic focus, an orderly and positive environment, high expectations, progress monitoring at all levels, parent and student commitment and teacher effectiveness. Teddlie and Stringfield (2007) also mention the development of capacities and instructional effectiveness of the pedagogical team as relevant factors in the process of achieving effective schools. Muijs et al. (2014) suggest that the most relevant factors in this regard are opportunities to learn, quality of teaching and interaction, school climate and teacher expectations. According to a meta-analysis by Hattie (2008), some factors that have a strong influence (effect greater than 0.6) are high teacher expectations regarding student performance, student-teacher relationships, feedback, the credibility of the teacher according to the students and the implementation of formative teacher

evaluations. This study shows that there are positive and significant relationships between students' academic performance and a climate of school coexistence, greater parental involvement, better leadership, and improved teacher feedback. In addition, the third regional comparative and explanatory study, or TERCE (Treviño et al., 2015b), identifies different factors associated with academic performance in Chile, including the classroom climate, availability of resources in the classroom, effective scheduling by teachers and school infrastructure, among others.

1.3.2 What knowledge gap do schools aim to help close?

Given that school improvement is a complex phenomenon, especially in vulnerable contexts, it is necessary to see teaching practices from different levels. First, this thesis questions the assumption of heterogeneity in the school, where it is assumed that the school is the change agent. In this regard, it is necessary to assess whether the quality of teaching practices varies within and between schools. Without ignoring the role of the school in possible changes in the quality of teaching practices (Santelices, 2015), we aim to identify which school factors are associated with teaching practices within a framework of heterogeneous teachers. So far, teaching practices have been seen as homogeneous in terms of student learning opportunities. Therefore, it is important to determine whether this heterogeneity also occurs at the level of teacher-student interactions. In the following sections we will address the need to study this topic in greater depth.

1.4 Variability of teaching practices

Several studies have suggested that what happens at the classroom level is a more important factor in student academic performance than the school level (Muijs and Reynolds, 2011), specifically the interactions between teachers and students (Muijs et al., 2014 ; Ortega et al. 2018). However, the heterogeneity of teaching practices has been little studied. Considering the importance of the classroom climate, which can be compared to the effect of the students' social context (Wenglinsky, 2004), the question arises of how homogeneous teaching quality is at the classroom level.

Evidence shows that teachers perform differently (Hanushek & Rivkin, 2012), or that they are not distributed randomly in schools (Valencia, E., & Taut, S., 2011; Meckes y Bascopé, 2012), which could indicate heterogeneity in teaching quality practices, but there is no quantification of this heterogeneity in the context of vulnerable schools. Ignoring this heterogeneity can lead to problems when developing more efficient public policies since a school could be classified homogeneously as bad, and its existing human capital would not be considered to improve practices among teachers through collaborative work. Considering the above, the question arises: To what degree does the quality of teaching vary at under-performing schools?

1.5 Teacher profiles and their relationship with school factors

Once it has been determined whether there is variability in the quality of teaching practices in a given school, it is important to characterize the possible heterogeneity of these practices. In this regard, it is necessary to take into account the complexity of teaching practices. This can be

understood as ceasing to see teacher quality as a one-dimensional attribute, and beginning to look for whether certain patterns of teaching practices are repeated among individuals. That is, to see if there are teacher typologies, where each typology represents a cluster of teachers who have similar values in their teaching practices.

Although there is evidence that school factors can influence student performance and that the school context in which teachers perform cannot be ignored (Moore Johnson, 2012; Santelices et al., 2015), as mentioned previously, this has a smaller effect than what happens in the classroom. Therefore, while the influence of the school level on student performance cannot be denied, as shown by the literature at the international level, it is necessary to assess whether there are school factors associated with certain teaching practices or teacher profiles in the school. This raises the following questions: 1) Can different teacher profiles be identified in low-quality index schools? 2) Is there a relationship between school-level factors and the probability of a teacher belonging to a specific teacher profile?

1.6 Variability of learning opportunities within the classroom

Another important aspect to consider is whether these teaching practices generate homogeneous learning opportunities among students, which would be the ideal case. Although there has always been heterogeneity in Chilean classrooms (for example, special needs students, gender gaps, or students from minority ethnic groups), in the last decade there has been significant intra-regional migration (INE, 2018). The number of migrants has tripled in the last three years, while the population of migrant students has also risen explosively. These students are concentrated in public schools, in areas with a low socioeconomic context (MINEDUC, 2018), which has

resulted in an increase in classroom diversity in adverse contexts. International literature has shown that migrant students tend to be discriminated against in their schools (den Brok & Levy, 2005, Casteel, 1998; den Brok, Wubbels, Veldman, & van Tartwijk, 2009; Fraser & Walberg, 2005; Irvine, 1985; 1986). However, in Chile studies of this recent phenomenon have been mainly qualitative, or based on a limited sample (Cerón, Pérez Alvarado, & Poblete, 2017; Mondaca, Muñoz, Gajardo, & Gairín, 2018; Tijoux, 2013a).

For these reasons, it is important to consider how diverse teaching practices are in terms of teacher-student interactions in different classrooms, emphasizing the quality and number of teacher interactions with different groups of students. This, in turn, raises the following questions: 1) Do students with an immigrant background have fewer teacher-student interactions compared to their non-migrant peers? If so, does this occur equally in groups of migrants from different countries of origin? 2) Is there variability among teachers regarding the inclusion of migrant students in their interactions? 3) Is there a relationship between the teacher's attitude towards diversity and their level of integration of these students?

1.7 Research Questions

Q1) How varied is the quality of teaching practices within and among under-performing schools?

Q2.1) Can we distinguish different teacher profiles in low-quality index schools?

Q2.2) Is there an association between school-level factors and the probability of a teacher to pertain to a specific teacher profile?

Q3.1a) Are students with immigration background significantly more peripheral in their class' teacher-student interaction networks?

Q3.1b) Do the interaction networks vary when considering the immigrant students' country of origin?

Q3.2) Is there a significant variation in the effect of student immigrant background on the frequency of interactions with the teacher across classrooms?

Q3.3) Can the immigrant composition of the classroom and the teacher's attitude towards diversity in the classroom partially explain the classroom-level variation on the effect of student immigrant background?

1.8 Hypothesis

H1) The variance of quality teacher practices within schools is greater than between under-performing schools.

H2.1) It is possible to distinguish different teacher profiles in low-quality index schools.

H2.2) There is an association between school-level factors and the probability of a teacher belonging to a specific teacher profile.

H3.1a) Students with an immigration background are significantly more peripheral in their class' teacher-student interaction networks than their non-immigrant peers, with teacher-immigrant-student interactions being less frequent than teacher-non-immigrant student interactions.

H3.1b) This difference will vary by immigrant students' country of origin.

H3.2) There is significant variation in the effect of student immigrant background on the frequency of interactions with the teacher across classrooms, with some classrooms being significantly more inclusive of immigrant students than others.

H3.3) The classroom-level variation on the effect of student immigrant background is partially explained by the immigrant composition of the classroom and the teacher's attitude towards diversity in the classroom.

1.9 Objectives

O1) To learn about the variance of quality in teaching practices within and between under-performing schools.

O2.1) To discover if different teacher profiles can be distinguish in low-quality index schools.

O2.2) To discover if there is an association between school-level factors and the probability of a teacher belonging to a specific teacher profile.

O3.1a) To discover if students with an immigration background are significantly more peripheral in their class' teacher-student interaction networks.

O3.1b) To discover if the difference of student-teacher interactions vary by immigrant students' country of origin

O3.2)To discover if there is significant variation in the effect of student immigrant background on the frequency of interactions with the teacher across classrooms.

O3.3) To discover if the classroom-level variation on the effect of student immigrant background is partially explained by the immigrant composition of the classroom and the teacher's attitude towards diversity in the classroom.

1.10 Results

R1) Only a small part of the variance (4.8%) is explained by school-level factors. This means that there is a huge variation in the quality of teaching practices within under-performing schools. Among under-performing schools, there is very little variation.

R2.1) Through a multilevel LCA, it is possible to distinguish the following teacher profiles: High-quality teaching practices (38.47%), low-quality practices (25.77%), content-oriented (18.88%), and climate-oriented (16.88%).

R2.2) There is a relationship between the “school support for behavioral issues” factor and the teacher categories.

R3.1a) Immigrant background was not found to be a significant predictor of the frequency of teacher-student interactions, especially after controlling for relevant student-level variables.

R3.1b) The situation of immigrant students heavily depends on their country of origin. Students from Peru were consistently more peripheral to their teacher-student networks than their non-immigrant peers.

R3.2) There is a significant variation in the effect of student immigrant background on the frequency of interactions with the teacher across classrooms.

R3.3) The immigrant composition does not explain the classroom-level variation on the effect of student immigrant background. However, it was found that teachers with more negative attitudes approach immigrant students significantly less frequently for pedagogical purposes.

1.11 Thesis structure

The structure of this thesis is based on the research questions, hypotheses, and objectives mentioned above. Figure 1.1 provides a model to demonstrate the connections between the aforementioned components.

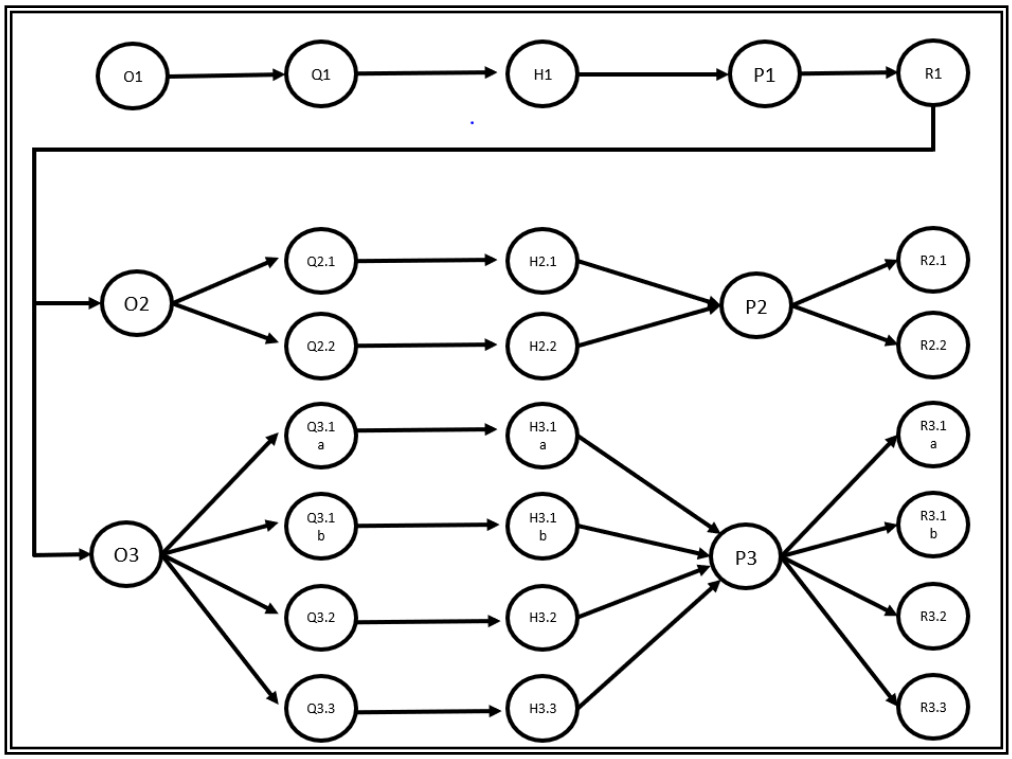

Figure 1.1. Thesis structure

The first objective of the thesis is: O1) To inquire about the variance in the quality of teaching practices within and between under-performing schools. This objective leads to: Q1) How varied is the quality of teaching practices within and among under-performing schools? The hypothesis for the study was the following: H1) The variance of quality teaching practices within schools is greater than between under-performing schools. The results are presented in

P1 (Chapter 2) and summarized in R1) Only a small part of the variance (4.8%) is explained by school-level factors. This means that there is a huge variation in the quality of teaching practices within under-performing schools. Among under-performing schools, there is very little variation.

The results from the first study generated two different objectives. Firstly, O2) To analyze teacher practices through the construction of teacher profiles based on classroom observations, which is disaggregated into O2.1) To discover if different teacher profiles can be distinguished in low-quality index schools, and O2.2) To discover if there is an association between school-level factors and the probability of a teacher belonging to a specific teacher profile. Each objective raises the following research questions: Q2.1) Can different teacher profiles be identified in low-quality index schools? Q2.2) Is there an association between school-level factors and the probability of a teacher belonging to a specific teacher profile? The corresponding hypotheses are: H2.1) It is possible to distinguish different teacher profiles in low-quality index schools, and H2.2) There is an association between school-level factors and the probability of a teacher belonging to a specific teacher profile. The results are presented in Paper 2 (Chapter 3) and are summarized as follows: R2.1) Through a multilevel LCA, it is possible to distinguish the following teacher profiles: Content-oriented, high-quality practices, climate oriented and low-quality practices, and R2.2) There is a relationship between the “school support for behavioral issues” factor and the teacher categories.

The second objective generated from the results of the first study is: O3) To discover whether teacher-student interactions in the classroom depend on the immigrant background of the students, which is disaggregated into: O3.1a) To discover if students with immigration

background are significantly more peripheral in their class' teacher-student interaction networks, O3.1b) To discover if the difference of student-teacher interactions vary by immigrant students' country of origin, O3.2)To discover if there is significant variation in the effect of student immigrant background on the frequency of interactions with the teacher across classrooms and O3.3) To discover if the classroom-level variation on the effect of student immigrant background is partially explained by the immigrant composition of the classroom and the teacher's attitude towards diversity in the classroom. Each of these objectives raised the following corresponding research questions: Q3.1a) Are students with immigration background significantly more peripheral in their class' teacher-student interaction networks?, Q3.1b) Do the interaction networks vary when considering the immigrant students' country of origin?, Q3.2) Is there a significant variation in the effect of student immigrant background on the frequency of interactions with the teacher across classrooms?, Q3.3) Can the immigrant composition of the classroom and the teacher's attitude towards diversity in the classroom partially explain the classroom-level variation on the effect of student immigrant background?. The corresponding hypothesis are the following: H3.1a) Students with an immigration background are significantly more peripheral in their class' teacher-student interaction networks than their non-immigrant peers, with teacher-immigrant-student interactions being less frequent than teacher-non-immigrant student interactions, H3.1b) This difference will vary by immigrant students' country of origin, H3.2) There is significant variation in the effect of student immigrant background on the frequency of interactions with the teacher across classrooms, with some classrooms being significantly more inclusive of immigrant students than others, and H3.3) The classroom-level variation on

the effect of student immigrant background is partially explained by the immigrant composition of the classroom and the teacher's attitude towards diversity in the classroom. Finally, the results, analyzed in Paper 3 (Chapter 4), are the following: R3.1a) Immigrant background was not found to be a significant predictor of the frequency of teacher-student interactions, especially after controlling for relevant student-level variables. R3.1b) The situation of immigrant students heavily depends on their country of origin. Students from Peru were consistently more peripheral to their teacher-student networks than their non-immigrant peers, R3.2) There is a significant variation in the effect of student immigrant background on the frequency of interactions with the teacher across classrooms, R3.3) The immigrant composition does not explain the classroom-level variation on the effect of student immigrant background. However, it was found that teachers with more negative attitudes approach immigrant students significantly less frequently for pedagogical purposes.

1.12 Thesis outline

This thesis is divided into five chapters. The first provides an introduction that describes the context of the study and outlines the research problem, which is supported by a review of the literature. Then, the research questions, hypotheses, objectives, results, and limitations are presented. Each of the following three chapters is a research article that was submitted, published, or will be promptly submitted to refereed journals. The list of articles is as follows:

Chapter 2: Escribano, R., Treviño, E., Nussbaum, Torres D., Carrasco, D (2020). How much does the quality of teaching vary at under-performing schools? Evidence from classroom observations in Chile, *International Journal of Educational Development*. This chapter

discusses a study that analyses the variance of quality of teaching practices within and between under-performing schools in Chile.

Chapter 3: Escribano, R., Mascareño, M., Timmermans, A., Bosker, R., Carrasco, D., Nussbaum, M. (work in progress). Teacher categories in Chilean low-quality index schools and their interaction with school practices: A multilevel latent class analysis. This manuscript will be sent to the journal *School Effectiveness and School Improvement*. This chapter discusses a study that aims to analyze teaching practices through the development of teacher profiles based on classroom observations, and their relation with school factors.

Chapter 4: Ortega, L., Boda, Z., Treviño, E., Arriagada, V., Gelber, D., Escribano, R. The Elephant in the (Class)Room: Analyzing the Inclusion of Immigrant Students within Teacher-Student Interaction Networks, Teaching and Teachers Education (under review). This chapter discusses a study that analyzes teacher-student interaction networks, aiming to discover whether the interactions depend on the immigrant background of the students.

The final chapter consists of a general conclusion, in which the research outcomes and future work are presented.

1.13 Research limitations

This study has limitations that should be considered to avoid the misinterpretation of the results. Firstly, these results cannot be generalized to schools in Chile, or in other countries. This limitation is due to the sample selection, which was not random or stratified. Moreover, the school sample analyzed in Chapters 2 and 3 was selected by the Agency for Quality Education to guide the schools to help improve their results, which means there are

restrictions on academic performance. In the case of the study presented in Chapter 4, the participating schools and teachers were invited to participate as volunteers. This could indicate a willingness to improve and to be more inclusive towards immigrant students.

Secondly, the observation instrument on which Chapters 2 and 3 are based was designed for public policy purposes, without all the technical recommendations that a research tool should have. Though this can lead to less precise data than the data collected based on a research guide, it is important to note that this instrument is aligned with the mandatory protocol that Chilean teachers must follow.

Thirdly, on the modeling of interactions in Chapter 4, there is a limitation on the types of dependencies between students and teachers. Even though the student nesting in classrooms is considered, the dependence of teacher-student probability within the classroom is not modeled. This means it is assumed that each teacher-student interaction is independent of other interactions, which is likely not the case in the classroom, where the networks between students can also influence the probability of interaction with the teacher.

Finally, there is a limitation on the sample size of the specific groups of interest in Chapter 4. Due to the intention of studying minority groups, in this case migrants, the small sample size of students might influence the scope of the study, and consequently might have prevented finding other important associations.

In the concluding chapter, some options are discussed to improve the results with future research.

.

2. HOW MUCH DOES THE QUALITY OF TEACHING VARY AT UNDER-PERFORMING SCHOOLS? EVIDENCE FROM CLASSROOM OBSERVATIONS IN CHILE

2.1. Introduction

Research into school effectiveness has tried to identify which factors are the most important for improving schools. However, this research has mainly focused on effective schools. Furthermore, it is normally assumed that ineffectiveness comes from an absence of the factors that make effective schools effective. However, this fails to take into account the idiosyncrasies of ineffective schools (Reynolds, 2014). The literature suggests that policies that are implemented without considering the specific context are unlikely to succeed (Lupton, 2005; Ladd, 2012). Thrupp and Lupton (2006) highlight the importance of taking context into account. In this sense, the challenges faced by schools in disadvantaged contexts, such as low socioeconomic status, are very different from those found in traditional research. The literature on school improvement tends to recommend promoting certain practices that have been observed in effective or improving schools in different contexts.

In their study of how teaching practices influence academic results, Treviño et al. (2016) show that teaching practices at improving schools do not differ significantly from those found at under-performing schools. These results suggest that such factors are highly sensitive to the specific characteristics of each context. Given this, the strategies used to

improve schools in less-advantaged contexts must be different from those adopted by schools in more favorable contexts. This is because the relevance of certain factors of school effectiveness often varies depending on the context (Muijs et al., 2004). Considering this, the present study will focus on schools from Chile classified as under-performing[1].

Chile has made huge efforts in recent decades to improve the quality of its education system. This led to improved results on national and international tests, such as the SIMCE, PISA, and TERCE tests, at the beginning and end of the 2000s. However, there are still significant learning gaps within the country (Mineduc, 2013a; OECD, 2016; Trviño et al., 2015b). Despite these improvements, the results of the 2015 PISA test suggest that almost 40% of students in Chile are below level two. According to the OECD, this means that they do not meet the minimum level required to develop the skills needed in modern society. Furthermore, only 2% of students in Chile reached the highest possible level on this test. Although the scores in Chile are low across the board, there is still significant inequality in education based on socioeconomic status (OECD, 2016).

[1] This is determined by the quality index established by the Chilean Agency for Quality in Education. This index compares academic results and other indicators of personal and social development with the school's socioeconomic status. Schools are classified as under-performing if the result of this index is lower than expected for their socioeconomic status.

The inequality within the Chilean education system may be explained by social or family-level factors, as well as by variables relating more specifically to the teaching/learning process itself (Martinic et al., 2013). However, the main factor in explaining this inequality is socioeconomic status (Munoz & Dossett, 2014; Treviño et al., 2015b; Hansen & Gustafsson 2018). This shows how socioeconomic differences can easily translate into learning gaps. Furthermore, following socioeconomic status, the next most important factor in academic achievement is the quality of teaching practices (Treviño et al., 2015b; OECD, 2016).

It is widely acknowledged that good teachers can have a significant and positive impact on a student's future (Barber & Mourshed, 2007; Ortega, Malmberg & Sammons, 2017). The quality of teaching practice has also been linked to student performance (Bill & Melinda Gates Foundation, 2013; Youn, 2016). Studying teacher effectiveness is, therefore, important, especially in disadvantaged contexts. If the effectiveness of a teacher improves, the first to benefit are the low-performing students (Nye et al., 2004; Sanders & Rivers, 1996; Torres, 2018). The aim of this study is, therefore, to explore how much the quality of teaching varies among teachers at underperforming schools.

2.2. Literature Review

2.2.1. Studying classrooms

Since the mid-90s, classroom-level variables have become increasingly important in models of school effectiveness. In this sense, one significant contribution has been Creemers' (1994) book

The Effective Classroom, where he proposes a model in which classroom-level results are essential for explaining school performance. Similarly, the classroom has consistently been shown to explain more of the variance in student performance than the school, with the teacher's actions being the biggest factor (Muijs & Reynolds, 2011). Barber & Mourshed (2007) suggest that having controlled for other external factors, what happens in the classroom, is one of the most significant factors in explaining student performance. A similar study by Wenglisky (2002) reveals that the effect of classroom practices, coupled with the characteristics of the teacher, is comparable to the effect of the student's context. This suggests that a teacher's actions have a significant impact on learning.

Broadly speaking, teacher effectiveness is defined as the overall effect of a teacher's observable classroom practices on student learning (Klassen & Tze, 2014). This is most commonly measured by looking at improvements in student learning outcomes (Seidel & Shavelson, 2007). Several studies have linked the scores obtained by teachers on teacher evaluations with student performance (Milanowski, 2004; Alvarado, 2012; Cruz-Aguayo et al., 2017; Taut et al., 2016). However, these types of measurements do not allow us to identify which specific practices lead to improved performance (Kane et al., 2011). Although this is the most widely-held view of teacher effectiveness within the field, there has been a push in recent years to expand the concept by focusing more on classroom processes (Goe, Bell, & Little, 2008) than on student performance.

2.2.2. How to measure the quality of teaching practices? Classroom observations

Classroom observations are one way of studying the teaching process in greater detail. Furthermore, the quality of teaching practices (measured directly through classroom observation) can vary greatly, even when student performance is similar (Blazar et al., 2016). Given this, observations have become increasingly important in both teacher evaluations as well as in research models. In general, classroom observation instruments are developed based on conceptual frameworks that try to define what a good teaching practice looks like (Seidel & Shavelson, 2007). This allows for a more comprehensive view of the teaching process. There are several classroom observation tools that have been validated in different contexts. This includes tools such as CLASS, MQI, and PLATO, among others. There have also been studies that have tried to link classroom observation tools with measures of value-added, such as the Measures of Effective Teaching project (MET). This study discovered certain links between the scores obtained by teachers using different classroom observation tools and student scores on standardized tests (Kane & Staiger, 2012; Kuhfeld, 2017). Kane et al. (2011) also mention that evaluations using well-implemented classroom observation models are an accurate measure of teacher effectiveness and can predict student performance in both mathematics and language arts. In the case of Chile, Taut et al. (2016) suggest that there is a link between measures of teacher effectiveness and scores on the national teacher evaluation system, which is partly based on classroom observation through a teacher portfolio.

Other studies in Chile have also involved classroom observation, both in preschool (Leyva et al., 2015; Treviño et al., 2015; LoCasale-Crouch, 2016), as well as a primary school (Treviño et al., 2016; Bruns, De Gregorio & Taut, 2016). These studies show that the quality of teacher-student interactions is positively linked to student learning.

There are several instruments for measuring the quality of teaching practices. However, the differences in their operationalization suggest that the definition of quality is still on discussion when it comes to teaching practices. Furthermore, Gill et al. (2016) show that although there is a common theme among the different models, there are significant differences in the coverage and the specific practices that are measured in each dimension. Similarly, the rubrics for these models also vary depending on the grade level in which they are applied. For example, different CLASS rubrics are used at different grade levels as the elements that are measured vary in their nature from one educational stage to another. Furthermore, there are other classroom observation tools that focus on specific educational stages and/or subjects, such as ECCERS, PLATO, and MQI, among others. We must, therefore, study in greater detail how the definition of quality in teaching practices is operationalized and applied.

2.2.3. Variance in the quality of teaching practices

De Gregorio & Bruns (2016) analyzed over 15,000 classroom observations in more than 3,000 schools from across six Latin America and Caribbean countries and found that there is huge variance in the amount of time spent on pedagogical activities, both among schools as well as within them. Furthermore, value-added models have revealed that the quality of teachers can vary drastically, even within a school (Hanushek & Rivkin, 2012). Several studies have shown

that students from low-income families tend to have a disproportionate exposure to lower quality teachers (Loeb & Reininger, 2004; Allen & Sims, 2018; Strunk, Marsh & Bruno, 2017). This is similar to the case of Chile, with several studies describing the inequality in the quality of teachers depending on socioeconomic status (Valencia, E., & Taut, S., 2011; Meckes & Bascopé, 2012; Cabezas et al., 2017). These studies show that there is a significant correlation between a student's socioeconomic status and the quality of their teachers. Furthermore, Palmer et al. (2016) suggest that teaching practices also vary greatly depending on the make-up of the classroom, especially in terms of the gender and educational level of the students. However, there is little literature on teaching practices at under-performing schools, both in Chile as well as other countries (Vanlaar et al., 2016).

Given the above, it is particularly important for us to study how teaching practices vary at under-performing schools. Doing so will allow us to understand whether the whole organization should be classified as ineffective or, alternatively, whether there is a mix of outstanding and not-so-outstanding practices. Studying this issue may shed some light on the challenge of improving not just at the individual, teacher level, but also at the organizational, whole-school level. Therefore, studying the variance of teaching practices within and among different schools is a first step towards understanding the level of teaching that can be found at under-performing schools. Consequently, our research question asks: "How varied is the quality of teaching practices within and among under-performing schools?"

As mentioned previously, a second issue to consider is the definition of quality in teaching practices, an issue that is still open for discussion in the literature. Therefore, it is important to

discover whether the quality rating for a teaching practice can be reduced to a single indicator instead of several sub-indicators spread across different dimensions. Doing so would be beneficial as it would facilitate the decision-making process in terms of public policy. Furthermore, as the observation rubrics are differentiated to accommodate different grade levels, this issue will also be explored based on the data that is obtained.

2.3 Methodology

2.3.1 Source of data

The data was obtained from the Chilean Agency for Quality in Education (http://www.agenciaeducacion.cl). This agency is a government body that safeguards the quality of education in Chile through student assessment and school evaluations. Its responsibilities include guiding schools toward educational improvement. In order to do so, the Agency visits schools that are classified as under-performing based on their results and/or other indicators of social and personal development in comparison to other schools of similar socioeconomic status. These visits are carried out by a team of experts who use different tools to gather their information. These tools include a school-wide observation rubric; interviews with the owner/administrator, principal, academic director, pedagogical coach, environment team, teachers, students, and parents/guardians; student, parent and teacher questionnaires; and a classroom observation rubric, which is applied in several classrooms during each visit. The classroom observations aren’t necessarily carried out by the same person each time, as either member of the visiting team is able to conduct an observation. Furthermore, there is a system of rotation in place to attempt that the team for each school inspection is different. By doing so, the

Agency looks to decrease the bias on item scores between school observations. Using the information gathered during these visits, the visitors then analyze the results and rate the school based on the Quality Standards used to guide schools (MINEDUC, 2013b). The Agency then writes a report highlighting the strengths and weaknesses of the school, as well as providing recommendations for improving performance. The information that is gathered during visits is stored in a database, to which researchers can request access. This database is the source of data for the present study.

2.3.2 Sample

This study will analyze the information gathered by the Agency during 260 school visits conducted between September 2014 and April 2016.[2] More specifically, the study will focus on the information gathered during the classroom observations that were carried out by the school visitors. The schools in this sample are mainly classified as under-performing (65.77%), with others classified as medium-low (30.38%) and some classified as average (3.85%). These categories are defined based on the students' academic performance (67% of the score) as well as other indicators of social and personal development (33% of the score). These scores are then adjusted according to the school's context (Agencia de Calidad de la Educación, 2014). In terms of administration type, 80.77% of the schools are public schools, while the other 19.23% are voucher schools. The sample does not include any private schools. Furthermore, only 10.38% of the schools were classified as being rural. In terms of socioeconomic status (SES), 31.54% of the

[2] Although the sample includes schools from every region of Chile, there is no way of ensuring that the sample is representative of all under-performing schools.

sample is low SES students, 56.54% are medium-low SES students, and 11.92% are medium SES students.[3]

An average of 4.87 classrooms was observed (SD = 2.23) at each of the schools that were visited. These classrooms were all between 1st and 8th grade, covering 8 different subjects: mathematics, language arts, natural sciences, history, English as a foreign language, music, art, and physical education. Table 2-1 shows the frequency with which each grade level and subject was observed.[4]

[3] Socioeconomic status is calculated by the Agency using cluster analysis. It is therefore not the same as the traditional classification based on quintiles. In total, there are five levels of socioeconomic status: low, medium-low, medium, medium-high and high. For more details see: http://www.agenciaeducacion.cl/estudios/biblioteca-digital/metodologia/

[4] Other subjects such as religion, guidance and extracurricular activities were not included as they were not observed frequently enough.

Table 2-1. Frequency of observation of classrooms by discipline and grade

	Grade								
Subject	1°	2°	3°	4°	5°	6°	7°	8°	Total
Mathematics	41	41	43	40	38	34	41	52	330
Language	57	47	43	35	28	42	40	51	343
Natural Sciences	20	11	12	33	23	24	21	30	174
History, geography & Social Sciences	7	17	20	22	25	23	34	35	183
Foreign language (English)	4	5	5	3	13	14	22	18	84
Music	3	6	7	7	11	3	6	4	47
Visual Arts	4	6	7	5	7	5	9	7	50
Physical Education	6	9	7	9	7	8	4	5	55
Total	142	142	144	154	152	153	177	202	1266

2.3.3. Variables

2.3.3.1 Indicators

The observations for this sample were carried out using the Ministry of Education's Quality Standards (MINEDUC, 2013b). As mentioned previously, the Quality Standards cover four dimensions: Leadership, Academic Management, Environment, and Resource Management. The aim of this study is to analyze the variance in the quality of teaching practices both among and within schools. Given this, only Academic Management and Environment were considered in our analysis, as these dimensions include indicators of the teachers' classroom practices.[5]

Each dimension comprises different standards, which specify the actions that must be observed in order for a teacher to be classified at a certain level for each of the criteria. The standards were broken down into a series of indicators, taking the different teaching practices as the unit of

[5] These dimensions are essential to our study. The standards that are used to define them draw heavily on the literature on school and instructional effectiveness. For more details on how these dimensions are defined see MINEDUC 2013c.

analysis. Each of these indicators was used as a variable in our study. Table 2-2 includes a description of each of the 21 indicators.

These 21 indicators were grouped into three factors based on common themes. The CLASS framework and the original fields used by the Agency were both taken into account when defining the three factors as Instructional Quality, Emotional & Pedagogical Support, and Classroom Environment. Instructional Quality groups together nine variables that refer to the way in which the teachers design and deliver their instruction.

The second factor, Emotional & Pedagogical Support, groups together eight variables that refer to the way in which the teachers interact with their students in order to provide them with support, beyond simply teaching them. The third factor, Classroom Environment, groups together four variables that refer to the teachers' efforts to establish a suitable classroom environment (Table 2-2).

The indicators used to evaluate the teaching practices are ordinal in nature, with four possible values between 1 and 4. In this case, a score of 1 suggests poor development of the practice in question, a score of 2 suggests initial development, a score of 3 suggests adequate development, and a score of 4 advanced development.

Table 2-2. Description of the 21 criteria observed that are used to measure the quality of teaching practice. To see the criteria in more detail, see appendix D and MINEDUC (2013b)

Indicator	Factors	Summary of criteria
Objective	IQ	Classes aligned with the learning objectives established by the national curriculum
Clarity	IQ	Clarity of the teacher when presenting
Content	IQ	Teachers master the necessary knowledge and skills and teach with conceptual rigor
Dynamism	IQ	The teachers deliver the class in a dynamic way
Introduction	IQ	The teachers explain and contextualize the lesson objective
Presentation	IQ	The teachers present the content and develop the students' skills using effective strategies
Production	IQ	The teachers encourage their students to produce information using effective strategies
Practice	IQ	The teachers encourage their students to practice their skills, providing them with guidance and autonomy
Review	IQ	The teachers do a review at the end of the class
Interest	EPS	The teachers show an interest in and concern for their students
Monitoring	EPS	The teachers constantly monitor their students' comprehension and performance
Feedback	EPS	The teachers give their students feedback on their performance, both individually and as a group
Praise	EPS	The teachers congratulate their students on their achievements, effort and perseverance
Timing	EPS	The teachers manage to spend most of the class time on the teaching/learning process
Dedication	EPS	The teachers manage to keep their students on task during class
Independence	EPS	The teachers encourage their students to work and study independently during class
Support	EPS	The teachers and assistants assist the students who need additional support
Expectations	CE	The teachers act according to the belief that their students can improve their behavior
Courtesy	CE	The teachers encourage and require their students to respect basic rules of common courtesy
Routines	CE	The teachers define set routines for managing certain activities and everyday situations
Opinion	CE	The teachers encourage their students to express their ideas and opinions

Note: IQ = Instructional Quality, EPS = Emotional & Pedagogical Support and CE = Classroom Environment

2.3.3.2. Coding process

As mentioned in the previous section, each teacher's practices are evaluated based on 21 indicators, with each indicator given a score from 1 to 4. In order to score each of the indicators, the written descriptions of the classroom observations were coded. These written descriptions were produced by the Agency's visitors and focus on the actions of both the teacher and the students.

The coding process, therefore, consisted of assigning a score from 1 to 4 based on evidence from the written description. The criteria used to assign these scores are defined in the Quality Standards, which specify the actions that are required by the teacher in order to be classified as poor (1), developing (2), adequate (3), or advanced (4). For more information on the rubric that was used, see Appendix E. Appendix D also shows which standard each variable came from.

The coding of the 21 indicators for the 1266 classroom observations was carried out by a primary school teacher. In order to ensure that the process was as rigorous as possible, the coding process was then performed a second time with a random sample of 20% of the observations. This second round of coding was carried out by the lead researcher in this study, and the results were used to calculate the inter-reliability index. This index is a measure of the degree to which two or more raters agree on the coding of a set of data. In our case, we calculate the degree of agreement between the two raters for each of the 21 indicators. We also calculate the interclass correlation coefficient (ICC) (McGraw & Wong, 1996), which is equivalent to the Weighted Kappa (Cohen, 1968; Fleiss & Cohen, 1973). The ICC was calculated using the *ICC* function with the *irr* package in R. Following the recommendations by Hallgren (2012), a two-way model

was specified in the function. This model takes into account any possible systematic deviation that may come from the specific group of raters used in this study. An absolute agreement was also used as it was important that the raters agreed on the same score for each indicator. Finally, the unit of analysis was specified as single as only a sub-sample of the study was coded by both raters, and the teacher's coding, therefore, had to be generalized.

The results of the ICC for each variable can be found in appendix A. The average for the ICC was 0.8 (SD=0.11), while the ICC for each of the variables was greater than 0.6. This means that there was a good or excellent level of agreement between the two raters (Cicchetti, 1994). These high levels of inter-rater reliability suggest that any measurement error introduced by the raters was small

2.3.4 Analysis

A three-stage analysis was conducted to evaluate the homogeneity of the quality of teaching. The first stage involved confirmatory factor analysis comparing different models. The distribution of variance between the general factor and the specific factors was then analyzed for the teaching quality index that was obtained from the factor analysis. Finally, the third stage involved calculating the distribution of variance for the quality of teaching.

Confirmatory factor analysis was carried out with the 21 indicators in order to confirm the dimensionality of the data. As highlighted previously, studying the dimensionality is important for determining whether the construct covers the proposed dimensions.

The analysis was conducted using a restricted bifactor model. In a restricted bifactor model, each of the items loads on the general factor, as well as on at least one additional factor, which must be orthogonal to the general factor (Reise, S. P., Moore, T. M., & Haviland, M. G., 2010). In this sense, Reise et al. (2010) suggest that, unlike the classic model, the restricted bifactor model has the advantage of being able to measure a general latent factor while also controlling for the variance that may come from other common factors. In some cases, these common factors might be sub-dimensions, while in others, they may just be measurement noise.

Estimates for one-factor and three-factor confirmatory factor analysis were used as alternatives. In both the bifactor and the three-factor confirmatory models, the three factors were informed from the same indicators and included the same number of items (Table 2-2).

This analysis was performed using Mplus v7. The exact same sample and WLSMV estimator were used in all three models. In this case, the WLSMV is considered appropriate for ordinal data (Li, 2016). The data hierarchy was designed using the CLUSTER command, with the classroom observations grouped by the school. For the alternative three-dimension model, the correlations between the different factors were estimated, while for the bifactor analysis these were set at 0 so as to meet the requirement for orthogonality (Reise, S. P., Moore, T. M., & Haviland, M. G., 2010).

Following this, the distribution of the variance between the general factor and the common factors in the bifactor model was analyzed. This allowed us to calculate an overall index for teaching practices should most of the variance be explained by the general factor (Rodriguez et al., 2016). Alternatively, in the case of a multidimensional model, we could calculate one index

for each dimension. In order to do so, the *Explained Common Variance (ECV_{gen})* (Reise, Moore, & Haviland, 2010) was calculated. This estimates the proportion of common variance in the variables that can be attributed to the general factor. The Percent of Uncontaminated Correlations (PUC) (Hammer, J. H., 2016a) was also calculated, as were the Omega (ω) and Omega Hierarchical (ωH) coefficients. The ω coefficient estimates the proportion of variance in the variables that can be explained by both the general factor as well as the specific factors (Reise et al., 2013). The ωH coefficient, on the other hand, refers to the total proportion of variance that can be attributed to the general factor, having first taken into account each of the specific factors (Zinbarg, 2005). Finally, the average relative parameter bias (ARPB) (Hammer, J.H., 2016b) was calculated, comparing the factorial loads of the specific factors with the factorial loads of the general factor in the bifactor model.

Once the dimensionality and distribution of variance had been proven, a quality index for teaching practices was then calculated. This index allowed for a comparison between classrooms. This calculation was performed using the *fscore* function in *Mplus Version 7*. The scores were calculated for both the overall sample of all classroom observations, as well as for each cycle of primary education (the first cycle, from 1st to 4th grade, and the second cycle, from 5th to 8th grade).

Finally, the *intraclass correlation index* (*ICC*) was calculated using an unconditional model so as to explore the variance among and within schools. For this study, the ICC (ρ) quantified the degree to which the quality index for teaching practices varied among schools. It is defined as

the proportion of variance of the *outcome variable* that is explained by the structure used to group together the observations and calculated as follows:

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \qquad \text{Eq. (1)}$$

Where σ_b^2 is the variance of the level 2 (school-level) residuals, and σ_w^2 is the variance of the level 1 (classroom-level) residuals. The ICC was calculated for the overall sample and each educational stage using R, with the *icest* function from the "icc" package.

2.4 Results

2.4.1 Confirmatory Factor Analysis and Bifactor Analysis

In the first stage of analysis, we will examine the goodness of fit of the different models: one-factor Confirmatory Factor Analysis (CFA), three-factor CFA, and three-factor bifactor analysis. We will also measure the goodness of fit of each of these models for the different cycles of primary education. The correlations between variables can be found in appendix B. The goodness-of-fit indices can be found in Table 2-3. Schreiber et al. (2006) suggest that for an adequate fit in models using categorical data they must meet the following conditions: RMSEA < 0.06, CFI > 0.95, TLI > 0.96 and WRMR < 0.90. In the case of the WRMR index, DiStefano et al. (2017) suggest that it tends to perform poorly with large sample sizes and models with small specification errors. In this case, they even propose increasing the conditional limit.

The one-factor model does not fit the data in any of the cases, i.e. for the overall sample or either cycle (Table 2-3). The three-factor model does not fit the overall sample or the sample for the second cycle. The fit for the first cycle in the three-factor model is acceptable. However, it is

worse than fit for the bifactor model, both for the overall sample as well as the sample for the first cycle.

The bifactor model shows the best fit for both the overall sample as well as the sample for each cycle. In this case, the fit is acceptable. The loads (λ_{gen}) associated with the general factor for the items in this model are $\lambda_{gen} > 0.5$, with the exception of the item for alignment between the lesson objective and the curriculum. The λ for each of the factors can be seen in Figure 2-1. Loads of these factors are noticeably smaller than the λ_{gen}, as they capture the variance of items not considered in the general factor for teaching quality.

Table 2-3. The goodness of fit for the 1 Factor model, 3 Factors model and the Bifactor model with 3 factors.

Model		X^2 [df]	RMSEA [90% C.I.]	CFI	TLI	WRMR
1 Factor						
	All grades	2086,035** *[189]*	0,089 *[0,086-0,093]*	0,954	0,949	2,711
	Cycle 1	770,777** *[189]*	0,073 *[0,067-0,078]*	0,972	0,969	1,694
	Cycle 2	1296,593** *[189]*	0,093 *[0,088-0,097]*	0,944	0,938	2,302
3 Factors						
	All grades	1237,414** *[186]*	0,067 *[0,063-0,070]*	0,974	0,971	2,049
	Cycle 1	542,778** *[186]*	0,055 *[0,049-0,061]*	0,984	0,982	1,299
	Cycle 2	921,938** *[186]*	0,076 *[0,071-0,081]*	0,963	0,958	1,856
Bifactor with 3 factors						
	All grades	607,151** *[168]*	0,045 *[0,042-0,049]*	0,989	0,987	1,287
	Cycle 1	288,734** *[168]*	0,035 *[0,028-0,042]*	0,994	0,993	0,844
	Cycle 2	562,015** *[168]*	0,059 *[0,053-0,064]*	0,980	0,975	1,306

Note: ** p<.001

The bifactor model for the first cycle is a good fit for the data, while it is also an acceptable fit for the second cycle (Schreiber, J. B et al., 2006). This is with the exception of the WRMR index, which is greater than 1 (DiStefano et al., 2017). The loads associated with the general factor for the items in both models are $\lambda_{gen} > 0.5$, except for the item for alignment between the lesson objective and the curriculum for the second cycle. The loads for the bifactor models for each cycle can be seen in the figures included in appendix C.

This shows that the rubric used to assess the quality of teaching is a good fit for the first cycle of primary school, i.e., 1st to 4th grade. However, this is not the case for the second cycle, i.e. 5th to 8th grade.

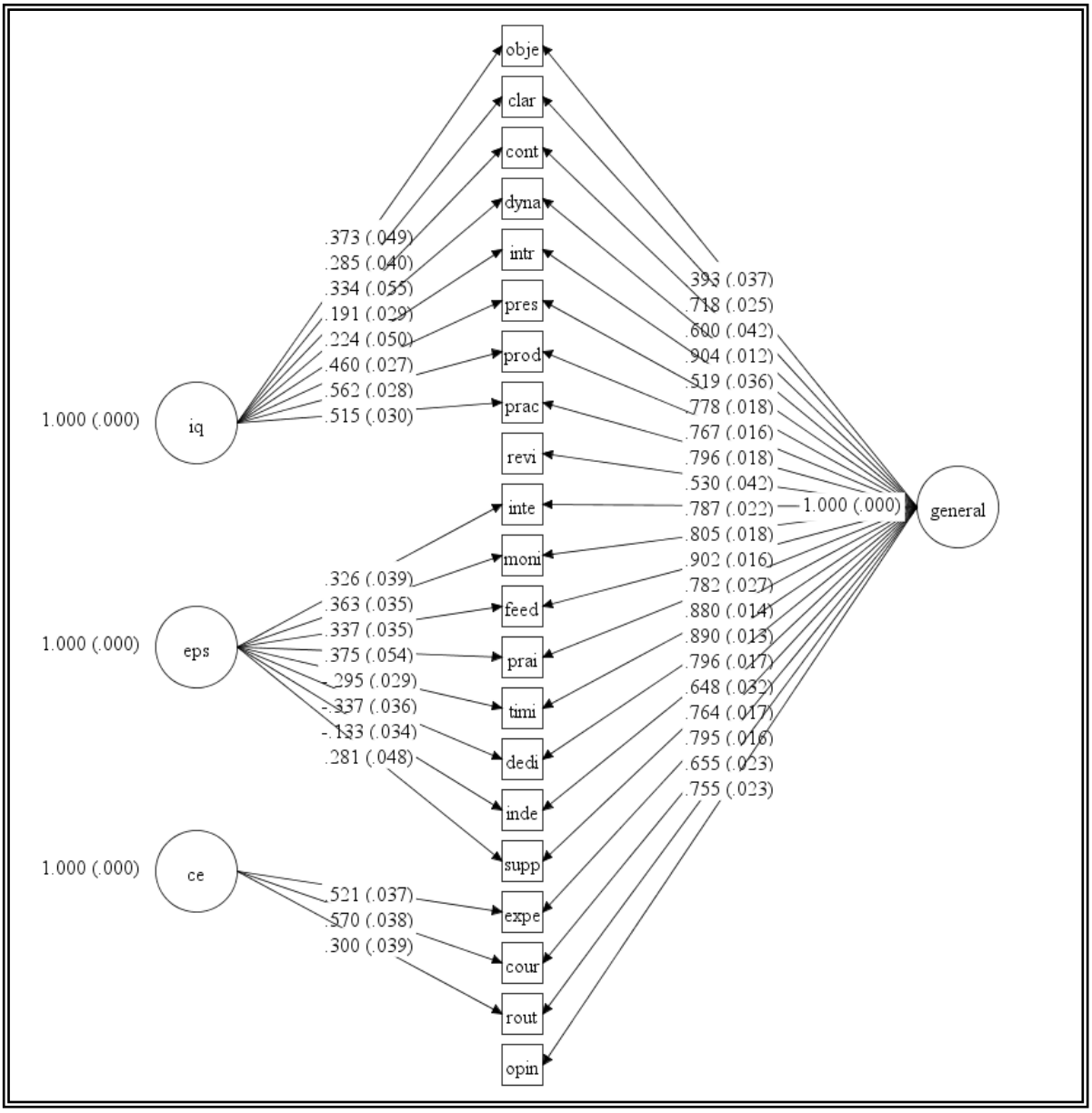

Figure 2-1. Loadings for the bifactor model, with the overall sample

Note: the abbreviations of the indicators correspond to the first four letters of the indicator names presented in table 2-2. Only significant cows are shown (p <.05)

2.4.2. Distribution of variance between the general factor and common factors in the bifactor model

In the previous section, we confirmed the dimensionality of teaching quality, with a bifactor structure giving the best fit, i.e., a structure with a general dimension and three common dimensions. However, it is also important to evaluate the substantive dimensionality. This means assessing whether the information provided by the general factor is enough to justify the calculation of an overall index for teaching practices or, alternatively, whether a separate index is required for each common dimension (Table2-4).

Table 2-4: Variance distribution measures for the general factor and the specific factors

	ECV	ω	ω_H	$\frac{\omega_H}{\omega}$	ARPB
All grades	0,814	0,978	0,937	0,96	0,056
First Cycle	0,831	0,981	0,946	0,96	0,04
Second cycle	0,786	0,979	0,927	0,95	0,073

Note: The ARPB statistic is a measure of comparison between factorial loads of the one-dimensional model and the general factor of the bifactor model.

Firstly, the Explained Common Variance (ECV) suggests that 81.4% of the variance is explained by the general factor for the model, including all grade levels. In the case of the models for the first and second cycles, the percentage of variance explained by the general factor is 83.1% and 78.6%, respectively. According to Rodriguez (2016) and Quin (2014), when ECV < .90, other indices must be considered in order to ensure unidimensionality.

For the model, including all grade levels, the ω and ω_H coefficients are .978 and .937, respectively, with the ratio $\frac{\omega_H}{\omega}$ = 0.96. This suggests that 96% of the explained variance can be attributed to the general factor for teaching quality. This is much higher than the level required

for treating data as coming from a unidimensional model (Reise et al., 2013). This is also the case for the models for the first and second cycles, with 96% and 95% of the explained variance attributable to the general factor.

The *Average Relative Parameter Bias* was also calculated. This index compares the differences between the factorial loads in the unidimensional model with the loads on the general factor in the bifactor model. The values obtained were $ARPB_{\text{All grades}}$ = 5.6%, $ARPB_{\text{First Cycle}}$ = 4.0% and $ARPB_{\text{Second Cycle}}$ = 7.3%. All of these values are within the acceptable limit (<10%), according to Rodriguez et al. (2016).

The final indicator to evaluate is the PUC, which, for all three models, was 0.67. The minimum recommended value for PUC is > .80. However, Reise et al. (2013) suggest that when PUC < .80, ECV_{gen} > .60 and ω_H > .70, the multidimensionality is not strong enough to stop the data from being treated as unidimensional (Reise, Scheines, Widaman, & Haviland, 2013b; Hammer, 2016a).

The analysis of the explained variance for the different factors reveals that the data from the classroom observations are an acceptable fit for the bifactor model. With this model, the general factor (i.e., the quality of teaching practices) can be considered a substantively unidimensional latent variable. Any remaining analyses were therefore conducted using only this dimension. In the case of the other three common factors (Instruction, Support, and Environment), these may be considered nuisance dimensions, which we assume comes from the original structure used to define the indicators for this study.

According to Zhang, Z. & Yuan, K. H. (2016), a robust Cronbach's alpha can be applied for samples with missing data or non-normal distribution. In the case of our study, the value for the sample with all grades is $\alpha_{robust\ ALL} = 0.938$, with a confidence interval between 0.934 and 0.941. The value of the samples for each cycle is $\alpha_{robust\ C1} = 0.937$ [C.I. = 0.869-0.966] for the first cycle and $\alpha_{robust\ C2} = 0.947$ [C.I. =0.879-203973] for the second cycle.[6]

2.4.3. Distribution of teaching quality at under-performing school

In this section, we first look at the average score for teaching quality using the general factor described in the previous section. This tells us the level of teaching practices included in the sample. Secondly, we then study the distribution of the variance between the general factor and the common factor of the instrument described in the previous section. The analyses are carried out by only taking into consideration the scores obtained for teaching quality for the general factor.

The analysis of variance suggests that only 4.8% of the variance in the quality of teaching practices for the overall sample is explained by the teacher's school (Table 2-5). In the case of the models for the first and second cycles, this percentage increases to 13.4% and 7.5%, respectively.

This result shows that most of the variance is not found among schools. Instead, it is found among teachers at the same school. In other words, any school-level factor that may explain the

[6] Parallel Analysis (Timmerman & Lorenzo-Seva, 2011) was also used to confirm the dimensionality. The results of this analysis can be found in appendix F.

quality of teaching practice within an establishment will only be associated with a minimum amount of the variability in the results.

Table 2-5: Intraclass correlation index for the complete sample and for each teaching cycle.

	All grades	**First cycle**	**Second Cycle**
ICC *[95% C.I.]*	0,048 *[0,007-0,097]*	0,134 *[0,041-0,231]*	0,075 *[-0,007 - 0,162]*
VARIANCE COMPONENTS			
Within variance	0,656	0,586	0,600
Between variance	0,033	0,091	0,048

2.5 Discussion

The main finding of this study is that, at under-performing schools, the teachers' membership in their schools only explains 4.8% of the variance in the quality of teaching practices. This suggests that most of the variance is not found among schools. Instead, it is found among teachers at the same school. The lack of variance explained by school-level factors in our study is in line with findings from other studies. In this sense, Mansfield (2015) shows that there is little difference in the distribution of teaching quality across schools of different socioeconomic status and that, in general, all students have both good and bad teachers. In their study of teacher mobility, Feng & Sass (2017) conclude that the worst teachers tend to get together with the worst teachers, while the best get together with the best. There is also constant movement within these two extremes. Feng & Sass' (2017) findings help us understand the dynamic of highly-stratified education systems. Our study focused mainly on the two lower socioeconomic groups and contributed to the literature with a description of these groups in a limited time period. Bruns & Luque (2015) suggest that there is considerable variance both among and within schools for a representative sample from Latin America and the Caribbean. Our study also reaffirms this

suggestion that there is huge variation in the quality of teachers. However, in our case, the level of variation is very low when comparing under-performing schools. This may be due to the fact that we mainly focused on low SES schools, as the education system in Chile is highly stratified and the results are heavily linked to socioeconomic status. This discrepancy highlights the importance of conducting studies in specific contexts, as the results vary depending on the population that is studied.

Furthermore, comparing the level of variance among and within schools seems to suggest that the school has little influence on the quality of its teachers' practices. The influence of a school on its teachers' practices has been studied previously, demonstrating that schools that promote collaborative teaching practices and academic coordination help improve their teachers' instructional practice (Kraft et al., 2015). In his book *Schools that Learn* (2012), Peter Senge suggests that having a shared mission is key to educational improvement. In order to achieve this, he recommends that collaboration and shared meaning among all members of the school community should be heavily promoted. These recommendations are similar to those provided by Fullan & Rincón-Gallardo (2016), who propose that a school system should focus on a small number of objectives and invest in the transfer capacity between peers. Jackson & Bruegmann (2009) find that teachers, and especially newly qualified teachers, benefit from collaborating with more experienced peers. Similarly, Ronfeldt et al. (2015) show that collaboration among peers may be a predictor of how quickly teachers improve their practices. A culture of learning is also important for substantial change to take place in schools, while school leaders can encourage a culture of peer-to-peer learning (Haiyan, Walker & Xiaowei, 2016). In the case of

Chile, and specifically in a low SES context, schools seem to have little influence on their teachers' practices as a whole. From Senge's (2012) perspective, the schools, in this case, are failing to work as an organization that develops its teachers' abilities, an essential component of school improvement. Hargreaves & Fullan (2012) propose that improving teaching at school level requires a systemic approach that promotes collaborative work among teachers for planning, executing, and assessing teaching practices, as well as spreading the practices of the most effective teachers within the school. This means that having effective teachers in a school—as in this study shows—should be seen as an advantage to start improving teaching through learning communities, but this requires a decided approach from school and system-wide leaders.

This is an important finding for public policy, especially in the area of continuing professional development, as it shows that quality teaching practices can be found at all types of schools. Consequently, collaborative work among teachers should be further promoted. In this sense, if there are teachers at a school who meet the necessary standards, public policy should focus on developing the skills of others by building on this (Haiyan, Walker & Xiaowei, 2016). Furthermore, given the level of variation that can be found in the quality of teaching practices within a school, it is not enough to judge the quality of a school-based on the observation of a single classroom. Similarly, any classroom observation that looks to assess the quality of a teacher's practices is more useful for surveying the quality of teaching on a national level than on a school level. These recommendations are in line with the findings by Darling-Hammond

(2010), who suggests that countries with good education systems promote collaboration among teachers at an institutional level.

One thing to come from our analysis is that the model does not fit the data as well in the second cycle of primary school (5th to 8th grade) as it does the first cycle (1st to 4th grade). This is important as the same observation rubric was used in the same way for all eight grade levels, with the same indicators used to provide the schools with feedback, regardless of the grade level that was observed. However, based on our analysis, the group of indicators used to evaluate teaching practice is more appropriate for use in the first cycle than the second. Some observation models, such as CLASS or the Classroom Observation System (Pianta, La Paro & Hamre, 2008; NICHD ECCRN, 2002; NICHD ECCRN, 2004), provide different rubrics for different grade levels. Observation rubrics that look to measure the quality of instruction use different criteria for different grade levels as the teaching strategies, type of support and classroom management techniques also differ (Pianta & Hamre, 2009). Meanwhile, other instruments, such as the Stallings classroom snapshot (Stallings, 1977; Stallings and Mohlman, 1988), have been validated in several different contexts using the same indicators. However, the reason that they work is that they are not intended for measuring quality. Instead, they are used to measure certain observable aspects of a classroom that can be easily quantified (Bruns & Luque, 2015). Given that the rubric used by the Agency is intended to measure quality, it is reasonable to expect that its level of effectiveness will not be the same for all grade levels. Furthermore, the difference in fit between the first and second cycles may be due to the fact that in Chile the same teacher teaches most subjects in 1st to 4th grade, while this is not the case from 5th to 8th grade. In other

words, teachers in the first cycle spend more time with the same group of students and, therefore, are more likely to know them well and to have established a routine. In addition to this, the children in this case are more likely to have adapted to their teacher's teaching style.

2.6 Conclusions

The present study posed the question "How varied is the quality of teaching practices within and among under-performing schools?" To answer this question, we measured the goodness-of-fit of a confirmatory bifactor model for measuring the quality of teaching practices. We then analyzed the distribution of variance between the general factor and the common factors in the bifactor model. This measured the construct of teaching quality for the rubric that was used in the study. Finally, the intraclass correlation index was calculated using an unconditional model to measure the proportion of variance in teaching quality both among and within schools. The results suggest that only a small part of the variance (4.8%) is explained by school-level factors. In other words, the answer to our research question is that there is huge variation in the quality of teaching practices within under-performing schools. Among under-performing schools, on the other hand, there is very little variation.

These findings are important for shaping public policy in Chile as they suggest that, in general, high quality teaching practices can be found at under-performing schools. However, we can also infer that when it comes to classroom practices, schools do not work in a coordinated fashion. This study provides evidence from a context that has not been as widely studied as effective schools (Reynolds et al., 2014; Muijs et al., 2004), demonstrating important differences with the traditional literature on the subject. Given the difficulty of directly transferring findings from one

context to another (Muijs et al., 2004), exploring these differences is key. The results of this study can be used for designing public policy relating to continuing professional development for teachers as it has been demonstrated that schools that work in a coordinated fashion can improve their results (Kraft et al., 2015).

One of the limitations of this study is that the observation rubric on which the results are based was not designed for use as a research tool; rather as a tool for providing schools with guidance. The psychometric aspects of this rubric could therefore be explored in future work, or indeed the study could be repeated using a previously validated rubric in order to confirm the findings. Future research could also look at exploring which of the factors in this study are associated with student achievement such as SIMCE scores (i.e. scores on a national standardized test). Furthermore, in order to better understand the results of this study, it would be important to look at the profile of the teachers and how these are distributed across the schools.In addition to this, it would also be interesting to look for any school-level factor that can explain this variance in teacher quality across schools, no matter how small.

2.7 Acknowledgements and Funding

2.7.1 Acknowledgements

This research was made possible thanks to the collaboration of the Chilean Agency for Quality in Education. The results, discussion and conclusions are the sole responsibility of the authors and do not in any way represent the opinion of the Agency.

2.7.2. Funding

This project was funded by the National Commission of Scientific and Technological Research of Chile, through the project Center of Advanced Studies on Educational Justice [**Proyecto CONICYT PIA CIE160007**]. Also, this work was funded by the CONICYT PFCHA/BECA DOCTORADO NACIONAL/2016 - 21160063

3. TEACHER CATEGORIES IN CHILEAN LOW-QUALITY INDEX SCHOOL AND THEIR INTERACTION WITH SCHOOL PRACTICES: A MULTILEVEL LATENT CLASS ANALYSIS

3.1 Introduction

Schools are the most basic organizational units in school systems and the focus of both policy and research. In the school efficacy framework, studies on schools focus on understanding how school characteristics are related to student outcomes. Teachers, on the other hand, are usually understood as the main providers of education to students through classroom interactions. Teachers are also studied in terms of their efficacy in promoting student outcomes. The fields of school and teacher effectiveness have provided valuable insights for understanding how to improve educational opportunities for children. However, these fields would benefit from complementary approaches that contribute to a better understanding of how teaching quality may vary, even within the same school. In this way, it would be possible to gain a better understanding of how schools are equipped to offer better educational opportunities to their students.

This study offers a different but complementary perspective to effectiveness for evaluating schools and teachers by using a person-centered approach. Effectiveness studies follow a variable-oriented approach that uses regressions to identify the main predictors of student outcomes, assuming that the population is homogeneous. Person-centered approaches, for example the latent class analysis (LCA) used here, look to understand how individuals are grouped into categories according to different patterns of survey answers or observed behaviors. This approach assumes that the population is heterogeneous in relation to the variables of interest, and focuses on understanding individual or group differences (Reichert, 2016).

Following a person-centered approach, this study has three main aims. First, it seeks to understand how teachers are classified according to the quality of their classroom interactions, as measured by an observation protocol. Second, the study analyzes how different categories of teachers are distributed into different categories of low-achieving schools. Finally, it analyzes the school variables that explain differences in the composition of schools in terms of the teaching quality categories.

Firstly, a literature review of variable and person-centered approaches is presented. This includes a mention of the school and teacher level practices identified as effective in the theoretical framework of school and teacher effectiveness research. Secondly, the sample and variables selection is described, as well as the statistical latent class analysis carried out in this study. Thirdly, the results of the multilevel latent class analysis at teacher and school level categories are presented, as well as the school covariates associated with different teacher categories. Fourthly, the research results are discussed in light of the existing literature. Finally, some limitations and future work opportunities are discussed.

3.1.1 Literature review

School and teacher effectiveness has been a cornerstone of educational research around the world. On the one hand, school effectiveness has identified educational quality as a school attribute. For example, in research on both school effectiveness (Bellei, Muñoz, Pérez, & Raczyinski, 2003; Goldstein & Woodhouse, 2000; Levin, 2006) and school choice (Chubb & Moe, 1988; Davis, 2014; Friedman, 1962; Gallego & Hernando, 2010; Hernández & Raczynski, 2015; Hsieh & Urquiola, 2003), it is implicitly assumed that the quality of education is a school-level feature.

The literature on school effectiveness has consistently found that there are key school attributes related to student outcomes. School attributes such as the role of the principal's leadership and expectations, shared pedagogical guidelines, collaborative school environments, good climate and parental participation, as well as the availability of basic infrastructure and resources, are commonly linked to student outcomes (Bellei, Muñoz, Pérez, & Raczynski, 2004; Bellei, Valenzuela, Vanni, & Contreras, 2014; Bryk, Bender-Sebring, Allensworth, Luppescu, & Easton, 2010; J Scheerens, 2000; Treviño et al., 2015).

The field of teacher or instructional effectiveness, on the other hand, has focused on the attributes of teachers that relate to higher levels of student achievement. For example, within the field of economics in education, there have been efforts to identify effective teachers based on value-added measures of student test scores, without necessarily focusing on observable behaviors related to the teaching process, arriving at seemingly tautological conclusions that high-quality teachers are those with the highest value-added levels (Everson et al., 2013; Rivkin, Hanushek, & Kain, 2005). The field of teacher effectiveness has also focused on studying generic characteristics of teachers and classrooms that are related to student outcomes and, more recently, has paid attention to differential teacher effectiveness (Muijs et al., 2014). This situation is well exemplified by research concentrated on understanding differential teacher effectiveness across subjects (Charalambous, Kyriakides, Kyriakides, & Tsangaridou, 2019). Also, in the US, the Measurement Effective Teaching (MET) project, probably one of the best-funded initiatives to focus on teaching, is aimed at studying how observed teacher practices are related to student achievement gains (Kane, MCcaffrey, Miller, & Staiger, 2013).

A review by Muijs et al. (2014) mentions the most relevant factors found in the area of teacher effectiveness. These are; learning opportunities, instruction and interaction, school climate, and teacher expectations. John Hattie (2012), in his book Visible Learning for Teachers, points out various factors and their corresponding effect on student achievement, according to 800 meta-analyses performed by the author (Hattie, 2008). Some of the factors that have a strong influence (effect greater than 0.6) are high teacher expectations for the performance of their students, the student-teacher relationship, feedback, the credibility of the teacher according to the students, and the implementation of teacher evaluations. Also, in the third regional comparative and explanatory study or TERCE (Treviño et al., 2015), the factors associated with academic performance in Chile are discussed, including classroom climate, availability of resources in the classroom, teachers' effective use of time and school infrastructure, among others.

As can be concluded from the above, research on both school and teacher effectiveness has followed a variable-centered approach which has offered key insights into the factors that relate to student outcomes. This research perspective can be complemented with a person-centered approach that is used in this study to understand how teacher quality varies across classrooms within the same school. In this way, it is able to offer insights to better understand schools as organizations with heterogeneous teaching capacities among their teachers. Furthermore, it also helps us to understand how schools are composed, identifying strengths and weaknesses within the schools' teaching force as a diagnosis for embracing improvement initiatives. This approach also allows us to offer recommendations for systemic changes in schools that take into account the differential teaching capacities within the school.

Although the person-centered approach has been widely used in areas such as psychology and medicine, its use is not yet common in the education area. Some recent studies have used latent class analysis to explore typologies of teachers or students, but the variables used as class indicators are generally sociodemographic measurements, or self-reports stemming from surveys (e.g., Boutin-Martinez et al., 2019; Fagginger Auer et al., 2016; Kim et al., 2018; Nasiopoulou et al., 2019).

To our knowledge, no research systematically analyzes both the variability of teaching quality within schools based on observation of teacher practices, and the school factors that may explain the teaching quality composition within each school. Given that there is scant research on the factors that explain the variation of teacher capabilities within a school, this study analyzes the factors used in school effectiveness research to determine whether student outcomes are related to the variability of teacher quality.

3.1.2 The study

As discussed above, the aim of the study is to fill the gap in the literature to understand how teacher quality varies within the same school, and which factors are related to this variation. Specifically, the study focuses on a sample of low-quality index schools in Chile. The school quality index is mainly defined by the performance of students in national standardized tests, combined with other school factors like climate, healthy lifestyle, academic motivation and self-esteem, and civic engagement, among others. Schools are defined as low-quality schools if their performance is lower than expected for their socioeconomic level and most of these schools teach students from the lowest socioeconomic levels. This study uses a person-centered approach to answer the following research questions:

1. Is it possible to distinguish different teaching quality profiles within low-achieving schools?

2. Is there an association between school-level factors and the probability of a teacher belonging to a specific teacher profile?

3.2 Method

3.2.1 Sample

The study used a sample of Chilean primary schools defined as low-quality by the national Agency for Quality Education (http://agenciaeducacion.cl), which are schools with academic or personal and social development indicators below what is expected given the socioeconomic level of their student population. This Agency is a public entity that was created to ensure educational quality in all Chilean schools. To accomplish its objective, the Agency measures student learning as an indicator for school performance through a standardized test known as the System for the Measurement of Educational Quality (SIMCE, by its Spanish acronym). Also, the Agency visits schools in the lowest range of the quality index. During these visits, an evaluation team composed of three educational experts uses different instruments to collect information through interviews and surveys with diverse agents in the schools (principal, parents, students, teachers, and others), as well as classroom observations. The lessons are encoded using an observation protocol based on an education policy guideline document called Educational Performance Standards (EPS) developed by the Ministry of Education (Ministry of Education, 2013). The EPS describes four levels of quality for school and teacher practices, ranging from

(1) weak to (4) advanced. The information collected during these visits constitutes the data of this study.[7]

Table 3-1

Frequency of classroom observations by grade and subject

	Subject					
Grade	Math	Language	Natural Sciences	Social Sciences	Other*	**Total**
1°	41	57	20	7	17	142
2°	41	47	11	17	26	142
3°	43	43	12	20	26	144
4°	40	35	33	22	24	154
5°	38	28	23	25	38	152
6°	34	42	24	23	30	153
7°	41	40	21	34	41	177
8°	52	51	30	35	34	202
Total	330	343	174	183	236	1266

* Other subjects observed are foreign language (English), music, visual arts, and physical education.

The teaching practices analyzed in this research comprise 1,266 teachers working in 258 schools visited between 2014 and 2016. The average number of observed classrooms per school is 4.87 (SD = 2), with grade level and discipline varying across the observed lessons (Table 3-1). Most of these schools (88%) serve students with low or medium-low socioeconomic status (SES),

[7] This article is part of a doctoral thesis. For this reason, the sample used in this study is the same as the sample of the unpublished paper *How much does the quality of teaching vary at under-performing schools? Evidence from classroom observations in Chile*, which is currently in press. Even though the sample is the same, the study purpose is different.

which is important contextual data to consider in the results and scope of this work. The Agency estimates the SES of the students through a cluster analysis, obtaining five unbalanced socioeconomic groups that are different from the traditional quintile method (Agency for Quality Education, 2013). Some characteristics of groups with low and medium-low SES are the incomplete schooling of parents (<12 years) and family income equal to or less than the national minimum wage. Regarding the dependency type, 80.77% of the schools in the sample are public, and the rest are private-subsidized.

3.2.2 Indicators and variables

3.2.2.1. Teacher level variables

During the Agency visits, the evaluation teams observe teacher practices included in the EPS document, which considers four dimensions: Leadership, Pedagogical Management, School Environment, and Resources Management. Each of these dimensions has three sub-dimensions. As our aim is to study teacher quality, teacher-level variables were selected from the dimensions of pedagogical management and school environment. Within these two dimensions, there are observed practices that refer to teacher actions in the classroom, classroom infrastructure or observations of teachers' notes in the classroom. Based on these two dimensions, we selected observed practices related to teacher actions that are effective in promoting student learning according to the literature (Hattie, 2008; Muijs et al., 2014). These practices are related to feedback and learning monitoring, the structure of instruction, clarity of presentation, classroom management, and classroom environment. These practices were operationalized in the EPS in 15 observed indicators, with each of them comprised of four ordinal categories: insufficient, weak, satisfactory, or advanced (see Table 3-2).

To obtain teacher values for the observed indicators, it was necessary to code each indicator to the corresponding level of quality of practice. The research team has raw classroom observation material collected by the Agency, so the coding process was conducted for the teacher level. This data was collected by the Agency using a structured observation instrument (Foster, 1996; O´Leary, 2013). Based on this information, and using the classification criteria from the rubric found in the EPS-document, it was possible to obtain the teacher values for each indicator. The EPS-rubrics have four levels of performance associated with each indicator: weak (1), incipient (2), satisfactory (3), or advanced (4). In the sample used for this study, level 4 (advanced) was recoded to the third level (satisfactory) due to its infrequent occurrence. For an example of an indicator rubric, see Appendix E[8].

Table 3-2
Variables used as teacher-level indicators

Variable name	Summary criteria
Clarity	The teacher presents the contents in a clear manner
Dynamism	Teacher conducts lessons dynamically
Exposition	Teachers presents content and develops skills through effective strategies
Elaboration	Teachers promotes student learning through effective strategies
Practice	Teachers allows students to practice skills in a guided and autonomous manner
Interest	Teachers shows interest and concern for students
Monitoring	Teachers continuously monitors students' understanding and performance
Feedback	Teachers provide feedback to students on their performance individually and in groups
Time management	Teachers shows good classroom management, using most of the lesson time in teaching-learning processes
Dedication	Teachers motivates students to work hard in their lessons

[8] For a detailed description of the coding process and their IRR indexes, see the paper mentioned in the first footnote

Independent study	Teachers helps students to develop the ability to work and study independently
Expectations	Teachers acts according to the belief that students can change
Courtesy	Teachers demands that students respect basic standards of civility and courtesy
Routines	Teachers establishes routines to regulate certain activities and everyday situations
Opinion	Teachers encourages students to express their ideas and opinions

Note: See Escribano, Treviño, Nussbaum, Torres & Carrasco (in press) for more details.

All the indicators were coded by a primary school teacher and then validated through the inter-class correlation coefficient (ICC), which uses a type of inter-reliability index (Hallgreen, 2012). The validation was done by comparing the codes of 20% of the sample, randomly selected and encoded by the lead researcher. The average of the ICC inter-reliability index of the 15 indicators was 0.8 (SD = 0.11), which is considered a good or excellent level of agreement between the two raters (Cicchetti, 1994).

3.2.2.2. School level variables

The school-level variables were used as covariates in this study. These variables are school practices observed by the Agency in its visits. Unlike the teacher-level variables, the values from each of the 79 variables present in the EPS document were encoded by the Agency team. These 79 variables are distributed among the four main dimensions mentioned in the previous section. Considering the need to limit costs, it was finally decided to consider 15 variables to explore the relationship between school factors and teacher class membership (see Table 3-3).

The process of variable selection was done by the research team, considering the following criteria: 1) theoretical relevance in relation to the context of the sample (Bellei et al., 2019;

Muijs, Harris, & Chapman, 2004); 2) 25% or less of missing data per variable; 3) similarity between the constructs of the variables[9] , and 4) risk of collinearity between variables.

Finally, the dimensions covered by the 15 school-level variables are principal leadership, curricular management, integral formation, various school climate factors, staff management, and educative resources management.

[9] For example, one variable that was excluded due to the similarity between constructs was "The establishment has a school library to support student learning". This variable can be understood as a sub-dimension of the more global variable, "Didactical resources", which is used as a school-level variable in this study.

Table 3-3
Regressors at the school level

Variable	Summary of criteria
Principal's expectations	The school principal establishes a culture of high expectations in the educational community.
Principal's leadership	The school principal effectively leads to the general operation of the establishment.
Shared pedagogical guidelines	The school principal and the technical-pedagogical team agree on the standard pedagogical guidelines for the effective implementation of the curriculum.
School leaders' involvement in the classroom	The school principal and the technical-pedagogical team support the teachers by observing lessons and reviewing textbooks and other educational materials to improve the learning opportunities of the students.
School leaders' involvement in assessments	The school principal and the technical-pedagogical team coordinate an effective system of learning assessments.
Conditions for teacher collaboration	The school principal and the technical-pedagogical team promote collaborative learning and the exchange of educational resources generated among the teachers.
Promotion of parental commitment	The leadership team and teachers actively promote the involvement of parents in the educational process of students.
Promotion of diversity	The leadership team and teachers value and promote diversity as part of the richness of human groups and seek to prevent any kind of discrimination.
Norms for student-teacher relationships	The school establishes norms for student-teacher relationships that help to organize school life by disseminating them in the educational community and ensuring they are respected.
Support for behavioral issues	The leadership team and the teachers are able to identify and correct antisocial behaviors of the students, from minor situations to more serious ones.
Selection of competent personnel	The school implements effective strategies to attract, select, and retain competent personnel.
Staff development management	The school manages the professional and technical development of the staff according to the pedagogical and administrative needs.
Work environment	The school has a positive work environment.
Infrastructure	The school has the infrastructure and equipment required by the normative, and these are in conditions to facilitate the learning of students and the welfare of the educational community.
Didactical resources	The school has sufficient didactical resources and inputs to enhance student learning and promote its use.

Note. Each of these variables has four possible levels of performance for each indicator: weak (1), incipient (2), satisfactory (3), or advanced (4). In this case, it was not necessary to re-assign the values for advanced cases to satisfactory,as was done at the teacher level.

3.2.3 Statistical analysis

In order to build the teacher typology[10] based on the quality of observed practices, a Multilevel Latent Class Analysis (MLCA) with covariates was estimated through a one-step approach (Asparouhov & Muthen, 2008; Henry & Muthen, 2010; Vermunt & Magidson, 2005), using the software Latent Gold v5.1.

The Latent Class Analysis (LCA) technique classifies subjects into groups — also named categories — based on the values of the observed indicators (Masyn, 2013). In each of these categories, the possible value of the ordinal indicator has a specific probability. Therefore, each category has a particular probability-based pattern. In this study, the subjects are the teachers, and the indicators are ordinal variables of 15 teachers' observed classroom practices. With this information, nominal latent groups can be estimated.

Commonly, observations in educational research violate the independence assumption as teachers are nested within schools. Leaving the dependence of observations unmodeled often leads to the inflation of Type I error (Snijders & Bosker, 2012). In the last two decades, efforts have been made for developing multilevel LCA models where it is possible to conduct LCA with nested data (Asparouhov & Muthén, 2008; Henry & Muthen, 2010; Vermunt, 2008).

In this study, we fitted a non-parametric model (Henry & Muthen, 2010), where school variance in teacher practices is modeled by discrete latent classes. In others words, this model defines groups at both levels — in our case, school and teacher levels — thus allowing mean probabilities of teacher level to change across schools. This means that the school typology will

be determined by the different probabilities of the school's teachers being classified in a specific latent class.

There is no consensus on the appropriate method to be used when covariates are introduced into LCA models (Bolck, Croon, & Hagenaars, 2004; Vermunt, 2010). The introduction of covariates in the model implies that it is necessary to estimate the measurement model and the structural part. The measurement model refers to the relation between the observed indicators with the latent classes. The structural part of the model refers to relations between the latent class variables and the covariates. Traditionally, there are two options to estimate this kind of model; the one-step approach or the three-step approach (Bakk, Tekle, & Vermunt, 2013). The one-step model estimates the measurement model and the structural part at the same time. This approach has the benefit of raising the precision of the classification of level 1 profiles, but its downside is that the level 2 covariates influence the probability of a teacher being assigned to a specific category. In contrast, the improved three-step model (Bakk, Tekle, & Vermunt, 2013) estimates the measurement part first, then, in the second step, the membership of the individuals is fixed to the estimated category in order to later treat the categories as observed variables with a misclassification error. In the third step, this model allows us to estimate the relation between the covariable and the classification of the teachers in a specific category. The problem with this method is that in this last step data cannot be treated as nested, ignoring the associations that exist between teachers within the same school. Thus, in the present study, the one-step approach was preferred. The model specification in the one-step procedure has the advantage of improving

[10] We asume that a typology is a collection of categories, i.e. different model solutions.

the classification of teachers in the different latent classes, while providing information regarding which school factors are contributing the most to this precision improvement.

Model selection was made based on two main criteria. Firstly, the evaluation of measurements of relative fit as AIC, individual-based BIC, and group-based BIC (Fagginger Auer et al., 2016; Vermunt, 2010)[11]. It is important to mention that the lowest value of these indexes is not always sufficient to identify the number of groups. In some cases, the small decrease of these measures across models does not necessarily justify the extraction of a new group. Secondly, and complementary to the first point, it is necessary to evaluate the model trough a classification diagnostic (Masyn, 2013). This diagnosis is commonly made by the Entropy indicator (EK), which assesses the precision of individuals' classification in their categories through the posterior probabilities. Finally, various authors (Henry & Muthen, 2010; Masyn, 2013) strongly recommend evaluating not only model fit indexes, but also the theoretical and substantial interpretability of the category as a whole, and also, to avoid an extremely disparate proportion of subjects in the different classes.

It is important to note that missing data was present in all teacher level indicators. The percentage of missing data ranged from 2.3% to 26.5%, with a mean of 13.2% (SD = 7.1). The software Latent Gold uses Full Information ML estimates with the missing data present on the observed indicators (Vermunt & Magidson, 2013). There was no missing data at school level covariates..

[11] Because the models were hierarchical, the Likelihood ratio test (LMR-LRT) cannot be computed in this case.

3.3. Results

. We estimated different models to determine the typology that best represents this sample. At the teacher level, the number of categories varied from 1 to 10, and at the school level the categories varied from 1 to 3. In Table 3- 4, the fit indices are shown for 18 estimated models. In all the models, both BIC indexes show that the solution of teacher-level categories should be with six teacher latent classes. But, from the graphs of the relative fit indices (Figure 3-1), it is possible to see that the curve of relative indexes stabilizes at four latent classes for both the one school-level category solution and the two school-level categories solution.

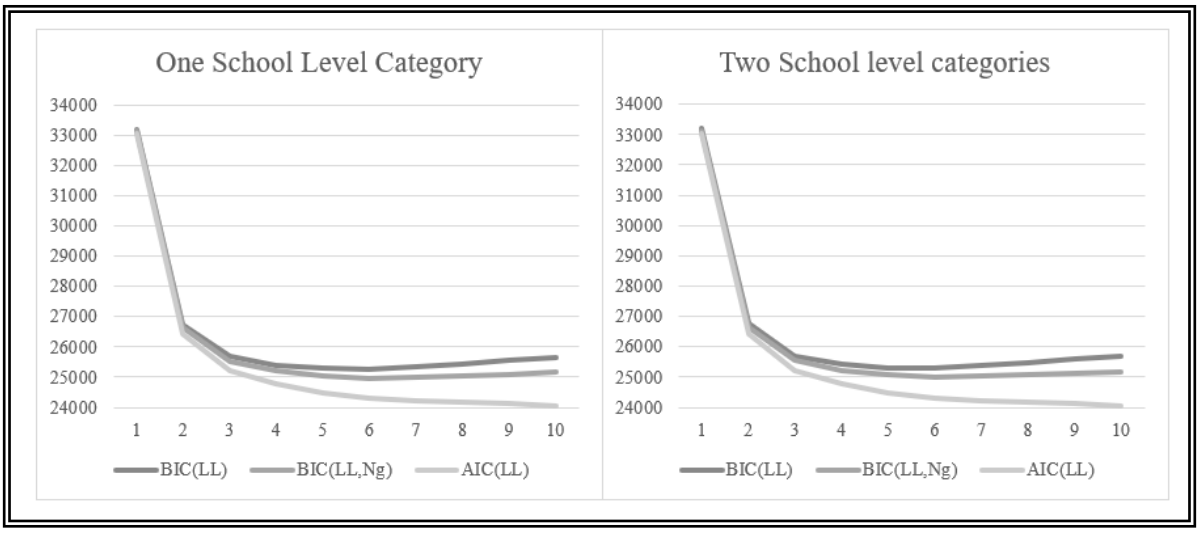

Figure 3-1. Relative fix indices values for the one and two school-level category. The teacher-level number of classes varies between 1 and 10

Although a solution with four categories, is not the solution with the minimum value between the relative fit indexes, this solution has a more substantive and parsimonious interpretation. When the resulting categories in the typologies of 5 and 6 categories at the teacher level are fitted, a duplication of classes is observed. That is, there are categories that have almost the same probability in each of the 15 indicators. Therefore, it was decided to opt for the model of 4 classes at the teacher level.

To decide the number of school-level latent categories, we inspected the models of 3 to 5 teacher classes. We carefully studied all possible models with 1 to 3 school-level latent classes, and 3 to 5 teacher-level latent classes. This implied the inspection of nine different model solutions (3x3). In all possible variations of the number of school-level categories, the four teacher level categories continued to be the best-fitting model. Regarding the selection of the number of school-level categories, in the case of the three school-level, the third category grouped only 7% of the schools. For this reason, the chosen typology was with two school-level categories and four teacher-level categories.

Table 3-4
Fit statistics for the non-parametric multilevel latent class models.

Latent Classes						BIC	
Schools	Teachers	Log-Likelihood	Parameters	Entropy	AIC	Individual-based	Group-based
1	2	-13150.98	61	0.93	26423.96	26737.48	26640.69
	3	-12511.38	92	0.88	25206.77	25679.62	25533.64
	4	-12261.24	123	0.88	24768.48	25400.66	25205.49
	5	-12094.28	154	0.86	24496.56	25288.06	25043.71
	6	-11972.82	185	0.84	24315.64	25266.48	24972.94
2	2	-13149.58	63	0.93	26425.17	26748.96	26649.01
	3	-12506.25	95	0.89	25202.50	25690.76	25540.03
	4	-12256.22	127	0.88	24766.45	25419.19	25217.67
	5	-12090.60	159	0.86	24499.20	25316.40	25064.12
	6	-11968.84	191	0.85	24319.68	25301.35	24998.29
3	2	-13149.49	65	0.93	26428.99	26763.06	26659.93
	3	-12504.75	98	0.89	25205.51	25709.19	25553.70
	4	-12252.33	131	0.89	24766.67	25439.96	25232.10
	5	-12086.18	164	0.86	24500.36	25343.27	25083.05
	6	-11962.38	197	0.85	24318.77	25331.28	25018.70

3.3.1. Interpreting the teacher-level categories

Figure 3-2 shows the class-specific item mean probability plot, which shows the average probability of a teacher in each category receiving a satisfactory score in each of the teacher practice items. The teacher categories can be defined as follows:

Cluster 1 or High-quality practices (38.47%): The teachers in this group have a high probability (>80%) of being evaluated at a satisfactory level in all 15 practices. It is remarkable in this case, considering the low socioeconomic school context, that this is the largest of the four categories identified.

Cluster 2 or Low-quality practices (25.77%): The second largest group corresponds to teachers who have particularly low probabilities (<20%) of obtaining a satisfactory level

in certain teacher practices. These teachers show a lack of dynamism in their lessons, and they do not provide feedback to their students. Their time and behavior management practices in the classroom are deficient. These teachers do not promote skill development practices or encourage their students to pay attention to their lessons.

Cluster 3 or Content-oriented practices (18.88%): The teachers assigned to this category have a high probability (≥70%) of being evaluated at a satisfactory level in practices related to content. These practices include the high-quality strategies used by the teachers to present the content, the clarity of their presentation of this content, and their ability to promote student participation in skill-development exercises. Also, these teachers express an interest in the academic development of their students. Another characteristic of this category is that, although all the indicators have a higher probability of being in the best quality-level, these are still smaller than for the teachers in Cluster 1. However, regarding classroom routines, time management, dedication of students, expectations, and courtesy indicators, the teachers of this category have low probability of being evaluated at a satisfactory level.

Cluster 4 or Climate-oriented practices (16.88%): This teacher group is characterized by higher probabilities in items relate to the climate in the classroom and the relationship between students and teachers. The indicators that differentiate this category, with a high chance (≥70%) of being evaluated as satisfactory quality level, are: appropriate time management, the dedication of students to their independent work, the establishment of routines, and behavioral rules in the classroom. Also, these teachers present a high level of interest and expectations for their students.

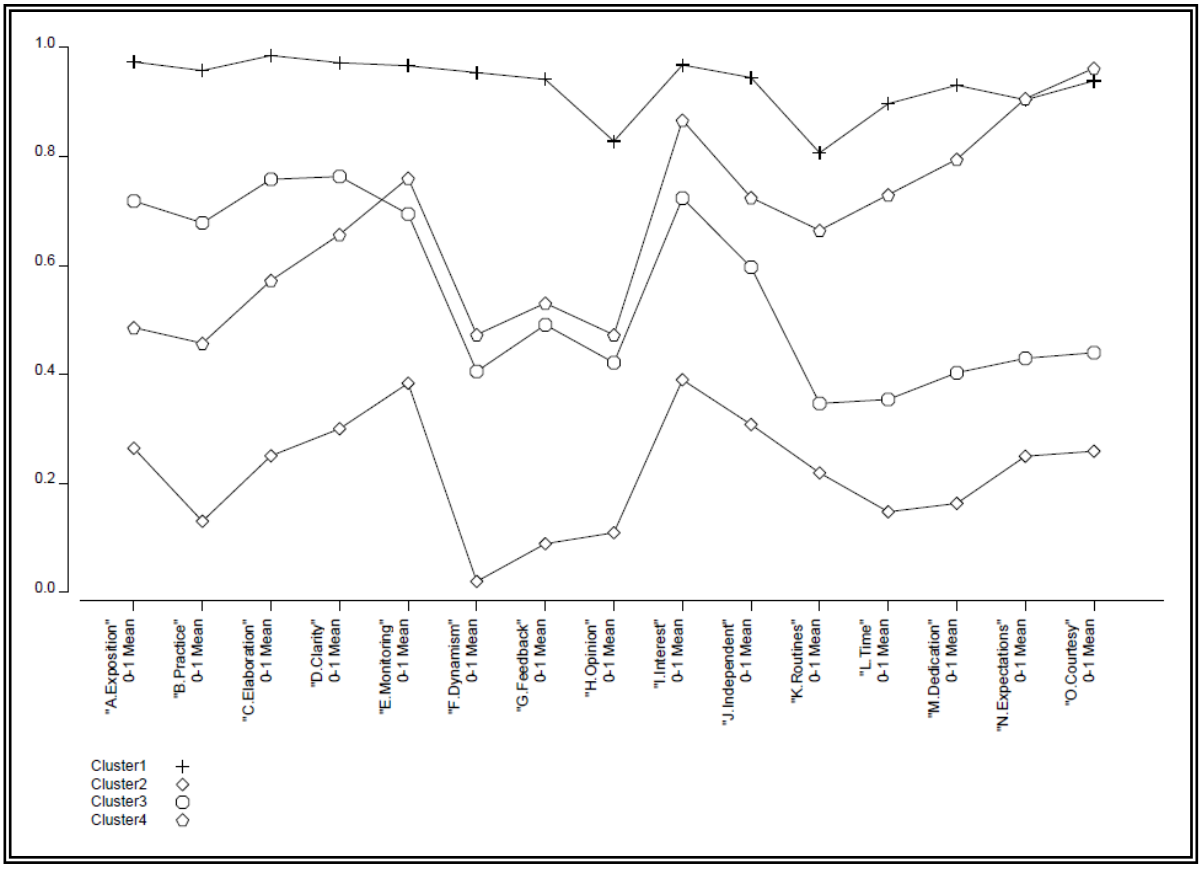

Figure 3-2. Class-specific item mean probability plot

3.3.2. School-level categories

Regarding the number of school categories identified, the results show two types of school groups (See Figure 3). The first school-category (A), in which 23.70% of the sample is present, there are only three of four types of teacher categories, with the *Content-oriented category* missing. Also, in this type of school, the proportion of teachers with *High-quality practices* is 46.50%, and the percentage of the *Climate-oriented category* is about 25.36%. In the second school-level category (B), all the teacher profiles are present, where the proportion of teachers with *High-quality practices* is 35.80%, while the proportion of teachers in the *Low-quality*

practices group and the *Content-oriented category* is 25.00%. These results have to be carefully interpreted because of the low level of entropy at the school level (<.4). For this reason, we would prefer to present this result as an exploratory finding without drawing any conclusions.

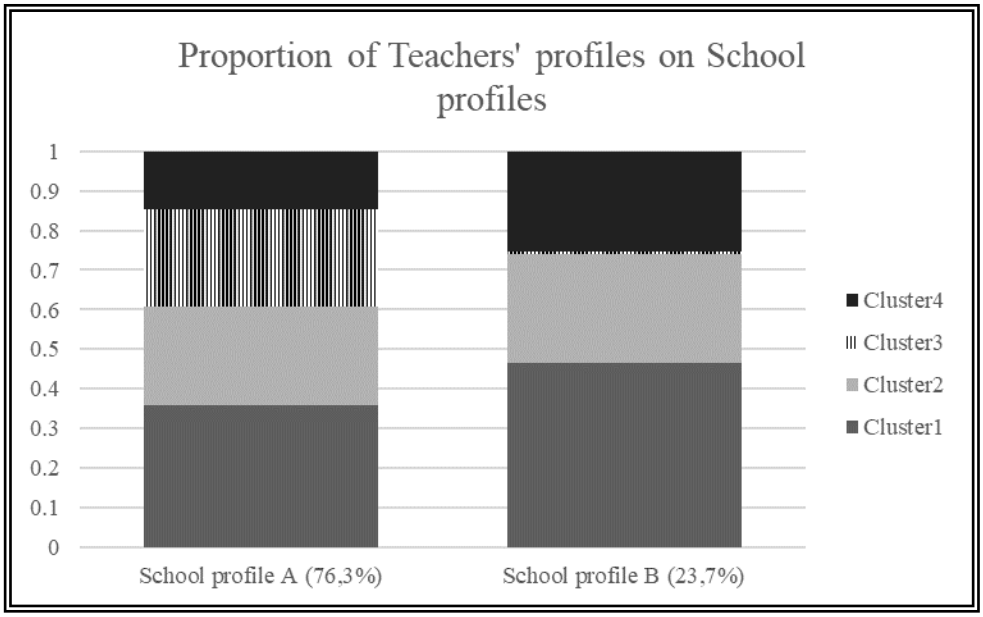

Figure 3-3. School-level description profiles based on the proportion of teacher categories

3.3.3. Covariates related to teacher category

The results of the Wald test are shown in Table 3-5. As we described in Section 2.3, the covariables were introduced considering the one-step method.[12] This implies that the variables that reject the null hypothesis ($p < 0.05$) are covariates that are informative for which school attributes are associated with the practices of teachers in their schools. That is, these covariates are relevant for which set of practices teachers are more likely to perform in these schools.

[12] The 3-step approach was also implemented, ignoring the hierarchical structure of the data, and the results of both the measurement model and the structural part did not present significant changes.

Table 3-5. Wald test results for determining which school-level covariates informing teacher level practices

Covariates	Cluster1	Cluster2	Cluster3	Cluster4	Wald	p-value
Principal's expectations	0	-0.086	-0.2127	0.0517	1.5834	0.66
Principal's leadership	0	-0.0096	0.0166	0.0667	0.1928	0.98
Shared pedagogical guidelines	0	0.0725	0.1811	0.0606	1.0647	0.79
Leadership's involvement with the classrooms	0	-0.0701	0.037	-0.0395	0.376	0.95
Leadership's involvement with assesments	0	-0.1231	-0.1893	-0.0666	1.2026	0.75
Conditions for teachers' collaboration	0	0.0714	-0.218	0.0152	2.4497	0.48
Promotion of parent´s commitment	0	-0.2813	-0.2085	0.0763	3.9241	0.27
Promotion of diversity	0	-0.0331	-0.2086	0.2171	3.3225	0.34
Normative for relations	0	-0.0922	0.0904	-0.3632	4.2318	0.24
Support for behavioural issues	0	-0.4965	-0.0654	0.1777	14.8889	0.0019
Competent personnel selection	0	0.1242	-0.0307	0.3004	3.9755	0.26
Staff development management	0	-0.0658	0.0399	-0.4144	5.4977	0.14
Work environment	0	-0.2641	-0.1712	0.0313	4.5215	0.21
Infrastructure	0	-0.0249	0.1002	-0.1488	0.9223	0.82
Didactical resources	0	-0.1078	0.1868	-0.2603	3.8939	0.27

"School support for behavioral issues" is one covariate associated with teachers' group membership. This variable refers to the school team's capacity to amend any antisocial behaviors of the students. Also, this covariate highlights the importance of the actions and goal alignment between the school authorities and teachers in order to accomplish their objective. The term "antisocial behaviors" groups two main types of actions: the first is minor antisocial behavior such as rude language, cheating on tests, destroying infrastructure, and theft, while the second type is dangerous behavior such as vandalism, cruelty, carrying guns, violent fights, assaults, and threats. This covariate has a positive association with the Climate-oriented teacher category. This cluster is characterized by teacher practices that focus on maintaining a pleasant classroom climate and a good relationship between teachers and students. This is a crucially important issue

in low-achieving and low-SES schools in Chile (Treviño et al., 2015). Specifically, when we test the contrast between the categories in pairs for the significant school-level covariable, the results shows a significant difference ($p = .0012$) between the *Climate-oriented practices* and the *Low-quality practices* clusters, where the value of support on behavioral issues is positively related with the probability of a teacher of being assigned to the *Climate-oriented practices* category instead of the *Low-quality practices* group.

3.4 Discussion and conclusions

This study aimed to offer a different perspective on how the efficacy of schools and teachers can be understood by using a person-centered instead of a variable-oriented approach. The first approach has the advantage that it assumes that the population is heterogeneous in relation to the variables of interest, and it focuses on understanding individual or group differences (Reichert, 2016). From this perspective, the purpose was to understand how teacher quality varies within the same school, and which factors are related to teacher quality variation within the school.

To fulfill this purpose, this research explores teacher typologies based on classroom quality indicators, the distribution of these typologies across types of schools by applying multilevel LCA, and also, the relation of these categories with school atributes. As the literature using this methodology in Educational Effectiveness Research is scarce, it is difficult to compare the findings described above with research results from similar studies. For this reason, the focus of the discussion will be on the importance of these results to better understand the strengths and weaknesses of the teaching force in low-achieving schools in Chile.

Interestingly, one unexpected result of this research is the large proportion of teachers in low-quality index schools found to be in the *High-quality practices* teacher category. Previous

research shows that there is great variability in the quality of practices across and within schools (Bruns & Luque, 2015; Mansfield, 2015), so it is expected that at least some teachers in these low-quality schools would show good performance in the classroom practices observed. However, what is remarkable is that the teachers with *High-quality practices* — usually teachers that perform as expected — represent almost 40% of the sample. It is important to note that, despite this fact, student achievement in these schools is low. This may be mainly due to two reasons. First, a high proportion of low-achieving schools are low-SES schools, with Chile being the country where the school SES average has the strongest influence on student achievement in the OECD — as shown in the PISA-studies (OECD, 2017). Second, academic achievement is measured mainly in 4th and 8th grades, and results in these grades may be reflecting the total effect of the quality of teaching in each school. Therefore, students in schools are subject to the differential quality of teaching throughout their school trajectory because they attend classrooms where they are taught by several teachers, not only by high-quality teachers. In other words, it can be hypothesized that even though students have some high-quality teachers, the effect of these teachers is not enough to compensate for the effects of their challenging context. This finding shows that, when analyzing schools, it is important to consider the distribution of capacities of the teaching force to identify with greater precision whether there are thresholds of installed teaching capacities — for example, a certain percentage of teachers with *Good-quality practices* — that would be a minimum for ensuring that a school in this context can improve.

Our results also complement the traditional way in which the National Teacher Assessment System in Chile defines, assesses, and acts to improve teacher quality. While the national system defines teacher practices as a continuous one-dimensional variable, teachers' typologies help to

better delineate the strengths and weaknesses of teacher practices and provide inputs for designing and implementing effective interventions to strengthen teacher capacities.

Both findings — the proportion of high-quality teachers and the teacher typologies — are relevant to policymakers, who can develop strategies to strengthen teachers' capacities in a school, for example, through programs that consider collaborative work between teachers with well-qualified teachers as tutors (Hargreaves & Fullan, 2012). This option can, at least partially, tackle the problem of decontextualization of educational policies.

Another finding is that the only school factor related to the teacher membership profile was the quality of the school environment regarding support for behavioral issues. This finding is complementary to the study of Kuhfeld (2017), which relates results from the Tripod students survey with teaching quality indicators . The Tripod survey is an instrument to measure students' perceptions of teacher quality based on students' observations scores in seven sub-dimensions (Fergurson, 2010; Kuhfeld, 2017). The results of this study show that an adequate classroom climate — and a high overall teacher score in the observed practices by students on the Tripod survey — is positively related to student achievement. Also, the results of our study show specifically that the quality of school support is positively associated with the probability of pertaining to the *Climate-support practices* category rather than the *Low-quality practices* category. This finding could be very useful in choosing which intervention school-factor to focus on, in order to mobilize teachers belonging to the undesirable *Low-quality* category, to a better category such as *Content-oriented*.

The schools in this study have a low-quality index, which is highly related to a low SES. In these challenging contexts, it is not surprising that maintaining a positive climate at the school level is

aligned with the probability of teachers maintaining a good classroom climate. The PISA In Focus document (OECD, 2013) points out that a good school disciplinary climate not only positively influences the academic achievement of the students, but it can also reduce the impact of SES on students' performance. Finally, it is common that in vulnerable contexts there are problematic antisocial behaviors that are also related to social exclusion and crime in the neighborhoods where students live. Therefore, alignment between teachers and schools is important, but may not be sufficient in all cases to improve school and classroom climate across schools and teachers.

3.5 Limitations and future work

The first limitation is the reduced range of the sample in terms of results in academic achievement, where students in most of the cases score below the national average. This implies that these results cannot be generalized to the whole school population. A second limitation is that the classroom observation instrument was designed for public policy purposes, not for research purposes. This can lead to data with less precision than data collected with an instrument developed for research purposes. However, as stated above, the observation protocol is part of the policy guidelines that teachers and schools in Chile must follow. Therefore, this policy instrument should influence the improvement of teacher practices in schools more than any research instrument. Future work should be oriented towards expanding the sample size, and considering different ranges of SES and learning achievement of the students in order to compare the categories of teachers that emerge from an MLCA

3.6. Acknowledgements and Funding

3.6.1. Acknowledgements

This research was made possible thanks to the collaboration of the Chilean Agency for Quality Education. The results, discussion, and conclusions are the sole responsibility of the authors and do not in any way represent the opinion of the Agency.

3.6.2. Funding

This project was funded by the National Commission of Scientific and Technological Research of Chile, through the project Center of Advanced Studies on Educational Justice [**Proyecto CONICYT PIA CIE160007**], and the CONICYT PFCHA/BECA DOCTORADO NACIONAL/2016 - 21160063

4. THE ELEPHANT IN THE (CLASS)ROOM: ANALYSING THE INCLUSION OF IMMIGRANT STUDENTS WITHIN TEACHER-STUDENT INTERACTION NETWORKS

4.1 Introduction

The proportion of students with immigrant background in Chile has increased rapidly in the last decade and tripled in the last three years, due to a growing influx of intraregional migration (INE, 2018). While unrestricted access of immigrant students to formal education has been ensured, little is known regarding their educational inclusion within schools and classrooms. The growing immigration trend and new regulations of the educational system, are confronting Chilean teachers with a more diverse student population than in the past. It is, thus, an increasingly relevant challenge to provide equal learning opportunities to all students regardless of their background. However, there is a lack of relevant evidence to inform teacher training and practices in this regard (European Union, 2017).

The aims of this study are (1) to investigate the inclusion of immigrant students within Chilean classrooms by analysing their relative position in teacher-student interaction networks[13] and, (2) to study and explain the variation in the inclusion of immigrant students across classrooms.

The international literature around this topic suggest that immigrant and ethnic minority students face prejudices and discrimination from their school community. However, this literature is usually based on very limited samples or teacher self-reports. Furthermore, no previous studies in Chile have systematically observed how teachers actually interact with immigrant students in

[13] We use the term teacher-student interactions to refer to interactions between a student and the teacher, regardless of the initiator of the interaction. When be distinguish by the interaction initiator, we specify this by referring to teacher-initiated or student-initiated interactions.

the classroom. The contribution of this study is three-fold. Firstly, the paper advances the field empirically, by providing systematic evidence –from a large sample– on the inclusion of immigrant students in Chile, a country with an unprecedented immigration trend. Secondly, it applies an innovative combination of advanced methods to investigate teacher-student interactions (i.e., systematic classroom observation, social network visualization and multilevel models). Thirdly, it investigates the within- and between-classroom variation in the inclusion of immigrant students across a variety of interaction contents while controlling for confounding variables and exploring compositional effects, setting a methodological benchmark for national and international studies.

4.2 Conceptual Framework

The 2030 Sustainable Development Goals aim to "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (United Nations, 2015: p 14). Eleven years away from this goal deadline, immigrant students perform at systematically lower levels than their non-immigrant peers (OECD, 2015a) and face various challenges in terms of access to, retention in, and completion of mandatory schooling (UNESCO, 2018). Among many factors contributing to educational inclusion, those operating within the classroom seem to be especially important (Loreman, 2014).

Educational effectiveness research has shown that a significant proportion of the variation in student outcomes is located at the classroom level. An important part of this classroom-level variation can be explained by the quality of teaching and interactions in the classroom (Muijs et al., 2014; Ortega et al., 2017). For instance, there is a clear association between the quality of interactions in the classroom and students' motivation and learning outcomes (e.g., Hattie, 2009;

Wubbels, Brekelmans, & Mainhard, 2016). This association has been confirmed empirically in Chile's pre-school education (Leyva et al., 2015), as well as in primary and secondary education (Taut et al., 2016). However, differential effectiveness has received less international attention, and has not yet been explored in Chile. This issue considers how schools, teachers, and particular teaching strategies, are differentially effective for different groups of students. For example, whether teachers promote different progress on educational outcomes for immigrant and non-immigrant students (Campbell, Kyriakides, Muijs & Robinson, 2003; Caro, Lenkeit, & Kyriakides, 2016; Strand, 2010).

Methods used to measure and evaluate classroom interactions include teachers' or students' self-reports – which can be biased (Desimone, Smith, & Frisvold, 2010; Mashburn, Hamre, Downer, & Pianta, 2006) –, and systematic classroom observation protocols. Most of the available protocols, such as the Classroom Assessment Scoring System (CLASS) (Pianta, Hamre & Mintz, 2011) and the International System for Teacher Observation and Feedback (ISTOF) (Teddlie et al, 2006), provide general measures of overall quality of classroom interactions. While these observation tools offer important insights, they do not distinguish how teachers interact with different students and, therefore, assume that students' experiences within a class are homogeneous (Martínez, Taut & Schaafa, 2016). Thus, the characteristics of each student (e.g., immigration background, gender and prior achievement), as well as potential teacher biases against students with certain characteristics, are hidden in instruments that report on the average quality of relationships between the teacher and the class.

In consequence, a complementary approach is required that allows the empirical study of educational inclusion in the classroom in terms of the interactions that the teacher establishes

with different groups of students within a class, and that might partially explain their differential effects. In this study, we propose a strategy for analysing differential teacher-student interaction patterns within classrooms, by focusing on differences by student immigrant background and country of origin.

4.2.1 The inclusion of immigrant students in teacher-student interactions

While there is a vast amount of research on the inclusion of immigrant students in peer networks (e.g., Moody, 2001; Vitorouslis & Georgiades, 2017), less is known about their inclusion in teacher-student interactions, a critical source of pedagogical and emotional support (Contini, 2013; Pianta, Hamre, & Mintz, 2011). Previous research suggests that the way teachers communicate vary by the immigrant background of their students (den Brok & Levy, 2005). For example, some early exploratory studies concluded that teachers establish less interactions with pupils of immigrant background than with other students and that the contents of the interactions also tend to differ, with teachers being more critical towards immigrant students and correcting their behaviour more frequently, compared to their non-immigrant peers (e.g., Casteel, 1998; den Brok, Wubbels, Veldman, & van Tartwijk, 2009; Fraser & Walberg, 2005; Irvine, 1985; 1986).

Quantitative studies in the field of educational psychology, have found evidence of differences in teacher-student relations by student ethnicity, with black students showing lower levels of closeness (Ladd, Birch, & Buhs,1999) and higher levels of conflict with their teacher that were also more likely to increase over time (Jerome, Hamre, & Pianta, 2008), compared to their peers. Also, a meta-analysis suggests that teachers' speech vary with students' ethnic backgrounds, with more positive and neutral speech for European American children than for African American and Latino/a (Tenenbaum & Ruck, 2007). Finally, a recent review of the literature on teacher

interactions with multilingual and monolingual children shows that multilingual students receive fewer learning opportunities, as compared with their monolingual peers (Langeloo, Mascareño, Deunk, Klitzing, & Strijbos, 2019).

These quantitative studies are usually not based on classroom observation data but on teachers' self-report, which may contain biases (Desimone, Smith, & Frisvold, 2010; Mashburn, Hamre, Downer, & Pianta, 2006). In addition, this body of literature does not distinguish among contents of teacher-student interaction, a key aspect in the analysis of the quality of classroom interactions (Brophy & Good, 1974).

4.2.1.1 The role of teacher attitudes in the inclusion of immigrant students

Previous research suggests that, in multicultural contexts, teachers hold differential expectations of educational attainment for their students, to the detriment of certain groups such as immigrant, black and Hispanic students (de Boer, Bosker, & van der Werf, 2010; Mahatmya, Lohman, Brown, & Conway-Turner, 2016; Tenenbaum & Ruck, 2007).

It is important to note that this field of research has also found associations between teachers' lower expectations towards immigrant students, the relative exclusion of these students from classroom interactions and lower progress on learning outcomes (e.g., de Boer, Bosker, & van der Werf, 2010; Glock, 2016; Mahatmya, Lohman, Brown, & Conway-Turner, 2016; Peterson, Rubie-Davies, Osborne, & Sibley, 2016).

4.2.1.2 The effect of classroom immigrant composition on the inclusion of immigrant students

While compositional effects have not been studied in teacher-immigrant student interaction research, the international peer interaction literature highlights the association of school and classroom ethnic composition with immigrant social inclusion (e.g., Cotini, 2013; Moody, 2001; Vitorouslis & Gerogiades, 2017). This research has generally found a non-linear relationship between the group's ethnic composition and different measurement of ethnic segregation (i.e., on friendship networks, racial victimization and educational attainment). Thus, it is important to investigate how the proportion of immigrant students in the classroom influences the interactions that teachers hold with students from immigrant and non-immigrant background.

4.2.2 The Chilean context and national research on immigrants' educational inclusion

The foreign-born population in Chile has increased significantly in the last decade due to a growing influx of intraregional migration (DEM, 2016). According to census data, the migrant population rate increased from 1,3% in 2002 to 4,4% in 2017, showing an explosive evolution since 2010 (INE, 2018). This new scenario is reflected in Chile's school enrolment trends. Official national reports show that 113,585 immigrant background students were enrolled in the school system in 2018, which is four times the number observed for this group in 2015. These students mainly originate from Latin American and Caribbean nations, such as Peru, Bolivia and, more recently, Colombia, Venezuela, and non-Spanish-speaking Haiti (MINEDUC, 2018).

Unrestricted access of immigrant students to formal education has been secured in Chile, through the promotion of multiple initiatives. These include the enactment of legislations that give

migrant children unconditional access to the educational system and the creation of educational resources and guidance for inclusive practices in the classroom (DEM, 2018; Mora, 2018). However, effective inclusion goes beyond access and formal participation, and little is known regarding the educational inclusion of immigrant students within schools and classrooms in Chile.

Immigrant students in Chile are unequally distributed both geographically and in terms of school sector. Migrant families tend to cluster in central areas of large cities, which leads to some schools in those areas showing higher proportions of pupils with immigrant background (Fernández, 2018). In addition, immigrant students in Chile are more likely to attend public schools (57.5% of immigrant students attend public schools, compared to 35.3% of non-immigrant students) (MINEDUC, 2018). In Chile's market-oriented and socially segregated education system, public schools usually serve low socio-economic status families (García-Huidobro, 2007; Valenzuela, Bellei, & De los Ríos, 2014) and show poorer educational results (Cox, 2004).

The immigration patterns described above have been shifting the composition of some schools dramatically, leading to more diverse classroom environments. However, little research has been carried out in Chile to understand immigrant children trajectories in the educational system. The available literature shows that, overall, immigrant students in public schools perform better than their national counterparts regarding grade repetition rates and they present no differences in the educational levels and expectations of their parents, compared to the national population (Castillo, Santa-Cruz, & Vega, 2018; Joiko & Vásquez, 2016).

Very few studies have investigated the social inclusion of students of foreign background in Chile. Qualitative studies conducted in Chilean schools suggest high levels of intolerance, the presence of negative stereotypes, as well as racism and discrimination against immigrant students, from both teachers and peers (Bravo, 2011; Joiko & Vasquez, 2016; Pavez-Soto & Chan, 2018; Salas et al, 2017; Stefoni, Acosta, Gaymer, & Casas-Cordero, 2008; Tijoux, 2013b).

A recent quantitative study investigating the position of immigrant students within their class' friendship network (Castillo, Santa-Cruz, & Vega, 2018) shows no significant overall differences between immigrant and non-immigrant students. However, differences emerge when distinguishing by the proportion of immigrant students in the class: in schools with low proportions of immigrants, immigrant students tend to be ignored or isolated, in schools with medium proportions of immigrant students they are better integrated, and in schools with higher proportion of immigrant students, immigrant children are more likely to be rejected. This is in line with the international evidence on non-linear compositional effects (e.g., Moody 2001).

While the Chilean literature has not systematically analysed the inclusion of immigrant students within teacher-student interactions, there are some recent exploratory case studies on teachers' attitudes towards immigrant students. This research suggests that teachers hold stereotyped attitudes and discourses towards immigrants, as they associate students from certain countries of origin with academic deficits and behavioural problems that, according to them, represent obstacles for their inclusion and participation in schools—understood as assimilation to the culture of the school and the country. For example, in qualitative studies teachers have express that students from Venezuela and Dominican Republic are more disruptive but also more likely to perform higher academically, while Peruvian students are seen as more introverted, less

prepared academically and less likely to conform to school rules (Cerón, Pérez Alvarado, & Poblete, 2017; Mondaca, Muñoz, Gajardo, & Gairín, 2018; Tijoux, 2013a).

In addition, there is evidence suggesting that Chilean teachers are not trained on dealing with diverse classrooms and do not have the necessary intercultural abilities, knowledge and pedagogical tools to support immigrant students in their processes of integration, as accounted in multiple observational studies (Carillo, 2014; Jimenez et al., 2017; Mondaca et al., 2018; Mora, 2018). Indeed, the percentage of teachers who report that they can cope well with the challenges of a multicultural classroom in Chile is significantly below the OECD average, according to the most recent results of the Teaching and Learning International Survey (TALIS) (OECD, 2019).

Based on the reviewed literature, the evidence on immigrant children inclusion, specifically in terms of teacher-student interactions, is very scarce in Chile, as it is elsewhere. This paper explores the relative position of immigrant students within their class' teacher-student interaction networks, as well as the magnitude and predictors of the variance in the inclusion of immigrant students across Chilean classrooms. This is the first study that provides systematic evidence on teacher-immigrant student interactions within Chilean classrooms. Based on the literature reviewed above, the analyses will be guided by the following hypotheses:

- **Hypothesis 1a**: Students with immigration background are significantly more peripheral in their class' teacher-student interaction networks than their non-immigrant peers, with teacher-immigrant-student interactions being less frequent than teacher-non-immigrant-student interactions.
- **Hypothesis 1b**: This difference will vary by immigrant students' country of origin.

- **Hypothesis 2**: There is significant variation in the effect of student immigrant background on the frequency of interactions with the teacher across classrooms, with some classrooms being significantly more inclusive of immigrant students than others.
- **Hypothesis 3**: The classroom-level variation on the effect of student immigrant background is partially explained by the immigrant composition of the classroom and the teacher's attitude towards diversity in the classroom.

All of the hypotheses above are formulated for both teacher-initiated and student-initiated interactions.

4.3 Methods

4.3.1 Sample

The study analyses data from a sample of 38 mathematics teachers and 933 seventh graders (typically aged 12 to 13) in 37 Chilean public schools. The schools were sampled based on their location within the five urban municipalities with the largest proportions of students with immigrant background in Santiago, Chile's capital. Table 4-1 shows that the average proportion of immigrant students in the participant classrooms was 42%, ranging from 3 to 90%. The socioeconomic status (SES) composition of the sampled schools corresponds mainly to Medium SES (n = 16) and Medium-Low (n = 16), followed by Low (n = 3) and Medium-High (n = 2) SES schools (Agencia de Calidad de la Educación, 2019).

Table 4-1. Demographics of the sample classrooms ($n_{classrooms}$ = 38).

	Number of students	**Percentage of immigrant students (%)**	**Number of countries of origin (Immigrant Diversity)**	**Number of total Teacher-student interactions**	**Number of teacher-initiated interactions**	**Number of student-initiated interactions**
Mean	24.6	42.0	4.61	132.4	73.0	59.4
Median	23.5	34.5	5.00	133.0	68.0	57.5

Standard Deviation	7.1	27.0	1.57	43.4	37.5	23.6
Minimum	14.0	3.3	2	33.0	6.0	23.0
Maximum	44.0	89.5	8	204.0	165.0	106.0

Tables 4-2 and 4-3 provide descriptive statistics on students' mathematics achievement[14] and their parent's educational level, by immigrant origin, language spoken at home and country of origin. We found no statistically significant differences in the achievement levels in mathematics of these children by immigrant background. However, when analysing mathematics achievement levels by *Country of origin*, students from Chile (M = 561.4, SD = 26.3) performed significantly higher that students from Haiti (M = 548.6, SD = 25.1; $t(548) = 2.199$, $p < .05$) and significantly lower than students from the category Other countries[15] (M = 574.2, SD = 29.6; $t(547) = -2.122$, $p < .05$).

With regard to parental educational level, we see that the parents of students of immigrant origin have attended university in a significantly larger proportion (30.5%) than parents of students that are not from immigrant origin (14.4%) ($\chi^2(1) = 35.08$, $p < .001$), which is the case of students from Colombia (37.3%), Haiti (33.3%), Venezuela (64.1%), Dominican Republic (52.9%) and those from the category Other countries (45.0%).

Finally, in line with the recent national immigration trends, the composition of our sample in terms of countries of origin is not equally distributed across immigrant generations. As shown in Table A4-1 (Appendix F), the students who are second-generation immigrants are mainly from Peru (71.3%) and from Other Countries (13.8%), whereas among first-generation immigrants, the percentage of students from Peru is lower (40.9%), with students from Venezuela (18.5%),

[14] Mathematics achievement refers to the score on the standardized SEPA mathematics test, which is introduced in section 4.3.4.2.

[15] Countries with less than 10 students in our sample were classified as 'Other countries'. This category corresponds to students from Argentina, China, Cuba, El Salvador, Spain, United States, Uruguay, Paraguay, Brazil and Nigeria

Colombia (13.8%), Haiti (5.9%) and Dominican Republic (5.0%) and other countries showing an increased presence.

Table 4-2. Descriptive statistics for teacher-initiated interactions by immigrant background, language spoken at home and country of origin.

					Teacher-initiated interactions				
	N Students	*% Parents' University Studies*		*Math Score*	*Behaviour Management*	*Instructions*	*Administrative*	*Pedagogical*	*Total*
Immigrant Background									
Students without Migrant Background	513	14.43	Mean	561.8	0.81	0.13	1.09	1.03	3.16
			SD	25.9	1.69	0.43	1.45	1.53	3.52
Students with Migrant Background	420	**30.48*** **	Mean	562.3	**0.59****	0.12	**0.97†**	0.98	**2.74***
			SD	30.0	**1.35**	0.43	**1.51**	1.59	**3.39**
Total	933	21.65	Mean	562.0	0.71	0.13	1.04	1.00	2.97
			SD	27.8	1.55	0.43	1.48	1.55	3.47
Language Spoken at Home									
Spanish	898	21.27	Mean	562.2	0.70	0.13	1.03	0.97	2.92
			SD	27.9	1.53	0.43	1.49	1.52	3.43
Other	35	31.43	Mean	556.1	1.00	0.11	1.20	**1.83*** **	**4.23***
			SD	25.6	2.03	0.40	1.30	**2.09**	**4.05**
Total	933	21.65	Mean	562.0	0.71	0.13	1.04	1.00	2.97
			SD	27.8	1.55	0.43	1.48	1.55	3.47
Country of Origin									
Chile	529	13.99	Mean	561.4	0.81	0.13	1.08	0.99	3.11
			SD	26.3	1.70	0.43	1.44	1.50	3.53
Peru	196	17.86	Mean	562.8	**0.52*** **	0.07	**0.84†**	**0.69*** **	**2.21*** **
			SD	31.0	**1.18**	0.30	**1.25**	**1.18**	**2.60**
Bolivia	11	27.27	Mean	564.6	1.09	0.18	0.64	1.18	3.09
			SD	32.1	1.81	0.60	0.81	1.17	3.39
Colombia	51	**37.26*** **	Mean	564.0	0.43	0.10	**0.76***	0.84	**2.25***
			SD	19.7	0.92	0.36	**1.26**	1.53	**2.98**
Haiti	21	**33.33***	Mean	**548.6***	1.24	0.14	**1.48†**	**2.05*** **	**5.00***
			SD	**25.1**	2.59	0.36	**1.40**	**2.27**	**4.24**
Ecuador	13	15.39	Mean	559.0	1.38	**0.54***	1.92	1.38	5.31
			SD	18.3	2.57	**0.97**	2.18	1.71	5.50
Venezuela	64	**64.06*** **	Mean	564.2	**0.39***	0.20	1.25	1.36	3.23
			SD	37.0	**0.85**	0.62	2.29	2.10	4.35
Dominican Republic	17	**52.94*** **	Mean	564.7	0.35	0.12	0.94	1.53	3.00
			SD	24.2	0.70	0.33	1.25	2.24	3.43

Other countries[1] (Argentina, China, Cuba, El Salvador, Spain, United States, Uruguay, Paraguay, Brazil, Nigeria)	20	**45.00*****	Mean	**574.2***	0.95	0.15	1.25	1.25	3.65
			SD	**29.6**	1.61	0.37	1.71	1.74	3.31
Missing	11	27.27	Mean	554.6	0.55	0.00	0.91	**1.82†**	3.55
			SD	18.4	0.82	0.00	1.14	**2.14**	2.98
Total	933	21.65	Mean	562.0	0.71	0.13	1.04	1.00	2.97
			SD	27.8	1.55	0.43	1.48	1.55	3.47

Notes: † $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$ in Chi-squared test (for parents' university studies), T-test (for math achievement score) and Mann-Whitney U Tests (for the rest of the variables).
Baseline for comparisons is Chile, for the Country of Origin variables.
[1] Countries with less than 10 students in our sample were classified as 'Other countries'.

Table 4-3. Descriptive statistics for student-initiated interactions by immigrant background, language spoken at home and country of origin.

					Student-initiated interactions				
	N Students	*% Parents' University Studies*		*Math Score*	*Behaviour Management*	*Task Instructions*	*Administrative*	*Pedagogical*	*Total*
Immigrant Background									
Students without Migrant Background	513	14.43	Mean	561.8	0.01	0.13	0.80	1.36	2.38
			SD	25.9	0.10	0.45	1.53	2.60	3.68
Students with Migrant Background	420	**30.48*** **	Mean	562.3	0.00	0.13	**0.65†**	**1.62***	2.47
			SD	30.0	0.05	0.55	**1.52**	**2.66**	3.80
Total	933	21.65	Mean	562.0	0.01	0.13	0.73	1.48	2.42
			SD	27.8	0.08	0.50	1.53	2.63	3.73
Language Spoken at Home									
Spanish	898	21.27	Mean	562.2	0.01	0.13	0.74	1.46	2.41
			SD	27.9	0.08	0.50	1.54	2.63	3.75
Other	35	31.43	Mean	556.1	0.00	0.09	0.69	1.83	2.60
			SD	25.6	0.00	0.38	1.08	2.50	3.26
Total	933	21.65	Mean	562.0	0.01	0.13	0.73	1.48	2.42
			SD	27.8	0.08	0.50	1.53	2.63	3.73
Country of Origin									
Chile	529	13.99	Mean	561.4	0.01	0.14	0.80	1.35	2.37
			SD	26.3	0.10	0.47	1.51	2.57	3.65
Peru	196	17.86	Mean	562.8	0.01	**0.05****	**0.49****	0.99	**1.59***
			SD	31.0	0.07	**0.21**	**1.68**	1.72	**3.00**
Bolivia	11	27.27	Mean	564.6	0.00	0.09	0.55	1.00	1.73
			SD	32.1	0.00	0.30	0.52	1.26	1.68
Colombia	51	**37.26*** **	Mean	564.0	0.00	0.39	0.90	**2.75****	**4.22***
			SD	19.7	0.00	1.22	1.57	**3.53**	**5.19**
Haiti	21	**33.33***	Mean	**548.6***	0.00	0.00	0.52	1.19	1.71
			SD	**25.1**	0.00	0.00	0.87	1.83	2.08
Ecuador	13	15.39	Mean	560.7	0.00	0.00	0.00	0.00	0.00
			SD	18.3	0.00	0.60	1.95	1.68	2.84
Venezuela	64	**64.06*** **	Mean	564.2	0.00	0.19	0.84	**2.84*** **	**3.98****
			SD	37.0	0.00	0.73	1.51	**3.64**	**4.72**
Dominican Republic	17	**52.94*** **	Mean	564.7	0.00	0.12	0.88	**2.53†**	3.53
			SD	24.2	0.00	0.33	1.41	**3.64**	4.52

Other countries[1] (Argentina, China, Cuba, El Salvador, Spain, United States, Uruguay, Paraguay, Brazil, Nigeria)	20	**45.00*** **	Mean	**574.2***	0.00	0.05	0.75	1.95	2.75
			SD	**29.6**	0.00	0.22	1.12	3.47	4.13
Missing	11	27.27	Mean	554.6	0.00	0.00	0.55	1.73	2.27
			SD	18.4	0.00	0.00	1.21	2.49	3.20
Total	933	21.65	Mean	562.0	0.01	0.13	0.73	1.48	2.42
			SD	27.8	0.08	0.50	1.53	2.63	3.73

Notes: † $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$ in Chi-squared test (for parents' university studies), T-test (for math achievement score) and Mann-Whitney U Tests (for the rest of the variables).
Baseline for comparisons is Chile, for the Country of Origin variables.
[1] Countries with less than 10 students in our sample were classified as 'Other countries'

4.3.2 Data collection

The study's data collection took place between March and April of 2018, at the beginning of the school year. Participant students completed a standardised mathematics test and a self-administered questionnaire that included items on demographic characteristics as well as their parents' and their own countries of birth. Participant mathematics teachers were also asked to complete a self-administered questionnaire that collected background information and scales regarding their attitudes towards diversity in the classroom. These data sets were complemented and validated by linking them to student- and school-level administrative data from public databases of the Ministry of Education.

In addition, a regular mathematics lesson conducted by each participant teacher was videotaped, lasting each approximately 80 minutes. The context of filming can be considered as low stakes for both teachers and students. The participants were not informed about the study's focus on the inclusion of immigrant students, in order to avoid introducing bias in the results and capture, to the greatest extent possible, classes and interactions that were typical of the observed classrooms.

4.3.3 Procedure

Each teacher-student interaction observed was coded by two professionals[16], who followed a classroom systematic observation protocol designed for this study. The protocol identified the individual student with whom the teacher interacted based on an ID, which indicated their position in the classroom sitting plan[17] In total, 5,031 teacher-student interactions[18] were identified, of which 2,187 (43.5%) were teacher-immigrant student interactions.

A random sample of 23.7% of the videos ($n_{classrooms}$ = 9, $n_{students}$ = 218) was double-coded. During the coding process, the level of agreement of the trained coders was monitored and meetings were held with the team to solve disagreements. Following Hallgren (2012), the reliability between coders was assessed for the different types of interactions analysed, using a

[16] Student-to-student interactions, although important, were not considered in the present study

[17] The spatial disposition of the students in the classrooms studied was mostly of the traditional type (i.e. desks in rows with students facing the blackboard and the teacher).

[18] This is the sum of the number of both teacher- and student-initiated interactions

two-way mixed consistency single-measures intra-class correlation (ICC), to evaluate the degree to which the coders provided consistency in their identification of the frequency of teacher-student interactions in the sample. As shown in the Appendix G (Table A4-2), the ICC was in ranges considered moderate to excellent (ICC = 0.53 - 0.82) (Koo & Li, 2016). The information obtained regarding each student's interactions with the teacher was linked to the collected student and classroom data.

4.3.4 Measures

4.3.4.1 Teacher-student interaction measures

The dependent variables in this study refer to the frequency of interactions with the teacher, in terms of the number of interactions initiated by the teacher and in terms of the number of interactions with the teacher initiated by the student. All exchanges that involved the teacher and that consisted of at least one verbal turn were identified as interactions. These were coded using code adapted from the Brophy-Good Dyadic Child Interaction System (Brophy & Good, 1974) and the TIMSS Videotape Classroom Study (Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999) instruments, which provide distinctions with regard to the initiator and content of the interactions. Thus, each interaction observed was coded, using the following indicators:

- *Initiator of interaction*: A dichotomous indicator registering if the first turn of the interaction was performed by the student (1) or the teacher (0).
- *Content of the interaction*: A categorical variable specifying the main focus of the interaction, among the following:
    - *Behaviour Management*: Interactions that focus on the rules of behaviour in the classroom, such as controlling and redirecting behaviour (e.g., calls to keep order, silence, etc.).
    - *Administrative*: Interactions that focus on administrative tasks or classroom management (e.g., taking attendance, announcing the structure of the class, announcing the activity to be carried out, etc.).

- *Task Instructions*: These interactions focus on the specific management of instructional activities, they inform how to carry out the activities without explaining the content (e.g., dictation of a guide, distributing groups for an activity, etc.)
- *Pedagogical*: Interactions that refer to the content or pedagogical skills addressed, they focus on the process of academic teaching (e.g. explaining the subject matter, verbally evaluating the contribution of a student, asking questions about the content addressed, etc.).

Distinguishing by content when analysing teacher-student interactions is particularly relevant in the Chilean context as, Chilean teachers declare to dedicate only 70% of their class time to teaching and learning (OECD, 2019), ranking fourth among the countries with the lowest percentage of time dedicated to those tasks. The remaining 30% of the time in the classroom is devoted to administrative tasks and maintaining discipline, activities that are not always considered in teacher-student interaction studies.

4.3.4.2 Student-level variables

- *Immigrant Origin:* This is a dichotomous variable indicating if the student as well as their father and their mother were born in Chile (0) or either the student, their father or their mother were born abroad (1). Therefore, in this study, we consider as students of immigrant origin those who are first- or second-generation immigrants.
- *Country of origin:* A series of dummy variables indicating the student's country of origin. Countries with less than 10 students in our sample were classified as 'Other countries'. Chile was used as the baseline category.
- *Language Spoken at Home:* A dichotomous variable distinguishing if the main language spoken at the student's home is Spanish (0) or other (1).
- *Female:* Gender is represented as a dichotomous variable that distinguishes boys (0) from girls (1). This is an important control variable as previous studies in the Chilean context have found that girls interact less frequently with their mathematics teacher (Ortega et al, in perss).
- *Mathematics Achievement*: The score on the standardized SEPA mathematics test, developed by the MIDE UC Measurement Center of the Pontificia Universidad Católica

de Chile (Manzi, García & Godoy, 2017). This standardised test is based on the national curriculum, consists of 40 multiple-choice items and presents satisfactory estimates of internal consistency (Cronbach's $\alpha > .85$). The scores on this test were standardised and centred on the group mean (this is, centred to the class's average achievement score).

- *Parents' University Education:* A dichotomous variable indicating if at least one of the student parents/guardians attended university (1) or not (0).
- *Years at School:* This variable indicates the number of years since the student entered their current school, and ranges from 0 to 6 years.
- *Sitting Row*: Corresponds to the row in which the student sat for most of the videotaped lesson, with values ranging from 1 (first row) to the total number of rows (8 at most, in this sample). We controlled for this variable as there is evidence suggesting that the physical distance of the student in the classroom from the teacher is related to class participation and academic achievement (Tagliacollo, Volpato, & Pereira, 2010).

4.3.4.3 Classroom-level variables

- *Proportion of Immigrant Students.* The number of students with immigrant background within the classroom divided by the total number of students in the classroom. This variable was centred to the mean.
- *Proportion of Immigrant Students*^{2}. The variable above was transformed to obtain its quadratic term, to assess non-linear effects of the immigrant composition of the classroom.
- *Teacher's attitude towards diversity in the classroom:* This scale was collected through the self-administered teacher questionnaire and consists of factor score calculated from four Likert scale items. Table A4-3 (Appendix H) lists the items that compose the scale, which were answered on a four-point scale and the response categories were 1 for "strongly disagree", 2 for "disagree", 3 for "agree", and 4 for "strongly agree". The scale showed high internal consistency in our sample (Cronbach's $\alpha = .88$). The higher the score in this scale, the more negative teachers' attitude towards diversity in the classroom are.

- *Classroom Size.* Number of students in the classroom. This was used as a control variable as it has been shown in the literature that it can affect teacher-student interactions (Blatchford, Bassett, & Brown, 2011).

4.3.5 Analysis

Teaching and learning are relational phenomena and research on these processes calls for theoretical perspectives and methodological approaches that go beyond the study of the individual and examine affiliations and interactions (ties) among actors within social structures. Due to this, educational research has started to rely more frequently on the analysis of social networks (Carolan, 2013), a conceptual framework and statistical technique used to study, quantify and visualize the connections and structures of relationships and interactions in organizations. This approach has been fruitfully applied to the study of the social inclusion of immigrants and minorities in educational settings (e.g., Author et al., 2015a; Moody, 2001).

In this study, the data derived from the coding process resembles partial ego-networks, with a focal node, "ego" (i.e., the teacher), and the nodes to whom ego is directly connected to, "alters" (i.e., the students). Teacher-student interaction networks were depicted via sociograms, which are graphical representations of social links, using the graphical capabilities of the *igraph* package in R (Csardi & Nepusz, 2006). Figure 4-1 shows the teacher-student interaction networks' sociograms for the 38 participant classrooms by students' country of origin.

Also, in-degrees (number of incoming ties) and out-degrees (number of outgoing ties) were calculated for each student and used as the *dependant variables* in this study, to investigate if immigrant students were significantly more peripheral to their class' teacher-student interaction networks than their non-immigrant peers. Given the non-parametric distribution of the frequency

of interactions (i.e., Poisson distribution), Mann-Whitney U tests and Poisson models were used to compare and model differences between immigrant and non-immigrant students.

Based on the review of the literature, it was decided to analyse separately the interactions initiated by students and the interactions initiated by the teachers. In addition, separate analyses were carried out for each type of interaction content. As depicted in Figure 4-2, the inclusion of immigrant students can vary by the initiator and content of the interactions considered. Therefore, we modelled separately the following seven types of interactions:

- Teacher-initiated interactions: Pedagogical, Administrative, Task Instructions, and, Behaviour Management.

Student-initiated interactions: Pedagogical, Administrative, and Task Instructions[19].

Then, given the hierarchical structure of the data, and to explicitly model this dependency, a multilevel approach (students nested within classrooms) was applied using a random intercept Poisson model with robust standard errors[20]. In the first model, named Model 1, only the fixed effect of the variable *Immigrant Origin* was included. Thus, as shown in Equation (1), the expected number of interactions with the teacher y_{ij} for student i in classroom j was specified as a log-linear model and a random classroom-level intercept u_{1j} was included.

Equation (1)

$$ln(y_{ij}) = \beta_1 + \beta_2 ImmigrantOrigin_{2ij} + u_{1j}$$

[19] Student-initiated Behaviour Management interactions were extremely rare and, therefore, not considered in this study

[20] Including a random intercept and using the sandwich estimator for the standard errors can, at least to some degree, address the common problem of overdispersion in count data (Rabe-Hesketh & Skrondal, 2008).

In this model, it is assumed that u_{1j} and the co-variable are independent, that u_{1j} are independent across the classrooms *j*, that the distribution of the random intercept is Gaussian and of variance $\sigma^2_{\mu 1}$, and that the conditional distribution of the dependent variable, given the random effect, is Poisson.

In Model 2, shown in Equation (2), the fixed effects of most of the student-level variables listed in Section 3.4.2, with the exception of the effects for country of origin, were included. This, to assess whether any significant effect of the variable *Immigrant* persisted or appeared after controlling for relevant student-level variables.

Equation (2)

$$ln(y_{ij}) = \beta_1 + \beta_2 ImmigrantOrigin_{2ij} + \beta_3 LanguageSpokenatHome_{3ij} + \beta_4 MathAchievement_{4ij} + \beta_5 Female_{5ij} + \beta_6 Parents'UniversityEducation_{6ij} + \beta_7 YearsinSchool_{7ij} + \beta_8 SittingRow_{8ij} + u_{1j}$$

Then, in Model 3, we assess if there are significant differences by students' country of origin, as shown in Equation (3).

Equation (3)

$$ln(y_{ij}) = \beta_1 + \beta_2 LanguageSpokenatHome_{2ij} + \beta_3 MathAchievement_{3ij} + \beta_4 Female_{4ij} + \beta_5 Parents'UniversityEducation_{5ij} + \beta_6 YearsinSchool_{6ij} + \beta_7 SittingRow_{7ij} + \beta_8 Peru_{8ij} + \beta_9 Bolivia_{9ij} + \beta_{10} Colombia_{10ij} + \beta_{11} Haiti_{11ij} + \beta_{12} Ecuador_{12ij} + \beta_{13} Venezuela_{13ij} + \beta_{14} Dominican\ Republic_{14ij} + \beta_{15} OtherCountries_{15ij} + u_{1j}$$

In a fourth model, the classroom-level variation on the effect of the student's immigrant background was explored. This is, we analysed whether some classrooms were significantly

more inclusive of immigrant students than others. For this, the goodness of fit of the random intercept model (Model 2), which assumes that the student immigrant-background effect on the frequency of interactions with the teacher is the same for all classrooms, was compared with that of a random slope model, which represents the differential effect of student immigrant background across classrooms (Model 4), by means of likelihood ratio tests. Thus, in Model 4, we introduced a random coefficient for the variable *Immigrant Origin* at the classroom level u_{2j}, as shown in Equation (4).

Equation (4)

$$ln(y_{ij}) = \beta_1 + \beta_2 ImmigrantOrigin_{2ij} + \beta_3 LanguageSpokenatHome_{3ij} + \beta_4 MathAchievement_{4ij} + \beta_5 Female_{5ij} + \beta_6 Parents'UniversityEducation_{6ij} + \beta_7 YearsinSchool_{7ij} + \beta_8 SittingRow_{8ij} + u_{1j} + u_{2j} Immigrant\ Origin_{2ij}$$

This specification allows the immigrant background effect $\beta_2 + u_{2j}$ to vary across classrooms j. We now assume that, given the covariates entered, the intercept and random coefficient have a normal bivariate distribution with zero mean and the following covariance matrix:

$$\begin{bmatrix} \sigma^2_{\mu 11} & \sigma^2_{\mu 12} \\ \sigma^2_{\mu 21} & \sigma^2_{\mu 22} \end{bmatrix}, \sigma_{21} = \sigma_{12}$$

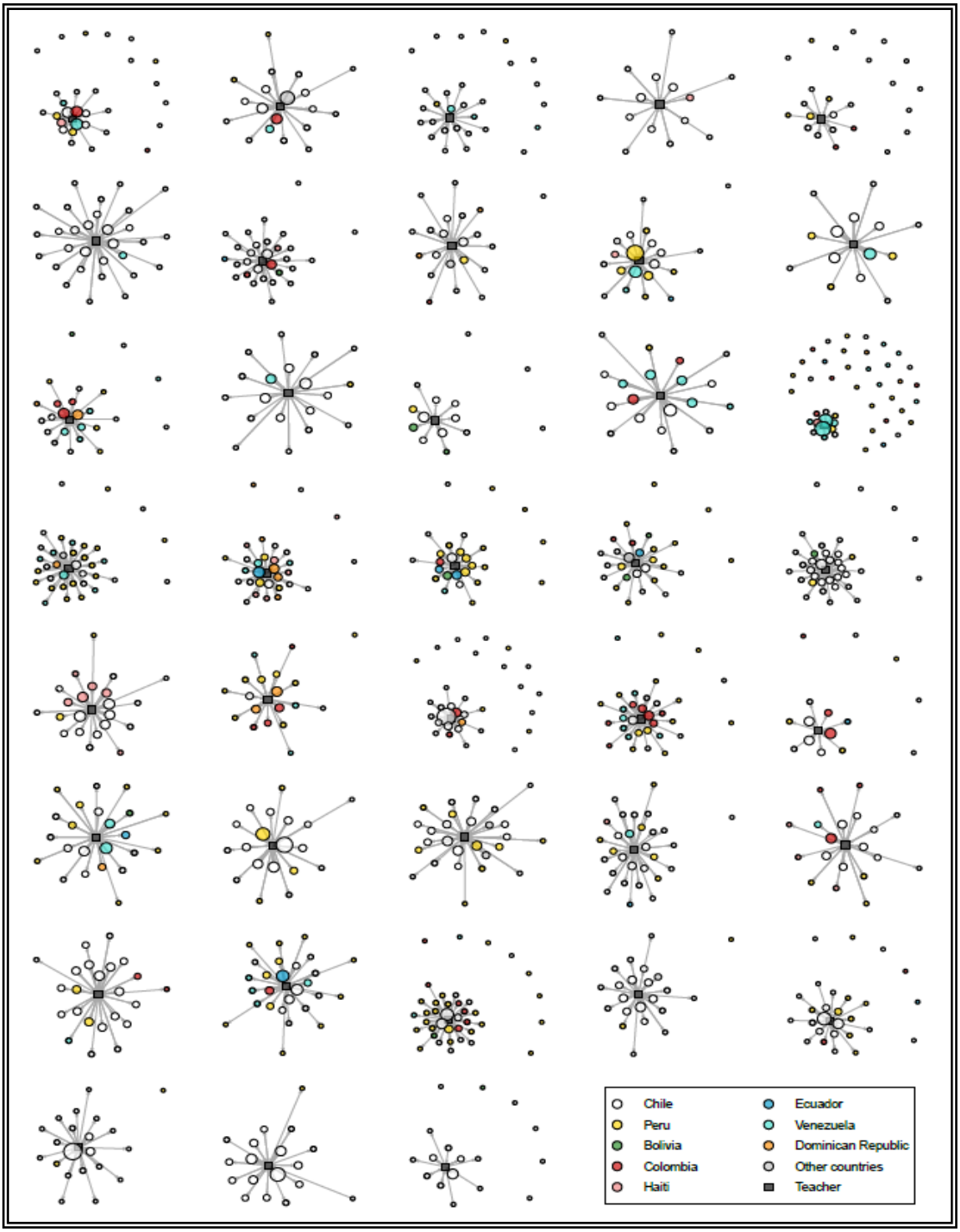

Figure 4-1. Teacher-student interaction networks in participant classrooms by students' country of origin (node size and arrow width weighted by degree of the student)

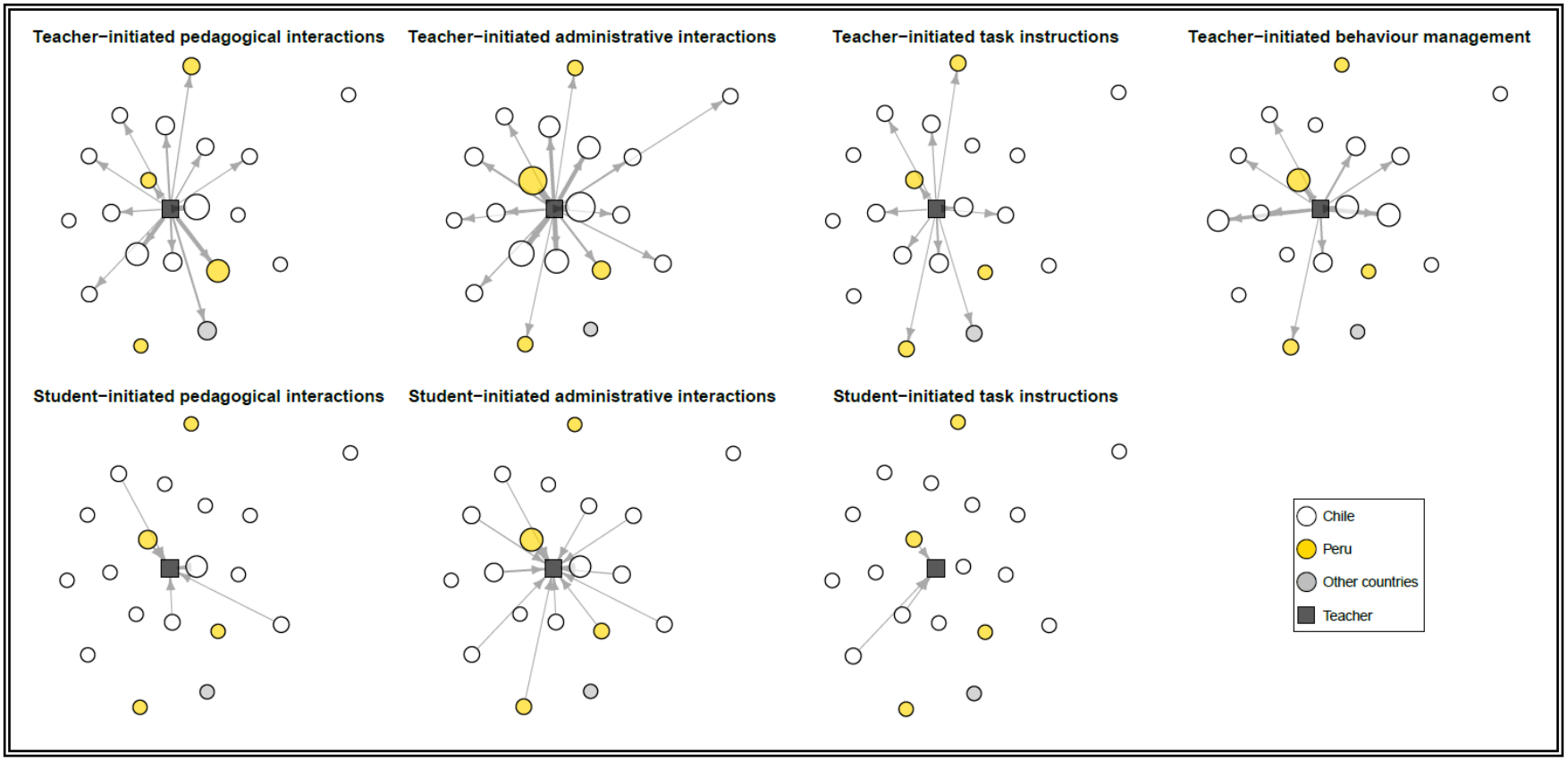

Figure 4-2. Teacher- and student-initiated interaction networks by interaction content and by students' country of origin in one of the participant classrooms (node size and arrow width weighted by in- and out-degree of the student, respectively).

Finally, in Model 5, we evaluate to what extent this variation is explained by the classroom-level variables listed in Section 3.4.3 by adding the fixed effects of these variables and the cross-level interactions between *Proportion of Immigrant Students*, *Proportion of Immigrant Students*^{2} and *Teacher's Attitude Towards Diversity* with the variable *Immigrant*. This specification is shown in Equation (5).

Equation (5)

$$ln(y_{ij}) = \beta_1 + \beta_2 ImmigrantOrigin_{2ij} + \beta_3 LanguageSpokenatHome_{3ij} + \beta_4 MathAchievement_{4ij} + \beta_5 Female_{5ij} + \beta_6 Parents'UniversityEducation_{6ij} + \beta_7 YearsinSchool_{7ij} + \beta_8 SittingRow_{8ij} + \beta_9 ProportionImmigrants_{9j} + \beta_{10} ProportionImmigrants^2{}_{10j} + \beta_{11} TeacherattitudeDiversity_{11j} + \beta_{12} Immigrant_{2ij} XProportionImmigrants_{9j} + \beta_{13} Immigrant_{2ij} XProportionImmigrants^2{}_{10j} + \beta_{14} Immigrant_{2ij} XTeacherattitudeDiversity_{11j} + \beta_{15} ClassSize_{15j} + u_{1j} + u_{2j} ImmigrantOrigin_{2ij}$$

These random-intercept and random-coefficient Poisson models were fitted via the *gllamm* command in Stata, using additional options to obtain exponentiated regression coefficients and robust standard errors (Rabe-Hesketh & Skrondal, 2008). The goodness of fit of these nested models was compared using likelihood-ratio tests.

4.4 Results

4.4.1 Differences in frequency of interactions with teachers by student immigrant origin

First, we depict and provide descriptive statistics on the sample classrooms. Figure 4-3 shows the teacher-student interaction networks for a random sample of 19 of the participant classrooms, considering separately teacher-initiated (sociogram on the right) and student-initiated (sociogram on the left) interactions. In these sociograms, teachers are the central grey squared node, students of immigrant origin are depicted in orange colour and non-immigrant students are shown in white. As in the previous figures, the size of each student node and the width of the arrows indicating the directionality of the interactions are based on the number of interactions between the teacher and the given student

Figure 4-3 shows that, in some of the networks, immigrant students are more peripheral to their teacher-student interaction networks than their peers. However, to tell if this difference corresponds to a systematic trend, it is necessary to conduct statistical tests.

In what follows, we present the results of the descriptive analysis and multilevel Poisson regression models for each type of interaction. These results are reported below in separate sections, according to the type of initiator of the interaction (i.e., the teacher or a student).

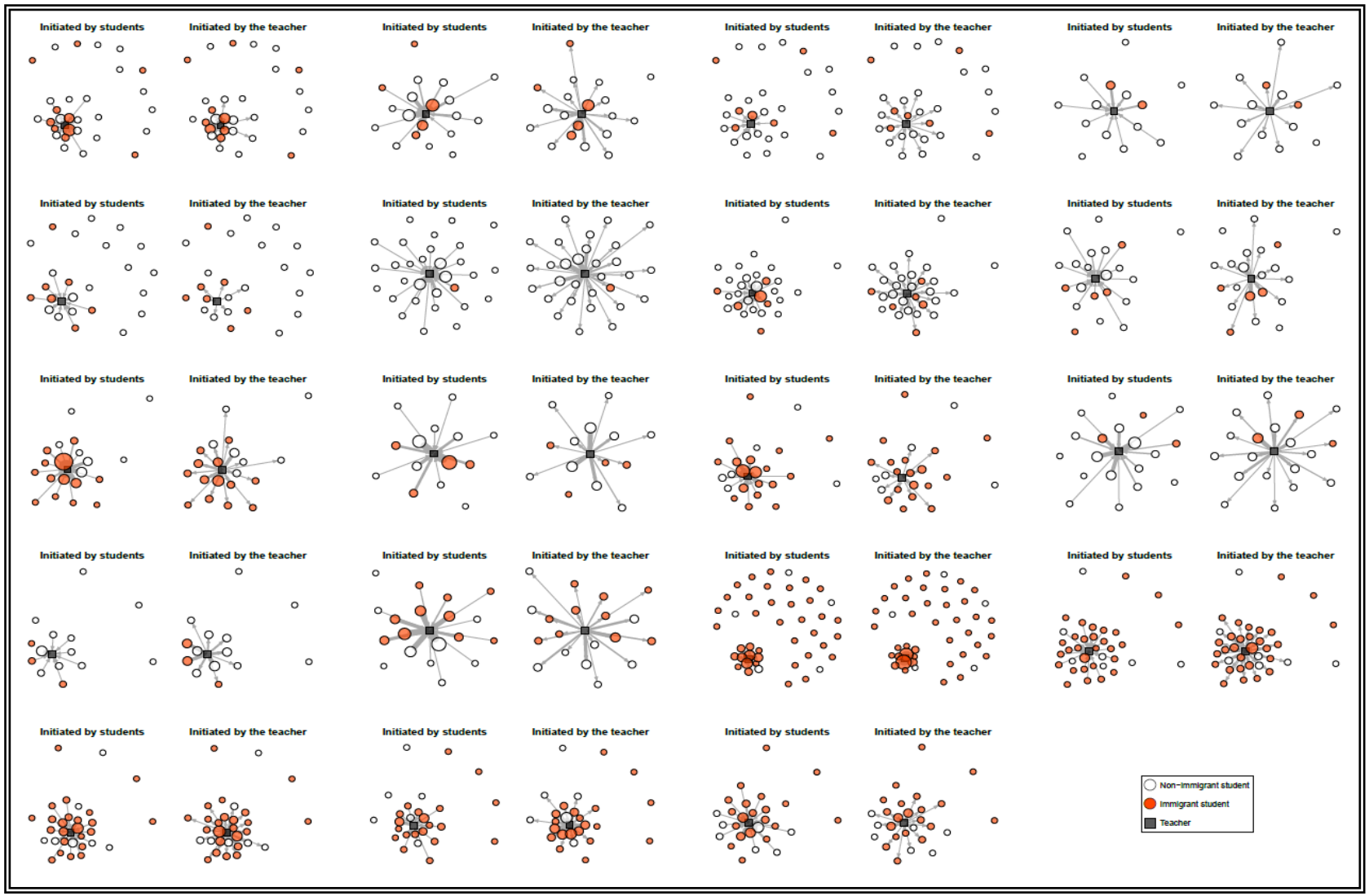

Figure 4-3. Teacher- and student-initiated interaction networks in 19 of the participant classrooms by students' immigrant background (node size and arrow width weighted by in- and out-degree of the student, respectively).

4.4.1.1 Teacher initiated interactions

As shown in Table 4-2, when comparing the average frequency of total teacher-initiated interactions, immigrant students were found to be significantly more peripheral to their class' teacher-student interaction networks than their non-immigrant peers, with an average of .42 less interactions (U = 97,343, p < .05). Interestingly, when analysing differences by *Language Spoken at Home*, we can see that students who report other languages than Spanish as their predominant language at home interact significantly more frequently with the teacher, in total teacher-initiated interactions (U = 12,607, p < .05).

Then, when looking specifically at *Country of Origin*, we observe that, overall, teachers tend to initiate significantly more interactions with students from Chile than with students from Peru (U = 43,980, p < .01) or Colombia (U = 10,871, p < .05). Conversely, teachers initiate significantly more interactions with students from Haiti than with students from Chile (U = 3,843, p < .05).

We now focus on the results of the multilevel Poisson regression models for teacher-initiated interactions, presented in Tables 4-4 and 4-5. The variance of the intercept was statistically significant for all adjusted models (p < .001). In addition, for Model 1, between 16 and 21% of the variance in the frequency of interactions of students with their teachers, initiated by the latter, is due to variation between classrooms[21].

[21] Therefore, it was decided to use and present only the results of the multilevel models. The variance partition coefficient was calculated as $\frac{\sigma_{\mu}^{2}}{\sigma_{\mu}^{2} + \pi^{2}/3}$, according to recommendations by Goldstein, Browne y Rasbash (2002).

Table 4-4. Results from multilevel Poisson models for teacher-initiated behaviour management and administrative interactions ($n_{classes}$ = 38; $n_{students}$ = 933).

	Behaviour Management Interactions					***Administrative Interactions***				
	MODEL 1	**MODEL 2**	**MODEL 3**	**MODEL 4**	**MODEL 5**	**MODEL 1**	**MODEL 2**	**MODEL 3**	**MODEL 4**	**MODEL 5**
	IR (SE)	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**
FIXED EFFECTS										
Intercept	.529 (.094)***	.939 (.304)	.995 (322)	1.008 (.334)	.717 (.565)	.807 (.128)	1.060 (.225)	.992 (.216)	.971 (.244)	1.117 (.635)
Student-Level Variables										
Immigrant	.858 (.098)	.835 (.098)		.721 (.116)*	.732 (.153)	1.046 (.101)	1.022 (.094)		1.151 (.178)	1.346 (.295)
Language Spoken at Home		1.257 (.250)	1.188 (.314)	1.172 (.266)	1.157 (.279)		1.292 (.309)	.970 (.226)	1.242 (.294)	1.241 (.294)
Math Achievement		.901 (.054)†	.902 (.055)†	.905 (.053)†	.902 (.054)†		1.015 (.047)	1.026 (.049)	1.014 (.048)	1.013 (.048)
Female		.494 (.081)***	.501 (.084)***	.491 (.081)***	.490 (.080)***		.882 (.085)	.910 (.081)	.891 (.088)	.891 (.088)
Parents' University Education		1.173 (.026)***	1.193 (.028)***	1.163 (.027)***	1.173 (.034)***		1.072 (.056)	1.060 (.059)	1.067 (.054)	1.071 (.054)
Years in the School		1.009 (.036)	.997 (.036)	1.007 (.035)	1.008 (.036)		1.001 (.019)	1.011 (.020)	1.002 (.020)	1.002 (.020)
Sitting Row		.885 (.037)**	.881 (.037)**	.885 (.037)**	.884 (.038)**		.926 (.027)**	.928 (.029)*	.928 (.027)*	.927 (.027)**
Peru			.727 (.114)*					.841 (.088)†		
Bolivia			1.410 (.357)					1.170 (.354)		
Colombia			.693 (.187)					1.001 (.163)		

Haiti			.719 (.310)					1.890 (.387)**		
Ecuador			1.843 (.924)					2.143 (.735)*		
Venezuela			.638 (.151)†					1.282 (.347)		
Dominican Republic			.710 (.573)					1.395 (.359)		
Other Country			1.541 (.384)†					1.165 (.422)		
Classroom-Level Variables										
Proportion of Immigrant Students					.629 (.410)					.547 (.352)
Proportion of Immigrant Students^2					.273 (.746)					20.010 (51.333)
Immigrant X Proportion of Immigrant Students					1.674 (.966)					1.186 (.668)
Immigrant X Proportion of Immigrant Students^2					.503 (1.207)					.100 (.207)
Teacher Attitude Diversity					.998 (.152)					1.060 (.129)
Immigrant X Teacher Attitude Diversity					.962 (.104)					.889 (.097)
Class Size					1.016 (.029)					.986 (.018)
RANDOM EFFECTS										
Variance (Intercept)	.882 (.282)	.884 (.276)	.876 (.270)	.742 (.227)	.735 (.227)	.629 (.185)	.646 (.185)	.662 (.189)	.859 (.321)	.785 (.288)
Variance (Immigrant)				.044 (.056)	.051 (.062)				.130 (.124)	.089 (.105)
CORR (Intercept, Immigrant)				.420	.441				-.714*	-.759*
Log likelihood	-1,149.40	-1,096.82	-1,086.30	-1,095.42	-1,094.30	-1,260.84	-1,249.53	-1,236.68	-1,247.08	-1,244.94

LR chi^2	-	105.16***	21.03**	2.81	2.23	-	22.61***	25.71**	4.91†	4.28

Notes: IR: Incidence Ratio. SE: Standard Error. CORR: Correlation. † $p < .10$: * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 4-5. Results from multilevel Poisson models for teacher-initiated task instruction and pedagogical interactions ($n_{classes}$ = 38; $n_{students}$ = 933).

	Task Instruction Interactions					***Pedagogical Interactions***				
	MODEL 1	**MODEL 2**	**MODEL 3**	**MODEL 4**	**MODEL 5**	**MODEL 1**	**MODEL 2**	**MODEL 3**	**MODEL 4**	**MODEL 5**
	IR (SE)	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**
FIXED EFFECTS										
Intercept	.089 (.020)***	.154 (.063)***	.132 (.062)***	.147 (.066)***	.307 (.259)	.843 (.131)	.970 (.176)	.840 (.152)	.881 (.176)	1.832 (.873)
Student-Level Variables										
Immigrant	.949 (.229)	.920 (.229)		.809 (.355)	.383 (.278)	.945 (.121)	.891 (.107)		1.006 (.136)	1.080 (.212)
Language Spoken at Home		.913 (.655)	.616 (.342)	.952 (.664)	1.044 (.569)		1.437 (.272)†	1.329 (.236)	1.486 (.290)*	1.500 (.306)*
Math Achievement		.942 (.085)	.943 (.091)	.939 (.087)	.942 (.091)		1.059 (.044)	1.058 (.044)	1.068 (.043)	1.065 (.044)
Female		.716 (.140)†	.770 (.161)	.724 (.107)	.727 (.141)		.754 (.089)*	.775 (.089)*	.762 (.091)*	.757 (.090)*
Parents' University Education		1.070 (.132)	1.090 (.129)	1.080 (.139)	1.099 (.169)		1.148 (.063)*	1.142 (.051)**	1.150 (.067)*	1.158 (.068)*
Years in the School		.999 (.043)	1.023 (.050)	.995 (.043)	.997 (.042)		.992 (.017)	1.012 (.017)	.988 (.017)	.987 (.017)
Sitting Row		.878 (.058)†	.878 (.061)†	.882 (.060)†	.893 (.054)†		.991 (.030)	.994 (.030)	.996 (.030)	.997 (.030)
Peru			.541 (.154)*					.711 (.105)*		
Bolivia			1.333 (1.174)					1.101 (.254)		
Colombia			.864 (.436)					1.014 (.273)		
Haiti			1.031 (.639)					1.047 (.233)		

Ecuador			5.058 (2.115)***					1.558 (.437)		
Venezuela			1.798 (.899)					1.536 (.276)*		
Dominican Republic			1.087 (1.073)					1.416 (.420)		
Other Country			1.086 (.544)					1.130 (.333)		
Classroom-Level Variables										
Proportion of Immigrant Students					.748 (.732)					1.636 (.834)
Proportion of Immigrant Students^2					.001 (.003)†					.948 (2.375)
Immigrant X Proportion of Immigrant Students					3.020 (4.087)					1.623 (1.078)
Immigrant X Proportion of Immigrant Students^2					125,044.6 (749,059.8)*					.193 (.501)
Teacher Attitude Diversity					1.374 (.328)					1.400 (.183)*
Immigrant X Teacher Attitude Diversity					.935 (.291)					.790 (.094)†
Class Size					.983 (.032)					.970 (.017)†
RANDOM EFFECTS										
Variance (Intercept)	.854 (.352)	.882 (.355)	.904 (.376)	.897 (.387)	.609 (.305)	.603 (.147)	.575 (.144)	.577 (.141)	.818 (.226)	.595 (.231)
Variance (Immigrant)				.716 (.867)	.063 (.152)				.302 (.140)	.250 (.104)
CORR (Intercept, Immigrant)				-.253	-.130				-.657*	-.594*
Log likelihood	-355.57	-351.40	-340.09	-350.57	-344.76	-1,285.44	-1,286.74	-1,268.96	-1,276.89	-1,271.04

LR chi^2	-	8.34	22.62**	1.66	11.61	-	38.35***	35.56***	19.72***	11.68

Notes: IR: Incidence Ratio. SE: Standard Error. CORR: Correlation. † $p < .10$: * $p < .05$; ** $p < .01$; *** $p < .001$.

Models 1 and 2 evaluate Hypothesis 1a. This is, if students with a migrant background are more peripheral, in terms of interactions with their teacher, than their non-migrant peers. The results of Model 1, for each type of interaction, show that, after the dependencies in the data (given its nested structure) are considered, there are no statistically significant differences in the average number of teacher-initiated interactions by student immigrant background. Furthermore, after we control for individual characteristics, in Model 2, student immigrant background remains as a non-significant predictor of teacher-initiated interactions[22]. Thus, there is no evidence to support Hypothesis 1a, in the case of teacher-initiated interactions.

However, when we introduce the country of origin dummy variables to evaluate the Hypothesis 1b, in Model 3, and control for the other variables in the model, interesting patterns appear. In particular, we observe that Peruvian students are significantly less frequently included in teacher-initiated interactions assessed, with 27% fewer behaviour-management interactions ($p < .05$), 16% fewer administrative interactions ($p < .10$), 46% fewer task-instruction interactions ($p < .05$) and 29% fewer pedagogical interactions ($p < .05$) than their non-immigrant peers. Haitian students, in turn, are approached by their teacher 89% more frequently in administrative interactions than their non-immigrant peers ($p < .01$). Similarly, Ecuadorian students are approached by their teacher 114% more frequently in administrative interactions ($p < .05$), and

[22] As shown in the results for Model 2, applied to teacher-initiated interactions, in Tables 4-4 and 4-5, other student-level factors positively associated with more frequent pedagogical teacher-initiated interactions are speaking a language other than Spanish at home ($p < .10$), being male ($p < .05$) and having at least one parent with university studies ($p < .05$). Student-level factors positively associated with more frequent teacher-initiated behaviour-management interactions are having a lower mathematics achievement level within the class ($p < .10$), being male ($p < .001$), having at least one parent with university studies ($p < .001$) and sitting in rows closer to the front of the classroom ($p < .01$). With regard to administrative teacher-initiated interactions, the only significant student-level predictors found was sitting in rows closer to the front of the classroom ($p < .01$). Finally, significant

also 406% more frequently in task-instruction interactions ($p < .001$). Finally, students from Venezuela show 36% fewer teacher-initiated behaviour-management interactions ($p < .10$) and 54% more pedagogical interactions ($p < .05$) than their non-immigrant peers. Furthermore, model fit improves significantly with the addition of Country of Origin dummy variables, in comparison to Model 2, for all of the teacher-initiated interactions analysed. Thus, we can confirm Hypothesis 1b, when analysing teacher-initiated interactions.

4.4.1.2 Student initiated interactions

As shown in Table 4-3, the average frequency of total student-initiated interactions does not vary by student immigrant background. However, when looking specifically at *Country of Origin*, we find that, on average, students from Chile tend to initiate .78 more interactions with their mathematics teacher that students from Peru ($U = 46{,}472$, $p < .05$). On the other hand, students from Colombia initiate, on average, 1.85 more interactions in with the teacher than students from Chile ($U = 10{,}662$, $p < .05$), as well as students from Venezuela, who initiate, on average, 1.61 more interactions in with the teacher than students from Chile ($U = 13{,}265$, $p < .01$).

When analysing student-initiated interactions using a random intercept Poisson regression model (Tables 4-6 and 4-7), it was again found that the variance of the intercept was statistically significant for all the fitted models ($p < .001$). In addition, for Model 1, between 9 and 22% of the variance in the frequency of student-initiated interactions resides between classrooms.

The results of Model 1 in Tables 4-6 and 4-7, for student-initiated interactions, show that immigrant origin predicts 29% more pedagogical interactions ($p < .05$), but is not associated with

---

predictors of more frequent teacher-initiated task-instruction interactions with the teacher are being male ($p < .10$) and sitting in rows closer to the front of the classroom ($p < .10$).

differential administrative or task-instructions student-initiated interactions. The significant effect of immigrant origin for student-initiated pedagogical interactions disappears after controlling for the individual-level variables in Model 2. Thus, we cannot confirm Hypothesis 1a in the case of student-initiated interactions.[23]

Model 3 evaluates the Hypothesis 1b that looks for differences by students' country of origin, while still controlling for relevant student-level variables. As shown in Table 4-6, the results coincide with those for teacher-initiated interaction, as they indicate that Peruvian students are more peripheral in all the student-initiated interactions studied. This is, they are significantly less likely to approach the teacher, with 38% fewer administrative interactions ($p < .10$), 61% fewer task-instruction interactions ($p < .01$) and 23% fewer pedagogical interactions ($p < .10$) initiated by them than by their non-immigrant peers[24].Also, while Colombian students show 136% more student-initiated task-instruction interactions ($p < .05$), the opposite is true for Haitian students, who show 44% fewer student-initiated task-instruction interactions ($p < .001$), than non-

[23] As shown in the results for Model 2, applied to student-initiated interactions, in Tables 4-6 and 4-7, other student-level factors positively associated with more frequent student-initiated pedagogical interactions with the teacher are having a higher mathematics achievement level within the class ($p < .001$) and sitting in rows closer to the front of the classroom ($p < .001$). With regard to student-initiated administrative interactions with the teacher, significant student-level predictors are having at least one parent with university studies ($p < .10$), being newer to the school ($p < .10$) and sitting in rows closer to the front of the classroom ($p < .001$). Finally, significant predictors of more frequent student-initiated task-instruction interactions with the teacher are having a higher mathematics achievement level within the class ($p < .10$), being male ($p < .01$) and sitting in rows closer to the front of the classroom ($p < .001$).

[24] While not the focus of this article, we assessed if the inclusion of Peruvian students varied by their immigration generation, by adding an interaction effect between the Peru dummy variable and a variable indicating if students were first (0) or second (1) generation immigrants. It was found that the frequency of interactions varied for teacher-initiated behaviour management interactions, where second-generation Peruvian students were approach by their teacher significantly more frequently than first-generation Peruvian students ($p < .05$), and for student-initiated pedagogical interactions, where second-generation Peruvian students approach their teacher significantly less frequently than first-generation Peruvian students ($p < .05$). These results are available upon request. We were unable to conduct this analysis for the other countries of origin due to their limited sample size. Indeed, the immigrant generation effect is collinear with that of country of origin.

immigrant students. With regard to student-initiated pedagogical interactions, students from Colombia, Venezuela and Dominican Republic approach the teacher significantly more frequently, with 143% ($p < .001$), 110% ($p < .01$) and 68% ($p < .10$) more interactions, respectively, that their non-immigrant peers. Finally, Model 3 fits the data significantly better than Model 2 for the three student-initiated interactions analysed, indicating that the students' country of origin is a relevant predictor of frequency of interactions with the teacher. We can, therefore, confirm Hypothesis 1b for student-initiated interactions.

Table 4-6. Results from multilevel Poisson models for student-initiated administrative and task-instruction interactions ($n_{classes}$ = 38; $n_{students}$ = 933).

	Administrative Interactions					***Task-Instruction Interactions***				
	MODEL 1	**MODEL 2**	**MODEL 3**	**MODEL 4**	**MODEL 5**	**MODEL 1**	**MODEL 2**	**MODEL 3**	**MODEL 4**	**MODEL 5**
	IR (SE)	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**
FIXED EFFECTS										
Intercept	.689 (.079)**	1.906 (.524)*	1.832 (.511)*	1.905 (.528)*	2.750 (1.035)*	.088 (.021) ***	.326 (.143)*	.315 (.127)**	.392 (.153)*	.532 (.391)
Student-Level Variables										
Immigrant	.932 (.150)	.801 (.116)		.779 (.113)†	.989 (.189)	1.071 (.311)	1.114 (.396)		.662 (.237)	1.010 (.472)
Language Spoken at Home		1.084 (.309)	1.130 (.348)	.996 (.299)	.968 (.218)		.932 (.741)	1.484 (.915)	1.002 (.830)	.967 (.820)
Math Achievement		1.068 (.060)	1.063 (.060)	1.066 (.060)	1.065 (.042)		1.237 (.140)†	1.207 (.132)†	1.225 (.142)†	1.216 (.144)†
Female		.882 (.163)	.895 (.159)	.875 (.161)	.874 (.073)		.444 (.122)**	.495 (.120)**	.439 (.122)**	.436 (.120)**
Parents' University Education		1.083 (.045)†	1.085 (.042)*	1.083 (.047)†	1.083 (.042)*		.659 (.198)	.640 (.179)	.614 (.218)	.614 (.218)
Years in the School		.931 (.034)†	.937 (.035)†	.930 (.035)†	.929 (.035)†		.994 (.061)	1.018 (.060)	.994 (.060)	.992 (.062)
Sitting Row		.801 (.036)***	.800 (.036)***	.805 (.036)***	.807 (.021)***		.729 (.049)***	.733 (.047)***	.728 (.050)***	.726 (.050)***
Peru			.617 (.161)†					.386 (.107)**		
Bolivia			.843 (.209)					.664 (.761)		
Colombia			1.144 (.265)					2.362 (.852)*		
Haiti			.555 (.232)					.463		

								(.128)***		
Ecuador			1.050 (.697)					1.919 (1.111)		
Venezuela			.956 (.302)					1.761 (.995)		
Dominican Republic			1.405 (.456)					1.056 (.822)		
Other Country			1.079 (.433)					.540 (.460)		
Classroom-Level Variables										
Proportion of Immigrant Students					.863 (.387)					1.029 (.844)
Proportion of Immigrant Students^2					1.262 (2.527)					30.275 (90.086)
Immigrant X Proportion of Immigrant Students					.873 (.512)					3.717 (4.666)
Immigrant X Proportion of Immigrant Students^2					.063 (.152)					.001 (.003)
Teacher Attitude Diversity					.983 (.100)					1.098 (.233)
Immigrant X Teacher Attitude Diversity					1.138 (.129)					1.019 (.338)
Class Size					.988 (.015)					.979 (.029)
RANDOM EFFECTS										
Variance (Intercept)	.312 (.074)	.290 (.071)	.304 (.073)	.282 (.082)	.274 (.093)	.918 (.315)	1.077 (.426)	.893 (.281)	.665 (.291)	.622 (.277)
Variance (Immigrant)				.316 (.156)	.206 (.121)				.714 (.547)	.621 (.512)
CORR (Intercept, Immigrant)				-.275	-.441				.474	.388
Log likelihood	-1,233.98	-1,175.12	-1,165.57	-1,169.03	-1,165.03	-380.81	-354.04	-339.31	-350.29	-348.93
LR chi^2	-	115.72***	21.10**	14.18***	8.00	-	53.54***	29.46***	7.50*	2.72

Notes: IR: Incidence Ratio. SE: Standard Error. CORR: Correlation. † $p < .10$: * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4-7. Results from multilevel Poisson models for student-initiated pedagogical interactions ($n_{classes}$ = 38; $n_{students}$ = 933).

	Pedagogical Interactions				
	MODEL 1	**MODEL 2**	**MODEL 3**	**MODEL 4**	**MODEL 5**
	IR (SE)	**IR (SE)**	**IR (SE)**	**IR (SE)**	**IR (SE)**
FIXED EFFECTS					
Intercept	1.149 (.128)	2.190 (.402)***	1.863 (.360)**	2.163 (.415)***	4.317 (2.151)**
Student-Level Variables					
Immigrant	1.294 (.138)*	1.177 (.130)		1.057 (.140)	.958 (.208)
Language Spoken at Home		1.212 (.293)	1.440 (.392)	1.169 (.286)	1.181 (.293)
Math Achievement		1.293 (.080)***	1.304 (.085)***	1.300 (.082)***	1.299 (.082)***
Female		.930 (.106)	.958 (.114)	.932 (.105)	.924 (.103)
Parents' University Education		1.003 (.048)	.975 (.054)	1.013 (.050)	1.012 (.049)
Years in the School		.974 (.027)	1.003 (.027)	.975 (.028)	.974 (.028)
Sitting Row		.848 (.036)***	.854 (.036)***	.854 (.036)***	.856 (.036)***
Peru			.768 (.118)†		
Bolivia			.702 (.260)		
Colombia			2.429 (.522)***		
Haiti			.817 (.366)		
Ecuador			.754 (.294)		
Venezuela			2.097 (.517)**		
Dominican Republic			1.676 (.453)†		
Other Country			1.108 (.385)		

Classroom-Level Variables					
Proportion of Immigrant Students					1.021 (.465)
Proportion of Immigrant Students^2					2.089 (5.056)
Immigrant X Proportion of Immigrant Students					1.434 (.736)
Immigrant X Proportion of Immigrant Students^2					3.848 (10.157)
Teacher Attitude Diversity					1.109 (.125)
Immigrant X Teacher Attitude Diversity					1.119 (.186)
Class Size					.970 (.019)
RANDOM EFFECTS					
Variance (Intercept)	.425 (.096)	.395 (.084)	.382 (.081)	.359 (.086)	.308 (.073)
Variance (Immigrant)				.234 (.111)	.239 (.105)
CORR (Intercept, Immigrant)				-.031	-.040
Log likelihood	-1,918.72	-1,825.38	-1,758.15	-1,815.45	-1,810.85
LR chi^2	-	186.69***	134.46***	19.85***	9.20

Notes: IR: Incidence Ratio. SE: Standard Error. CORR: Correlation. † $p < .10$: * $p < .05$; ** $p < .01$; *** $p < .001$.

4.4.2 Variation in the inclusion of immigrant students across classrooms

Hypothesis 2 states that there is significant variation in the effect of student immigrant background, on the frequency of interactions with the teacher, across classrooms. Figure 4-3 provides a first indication of variation in this regard across classes. We formally tested the significance of this classroom-level variation, by comparing the fit of Models 2 and 4. This is, we compare the fit of the random intercept only Poisson model with that of the random coefficient (for the *Immigrant* variable) Poisson model.

As shown in Tables 4-4 and 4-5, when considering teacher-initiated interactions, Model 4 fits the data significantly better than Model 2 only in administrative and pedagogical interactions ($\chi^2(1) = 4.91$, $p < .10$ and $\chi^2(1) = 19.72$, $p < .001$, respectively). In the case of student-initiated interactions, Model 4 fits the data significantly better than Model 2 in all the types of interactions analysed; this is, administrative ($\chi^2(1) = 14.18$, $p < .001$), task-instruction ($\chi^2(1) = 7.50$, $p < .05$) and pedagogical ($\chi^2(1) = 19.85$, $p < .001$) interactions (see Tables 4-6 and 4-7). Thus, we find evidence to support Hypothesis 2; some classrooms are significantly more inclusive of immigrant students than others, particularly in terms of student-initiated interactions and in teacher-initiated administrative and pedagogical interactions.[25]

[25] As shown in Tables 4-4 and 4-5, the high and statistically significant, negative correlations between the intercept and the random coefficient in Model 5 for teacher-initiated administrative and pedagogical interactions ($r_p = -.71$, $p < .05$ and $r_p = -.66$, $p < .05$, respectively) are also interesting. This trend indicates that, in classrooms with a greater number of teacher-initiated administrative and pedagogical interactions, there is a lower effect of student's immigrant origin.

4.4.3 Predictors of the variation in the inclusion of immigrant students across classrooms

Hypothesis 3 states that the classroom-level variation on the effect of student immigrant background is partially explained by the immigrant composition of the classroom and the teacher's attitude towards diversity in the classroom. In Model 5, the main effects of the variables *Proportion of Immigrant Students*, *Proportion of Immigrant Students*^{2} and *Teacher's attitude towards diversity in the classroom* are included, as well as the cross-level interactions between these variables and the student-level variable *Immigrant Origin*. We also controlled for *Classroom Size*[26].

As shown in Table 4-5, when considering teacher-initiated task-instructions interactions, we find a significant positive interaction effect between student *Immigrant Origin* and *Proportion of Immigrant Students*^{2} ($p < .05$), after controlling for the rest of the variables in the model. This suggests that there is quadratic (U-shaped) dependence of the number of teacher-initiated task-instructions interactions for a migrant child on the proportion of migrants in the classroom. In this case, an immigrant student is approached more frequently by the teacher in task-instructions interactions when the proportion of immigrant students in the classroom is low and high, as opposed to medium.

Also, when considering pedagogical teacher-initiated interactions, we find a significant interaction effect between student *Immigrant Origin* and *Teacher's Attitude towards Diversity in the Classroom* ($p < .10$), indicating that teachers with more negative attitudes towards diversity in the classroom approach immigrant students less frequently for pedagogical purposes than teachers who report more positive attitudes towards diversity. However, when we consider

behaviour management and administrative teacher-initiated interactions, none of the interaction effects between student immigrant origin and the classroom-level variables are statistically significant (see Table 4-4). Furthermore, the inclusion of the classroom-level variables and cross-level interaction effects does not significantly improve model fit for any of the teacher-initiated interactions studied.

The results of Model 5 for student-initiated interactions, in Tables 4-6 and 4-7, show that none of the cross-level interactions are statistically significant after controlling for the rest of the variables in the model. Also, Model 5 does not fit the data better than Model 4 in any of the student-initiated interactions modeled.

Thus, we find little support for Hypothesis 3 in our data: the classroom-level variation on the effect of student immigrant background is partially explained by the proportion of immigrant students in the classroom only in teacher-initiated task-instruction interactions, and by the teacher's attitude towards diversity in the classroom only in teacher-initiated pedagogical interactions.

4.5 Discussion

This study investigated the inclusion of immigrant students at 38 classrooms in Chilean public schools. Overall, the immigrant background was not a significant predictor of the frequency of teacher-student interactions, especially after controlling for relevant student-level variables, such as mathematics achievement level within the class, parental educational level and the position of the student in the classroom. However, our detailed analysis suggests that the situation of immigrant students heavily depends on their country of origin.

---

[26] The correlations between these classroom-level variables do not suggest problems of collinearity,see Appendix I.

Worryingly, students from Peru were consistently more peripheral to their teacher-student networks than their non-immigrant peers across all of the interactions assessed. This is in line with previous Chilean studies suggesting that Peruvian students are perceived as lower status immigrants and are discriminated against on the basis of their phenotype, skin color and way of speaking (Mondaca et al., 2018; Pavez, 2012). On the other hand, students from Caribbean countries (Venezuela, Colombia and Dominican Republic) are particularly active when it comes to initiating pedagogical interactions with their mathematics teacher. In the case of students from Venezuela, they are also more frequently addressed by the teacher in pedagogical interactions and less frequently addressed in behavior-management interactions. Finally, students from Haiti and Ecuador are more frequently approached by their teacher than their non-immigrant peers, but only in administrative or task-instruction interactions.

Although beyond this study focus, a possible explanation for the differential inclusion of immigrant students by country of origin in the literature suggests that, as immigration grows in numbers and diversity, immigrant students are compared against each other and are perceived differently by the school community (Cerón, Pérez Alvarado, & Poblete, 2017).

In addition, we identified significant variation at the classroom level in the effect of student immigrant-background on the frequency of interactions with the teacher, for five of the seven types of interactions investigated, which include both teacher- and student-initiated pedagogical interactions, with some classrooms being significantly more inclusive of immigrant students than others. This variation was not consistently explained by the immigrant composition of the classroom, as expected from the literature. However, we explored the relationship between

teachers' attitudes towards diversity in the classroom and found that teachers with more negative attitudes approached immigrant students significantly less frequently for pedagogical purposes.

4.6 Conclusions

Worldwide, increased diversity in the student population is becoming an enormous challenge for teachers, schools and education systems. Recent influxes of immigration in different regions of the world are putting pressure on both the school system and the research agenda to better understand how to pedagogically deal with this diversity in effective and inclusive ways. This study proposes a new way of looking at inclusion in the classroom by exploring teacher-student interaction networks within school classes. The study advances the field by (1) exploring the educational inclusion of immigrant students in an emerging country with an accelerated and unprecedented trend of immigration, (2) demonstrating the combined use of systematic classroom observation, descriptive social network analysis and multilevel models for investigating equality in learning opportunities, and (3) contributing with further evidence on the within- and between-classroom inclusion of immigrant students and their predictors. Thus, the study is significant in presenting a new and fertile approach to important questions about inclusion in the classroom and teacher-student interactions, which stand in need of better research.

The growing immigration trends around the world will require schools and teachers to work in diverse classrooms. Furthermore, in the case of Chile the recently passed Inclusion Law (20.845; 2015) eliminates the barriers that limited equitable access of students to all state-subsidized schools. The reform removes economic barriers and school's arbitrary selection of students, and is expected to increase social and ethnic diversity within schools and classrooms. In this new

context, Chilean teachers face a significant challenge: to provide equal learning opportunities to students regardless of their background.

The international community states that immigrant inclusion should be at the centre of education policies and systems (UNESCO, 2018). The results presented suggest that particular emphasis should be put on initiatives that support immigrant groups that are at high risk of pedagogical exclusion (e.g., Peruvian students).

Previous evidence shows that Chilean teachers are not well equipped to work with diversity. Our study also suggests that teachers differ in the extent to which they include immigrant students: due to the lack of formal training in this direction, their personal attitudes matter. This suggests that it would be important to develop tools that help teachers better understand how they can improve the quality of their interactions and their distribution among the students in the classroom. There is also a need for interventions that improve teachers' attitudes towards diversity, that help them to become aware of their biases and that develop skills and strategies for working in diverse classrooms. The results are highly relevant for advancing our understanding of educational inclusion in classrooms, fostering school equity policies and teacher development programmes, both in Chile and internationally.

As with any piece of research, this study has limitations that are important to discuss. Firstly, it should be noted that these conclusions cannot be generalized to all types of classrooms/schools in (or outside of) Chile nor to all types of teacher-student interactions. Indeed, similar to the majority of international studies on teacher-student interactions, we worked with volunteer schools and teachers, which may lead to a selection of teachers more open to evaluation and

research, and to classrooms in which teacher-student relationships are more positive and less biased toward certain groups of students (den Brok & Levy, 2005). If so, the estimated differences in the inclusion of students by country of origin are conservative in relation to those that could be found in a representative sample of Chilean classrooms.

Secondly, we analysed differential patterns of teacher-student interactions by immigrant background but did not look at culturally responsive teaching, another important approach to assess immigrant educational inclusion (e.g., Civitillo, Juang, Badra, & Schachner, 2019; Jensen, Grajeda, & Haertel, 2018). Indeed, Chilean scholars have also called for intercultural education approaches and programs to address the needs of immigrant children and promote their social and educational inclusion (Riedemann & Stefoni, 2015; Stefoni et al., 2016).

Thirdly, while our study considers dependencies in the data that are related to its nested structure (i.e., students nested within classrooms) by using appropriate multilevel models, it ignores other types of dependencies. Most importantly, we assumed that within each classroom, interactions happen independent of each other. This is most likely not the case: those who already interacted with the teacher once may repeat this more easily, or those approached by the teacher may be more likely to initiate interactions themselves in the future. Teachers may specifically approach those students who seem passive in the classroom. These and similar tendencies could be modelled using statistical techniques that take the timing of events into account and model them based on patterns of past events. An example of such an approach, applied to (student-student) classroom dynamics data, can be found in DuBois et al. (2013). Unfortunately, existing methods, such as relational event models (e.g. Butts, 2008) or dynamic network actor models (e.g. Stadtfeld, Hollway & Block, 2017) cannot be applied to our data (i.e., network data with two

types of nodes: teachers and students, where there is only one teacher) in their current forms in a straightforward way. Finally, the small sample sizes for some countries of origin and at the limited number of classrooms studied might have prevented us from finding other important associations.

The study also provides important implications for research. It supports the importance of using a disaggregated perspective to the study of educational processes. It also highlights the relevance of distinguishing by initiator (directionality) and content of interaction when studying teacher-student relationship. Finally, it shows that studying the inclusion of immigrant students by treating this group as a whole can hide important differences, and it is necessary to distinguish by country of origin, as well as controlling for important student-level and contextual characteristics.

In the future, this approach could be extended by assessing the inclusion of other traditionally excluded or underperforming student groups (e.g., students with special educational needs, indigenous students, etc.), investigating potential biases in teacher-student interactions at other educational levels (e.g., pre-school education, higher education, etc.), exploring education inclusion in other school subjects (e.g., language, science, etc.) and by exploring the association between differential patterns of teacher-student interactions by sub-groups of students and progress in educational outcomes (i.e., teacher differential effectiveness), using longitudinal data.

5. FINAL CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

The aim of this thesis is to investigate the quality of teaching practices in schools with low academic performance. In order to contribute to the body of existing literature, this research seeks to explore in-depth the quality of teaching practices from different perspectives.

First, we conducted a study that investigates the variability in the quality of teaching practices in low-performing schools. This research was carried out through the creation of a quality index that considers multiple teaching practices observed in the classroom. This study confirmed that, although schools are relatively homogeneous in their academic results, the quality of the teaching practices within these schools varies greatly. This result can be considered counter-intuitive since, being low-performing schools, one might expect that it is not possible to find teachers with a good-quality index. However, this finding is in line with the literature that indicates that the context is an essential research aspect to be considered. The main result of this research indicates that only 4.8% of the quality variance of teaching practices is explained by the teacher's school context. This finding is relevant to public policy since it indicates that, in most of these low-performing schools, it is possible to find teachers with good teaching performance according to the standards of the Agency for Quality Education.

Second, we explored teaching quality from a person-centered approach. This approach is characterized by understanding teaching quality as a multidimensional concept in terms of the variables that define quality and heterogeneous in terms of the subjects of study (Masyn, 2013). The results of this study show four teaching categories, with the category of high-quality

practices grouping the highest proportion of teachers (38.47%). The category that groups the second highest proportion of teachers is low-quality practices, with 25.77%. The remaining two categories are content-oriented practices (18.88%) and climate-oriented practices (16.88%). This finding shows that there are different patterns in the quality of teaching practices, especially among those teachers who are in the middle range of teaching quality. Despite the little variance in teaching quality explained by the school context, the previous study identified a school factor related to the probability of teachers belonging to a specific category, which is related to school support for behavioral issues. This factor is important since it is a minimum operating condition for schools with a high vulnerability index (Kuhfeld, 2017).

The results of Chapters 2 and 3 are relevant for the development of public policies, as they provide useful information to guide future support strategies for these types of schools. One of these strategies could be a collaborative work system (Hargreaves & Fullan, 2012) that considers teachers with a high level of quality as tutors, which would mean taking advantage of the capacities that already exist in schools in difficult contexts.

Third, this thesis helps to address a recent challenge mentioned in the literature. This challenge is to evaluate, through observations and ego-networks, whether student learning opportunities are homogeneous within the classroom. This implies not considering teaching quality as an attribute that is distributed homogeneously among all students in the classroom, as is common in the literature and public policies. The results of this study show that there are differences in learning opportunities provided by the teacher to different groups of students in a classroom. Specifically, the group of Peruvian students was consistently less included in teacher-student interactions, which is aligned with the results of previous studies in Chile (Mondaca et al., 2018; Pávez,

2012). Finally, another result consistent with the study presented in Chapter 2 is that there is a significant variation in the level of student inclusion among teachers. Given that the migration phenomenon in Chile is expected to be maintained in the coming years due to political contexts elsewhere in Latin America and the Caribbean, the results of this study show there is an urgent need to raise awareness about migration among teachers and management teams, as well as to deliver concrete tools to teachers so that they can deal with diverse classrooms.

To sum up, we can conclude that it is important to study schools in their social context, as well as to study the heterogeneity of the quality of teaching practices at different levels. That is, to study in-depth whether teaching practices are distributed heterogeneously at the school level since the results of this thesis show that the variance is considerable. Also, it is important to question whether this heterogeneity can be systematized in certain profiles to achieve greater impact on policies designed to improve learning, and finally, to study the heterogeneity of teaching quality in terms of evaluating the inclusion of students in the process of learning in the classroom.

5.2 Future work

Based on the conclusions described above, there are lines for future work and aspects that could be studied in greater depth. An important improvement for the generalization of the results could be to perform the analysis described in Chapters 2, 3, and 4, but with a representative sample of Chilean schools. That is, increase the number of schools by diversifying academic performance and socioeconomic status. It would also be an improvement to stratify the sample in order to make inferences in minority groups.

Regarding the research proposed in Chapter 2, the psychometric aspects of the EPS document could be explored. Another option could be to use a previously validated observation rubric for the Chilean population with the same sample from the study, and verify associations of the results with both instruments. Also, in the future, analysis linking the observed teaching practices with sociodemographic variables of students, academic performance, or classroom composition could be included.

In the research described in Chapter 3, the next step would be to add covariates related to sociodemographic aspects, initial training, and working conditions of teachers. Students' sociodemographic characteristics could also be added, along with exploring a possible effect of classroom composition on teaching practices.

Finally, regarding the research presented in Chapter 4, one aspect to improve would be to consider secondary schools with three or more classrooms per level, and thus be able to compare the variability of the inclusion of students between classrooms of the same school, as well as between schools. Also, considering another age range of students would give information about inclusion patterns at different school levels. Another aspect that could be evaluated is whether there are differences in the level of inclusion of teachers who teach specific disciplines. The analysis could also be extended to other minority groups, such as students with special needs or indigenous students. Finally, similar measures could be taken to evaluate teachers in order to identify temporary patterns of inclusion.

6. REFERENCES

Agencia de Calidad de la Educación (2014) Guía para comprender la categoría de desempeño y orientar las rutas de mejora. Santiago, Chile

Agencia de Calidad de la Educación (2019). Resultados SIMCE 2018 [SIMCE 2018 Results]. [Data file and code book]. Retrieved from https://www.agenciaeducacion.cl/evaluaciones/que-es-el-simce/

Allen, R., & Sims, S. (2018). Do pupils from low-income families get low-quality teachers? Indirect evidence from English schools. *Oxford Review of Education*, 1-18.

Alvarado, M., Cabezas, G., Falck, D., & Ortega, M. E. (2012). La evaluación docente y sus instrumentos: Discriminación del desempeño docente y asociación con los resultados de los estudiantes. Artículo de investigación, Centro de Estudios Ministerio de Educación de Chile y Programa de Desarrollo de las Naciones Unidas [PNUD], Santiago.

Angell, A. (2002). The Politics of Education Reform in Chile: The Programa 900 Escuelas and the MECE-Básica. In *Social Policy Reform and Market Governance in Latin America* (pp. 147-164). Palgrave Macmillan, London.

Asparouhov, T., & Muthen, B. (2008). Multilevel mixture models. *Advances in latent variable mixture models*, 27-51.

Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools systems come out on top*. McKinsey & Company.

Bell, C. A., Dobbelaer, M. J., Klette, K., & Visscher, A. (2019). Qualities of classroom observation systems. *School effectiveness and school improvement*, *30*(1), 3-29.

Bellei, C., Muñoz, G., Pérez, L. M., & Raczynski, D. (2003). Escuelas efectivas en sectores de pobreza. *La educación en Chile, hoy*, 347-372.

Bellei Carvacho, C. I., Morawietz, L., Valenzuela, J. P., & Vanni, X. (2015). *Nadie dijo que era fácil. Escuelas efectivas en sectores de pobreza, diez años después*. LOM ediciones.

Bellei, C., Cabalin, C., & Orellana, V. (2018). The Student Movements to Transform the Chilean Market-Oriented Education System 1. In *Civil Society Organizations in Latin American Education* (pp. 63-84). Routledge.

Bellei, C., Morawietz, L., Valenzuela, J. P., & Vanni, X. (2019). Effective schools 10 years on: factors and processes enabling the sustainability of school effectiveness. *School Effectiveness and School Improvement*, 1-23.

Bill and Melinda Gates Foundation (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. MET Project Research Paper.

Blatchford, P., Bassett, P., & Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher–pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction*, *21*(6), 715–730.

Blazar, D., Litke, E., & Barmore, J. (2016). What does it mean to be ranked a "high" or "low" value-added teacher? Observing differences in instructional quality across districts. *American Educational Research Journal*, 53(2), 324-359.

Boda, Z., & Néray, B. (2015). Inter-ethnic friendship and negative ties in secondary school. *Social Networks*, *43*, 57-72.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*(1), 3-27.

Boutin-Martinez, A., Mireles-Rios, R., Nylund-Gibson, K., & Simon, O. (2019). Exploring Resilience in Latina/o Academic Outcomes: A Latent Class Approach. *Journal of Education for Students Placed at Risk (JESPAR)*, *24*(2), 174-191.

Bravo, R. (2011). Inmigrantes en la escuela chilena: Ciertas representaciones para ciertas políticas en educación [Immigrants in the Chilean school: Certain representations for certain education policies]. *Revista Latinoamericana de Educación Inclusiva*, *6*(1), 39–52.

Brophy, J., & Good, T. (1974). *Teacher-Student Relationships: Causes and Consequences*. New York: Holt, Rinehart, & Winston.

Bruns, B. De Gregorio, S., Taut, S. (2016). *Measures of effective teaching in developing countries*. Oxford: Estudio para el programa RISE

Bruns, B., & Luque, J. (2014). *Great teachers: How to raise student learning in Latin America and the Caribbean*. The World Bank.

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, *38*(1), 155–200.

Cabezas, V., Paredes, R., Bogolasky, F., Rivero, R., & Zarhi, M. (2017). First job and the unequal distribution of primary school teachers: Evidence for the case of Chile. *Teaching and Teacher Education*, 64, 66-78.

Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education*, *29*(3), 347–362.

Caro, D. H., Lenkeit, J., & Kyriakides, L. (2016). Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012. *Studies in Educational Evaluation*, *49*, 30–41.

Carolan, B. V. (2013). *Social Network Analysis and Education: Theory, Methods & Applications*. Thousand Oaks, CA: Sage Publications.

Casteel, C.A. (1998). Teacher-student interactions and race in integrated classrooms. *The Journal of Educational Research*, *92*(2), 115–120.

Castillo, D., Santa Cruz-Grau, E., & Vega, A. (2018). Estudiantes migrantes en escuelas públicas chilenas [Migrant students in Chilean public schools]. *Calidad en la Educación, 49*, 18–49.

Cerón, L., Pérez Alvarado, M., & Poblete, R. (2017). Percepciones docentes en torno a la presencia de niños y niñas migrantes en escuelas de Santiago: Retos y desafíos para la inclusión [Teacher perceptions about the presence of migrant children in schools in Santiago: Challenges for inclusion]. *Revista Latinoamericana de Educación Inclusiva*, *11*(2), 233-246.

Charalambous, C. Y., Kyriakides, E., Kyriakides, L., & Tsangaridou, N. (2019). Are teachers consistently effective across subject matters? Revisiting the issue of differential teacher effectiveness. *School Effectiveness and School Improvement*. https://doi.org/10.1080/09243453.2019.1618877

Chubb, J. E. & Moe, T. M. (1988) Politics, Markets and the Organization of Schools. *American Political Science Review*, 82, 1065-87.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, 6(4), 284

Civitillo, S., Juang, L.P., Badra, M., & Schachner, M. K. (2019). The interplay between culturally responsive teaching, cultural diversity beliefs, and self-reflection: A multiple case study. *Teaching and Teacher Education*, *77*, 341–351.

Contini, D. (2013). Immigrant background peer effects in Italian schools. *Social science research*, *42*(4), 1122–1142.

Cox, C. (2003). Las políticas educacionales de Chile en las últimas dos décadas del siglo XX. *Políticas educacionales en el cambio de siglo. La reforma del sistema escolar en Chile*, 19-113.

Cox, C. (2004). Innovation and reform to improve the quality of primary education: Chile. *Paper commissioned for the EFA Global Monitoring Report 2005, The Quality Imperative*.

Cox, C. (2012). Política y políticas educacionales en Chile 1990-2010. *Revista Uruguaya de Ciencia Política*, *21*(1), 13-43.

Creemers, B.P.M. (1994), *The Effective Classroom*, London: Cassell.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems, 1695*(5), 1-9.

Davis, T. M. (2014). School Choice and Segregation: "Tracking" Racial Equity in Magnet Schools. *Education and Urban Society*, *46*(4), 399-433.

De Boer, H., Bosker, R. J., & van der Werf, M. P. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology, 102*(1), 168–179.

De Gregorio, S. and Bruns, B. (2016). Práctica docente en el aula: Una mirada al interior de las salas de clase en América Latina y el Caribe. In: J. Manzi and M. García, ed., *Abriendo las puertas del aula: Transformación de las prácticas docentes*, 1st ed. Santiago, Chile, pp.59-92.

DEM [Departamento de Extranjería y Migración] (2016). *Migración en Chile 2005-2014 [Migration in Chile, 2005-2014]*. Retrieved from: https://www.extranjeria.gob.cl/media/2016/02/Anuario-Estad%C3%ADstico-Nacional-Migraci%C3%B3n-en-Chile-2005-2014.pdf

DEM [Departamento de Extranjería y Migración] (2017) Población Migrante en Chile [Migrant Population in Chile] Retrieved from: https://www.extranjeria.gob.cl/media/2019/04/Reporte-Migratorio-Poblaci%C3%B3n-Migrante-en-Chile.pdf

den Brok, P., & Levy, J. (2005). Teacher-student relationships in multicultural classes: reviewing the past, preparing the future. *International Journal of Educational Research, 43*(1-2), 72–88.

den Brok, P., Wubbels, T., Veldman, I., & Van Tartwijk, J. (2009). Perceived teacher-student interpersonal relationships in Dutch multi-ethnic classes. Educational Research and Evaluation, 15(2), 119-135.

Desimone, L. M., Smith, T. M., & Frisvold, D. E. (2010). Survey Measures of Classroom Instruction: Comparing Student and Teacher Reports. *Educational Policy*, *24*(2), 267–329.

DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness?. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 453-466.

Drew H. Gitomer (2019). Evaluating instructional quality. *School Effectiveness and School Improvement,* 30:1, 68-78, DOI: 10.1080/09243453.2018.1539016

DuBois, C., Butts, C. T., McFarland, D., & Smyth, P. (2013). Hierarchical models for relational event sequences. *Journal of Mathematical Psychology*, *57*(6), 297–309.

Elacqua, G., Martinez, M., Santos, H., & Urbina, D. (2016). Short-run effects of accountability pressures on teacher policies and practices in the voucher system in Santiago, Chile. *School Effectiveness and School Improvement*, *27*(3), 385-405.

Escribano, R., Treviño, E., Nussbaum, M., Torres Ibarra, D., Carrasco, D. (2020). How much does the quality of teaching vary at under-performing schools? Evidence from classroom observations in Chile. *International Journal of Educational Development.72.* https://doi.org/10.1016/j.ijedudev.2019.102125

European Union (2017). *Preparing Teachers for Diversity. The role of Initial Teacher Education. Final Report*. Retrieved from: https://doi.org/10.2766/637002

Everson, K. C., Feinauer, E., & Sudweeks, R. (2013). Rethinking teacher evaluation: A conversation about statistical inferences and value-added models. *Harvard Educational Review*, *83*(2), 349-370.

Fagginger Auer, M. F., Hickendorff, M., Van Putten, C. M., Béguin, A. A., & Heiser, W. J. (2016). Multilevel latent class analysis for large-scale educational assessment data: Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education*, *29*(2), 144-159.

Feng, L., & Sass, T. R. (2017). Teacher quality and teacher mobility. *Education Finance and Policy*, *12*(3), 396-418

Fernández, M.P. (2018). Mapa del Estudiantado Inmigrante Escolar Chileno. Documento de trabajo N° 12 [Map of Chilean Immigrant School Students. Working Paper N° 12]. Retrieved from: https://www.mineduc.cl/wp-content/uploads/sites/19/2018/05/MAPA_ESTUDIANTES_EXTRANJEROS_SISTEMA_ESCOLAR_CHILENO_2015_2017.pdf

Fraser, B. J., & Walberg, H. J. (2005). Research on teacher–student relationships and learning environments: Context, retrospect and prospect. *International Journal of Educational Research, 43*(1-2), 103-109.

Friedman, M. (1962). The role of government in education. In M. Friedman. (Ed.). *Capitalism and Freedom*. Chicago, Illinois: University of Chicago Press.

Fullan, M., & Rincon-Gallardo, S. (2016). Developing high-quality public education in Canada: The case of Ontario. In *Global Education Reform* (pp. 169-193). Routledge.

Gallego, F. A., & Hernando, A. (2010). School choice in Chile: Looking at the demand side. *Pontificia Universidad Catolica de Chile Documento de Trabajo*, (356).

García-Huidobro, J. E. (2007). Desigualdad educativa y segmentación del sistema escolar. Consideraciones a partir del caso chileno [Educational inequality and segmentation of the school system. Considerations from the Chilean case]. *Pensamiento Educativo, 40*(1), 65–85.

Gill, B., Shoji, M., Coen, T., and Place, K. (2016). The content, predictive power, and potential bias in five widely used teacher observation instruments (REL 2017–191). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory MidAtlantic. Retrieved from http://ies.ed.gov/ncee/edlabs.

Glock, S. (2016). Does ethnicity matter? The impact of stereotypical expectations on in-service teachers' judgments of students. *Social Psychology of Education, 19*(3), 493–509.

Goe, L., Bell, C., & Little, O. (2008). Approaches to Evaluating Teacher Effectiveness: A Research Synthesis. National Comprehensive Center for Teacher Quality.

Goldstein, H., & Woodhouse, G. (2000). School effectiveness research and educational policy. *Oxford Review of Education, 26*(3-4), 353-363.

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in generalized linear multilevel models. *Understanding Statistics, 1*(4), 223–232.

Haiyan, Q., Walker, A., & Xiaowei, Y. (2017). Building and leading a learning culture among teachers: A case study of a Shanghai primary school. *Educational Management Administration & Leadership*, 45(1), 101-122.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23–34.

Hammer, J. H. (2016, October). Percent of Uncontaminated Correlations (PUC) Calculator: A Microsoft Excel-based tool to calculate the PUC statistic. Available from http://DrJosephHammer.com/.

Hansen, K. Y., & Gustafsson, J. E. (2018). Identifying the key source of deteriorating educational equity in Sweden between 1998 and 2014. *International Journal of Educational Research*.

Hanushek, E.A. (1989). The impact of differential expenditures on school performance. Educational Researcher, 18(4), 45–51, 62.

Hanushek, E. A., & Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. *Annu. Rev. Econ*., 4(1), 131-157.

Hargreaves, A., & Fullan, M. (2012*). Professional Capital : Transforming Teaching in Every School*. New York: Teachers College Press.

Hattie, J. A. C. (2009). *A synthesis of over 800 meta-analyses relating to achievement*. London, England: Routledge.

Henry, K. L., & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling*, *17*(2), 193-215.

Hernández, M., & Raczynski, D. (2015). Elección de escuela en Chile: De las dinámicas de distinción y exclusión a la segregación socioeconómica del sistema escolar. *Estudios pedagógicos (Valdivia)*, *41*(2), 127-141.

Hsieh, C. T., & Urquiola, M. (2003). *When schools compete, how do they compete? An assessment of Chile's nationwide school*

Inclusion Law 2015 [Ley de inclusion Escolar 2015] (Chile). Retrieved from: https://www.leychile.cl/Navegar?idNorma=1078172

Instituto Nacional de Estadísticas (INE) (2018) Síntesis de Resultados Censo 2017 [Census Results summary 2017]. Retrieved from: https://www.censo2017.cl/descargas/home/sintesis-de-resultados-censo2017.pdf

Irvine, J. (1985). Teacher communication patterns as related to the race and sex of the student. *The Journal of Educational Research*, *78*(6), 338–345.

Irvine, J. (1986). Teacher–student interactions: Effects of student race, sex, and grade level. *Journal of Educational Psychology*, *78*(1), 14–21.

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1(4), 85-108.

Jensen, B., Grajeda, S., & Haertel, E. (2018). Measuring cultural dimensions of classroom interactions. *Educational Assessment*, *23*(4), 250–276.

Jerome, E. M., Hamre, B. K., & Pianta, R. C. (2008). Teacher-child relationships from kindergarten to sixth grade: Early childhood predictors of teacher-perceived conflict and closeness. *Social Development, 18*(4), 915–945.

Joiko, Sara, & Vásquez, Alba. (2016). Acceso y elección escolar de familias migrantes en Chile: "No tuve problemas porque la escuela es abierta, porque acepta muchas nacionalidades". [School access and choice of migrant families in Chile: 'I had no problems because the school is open, because it accepts many nationalities']. *Calidad en la Educación, 45,* 132–173.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. In *Research Paper. MET Project. Bill & Melinda Gates Foundation*.

Kane, T. J., Staiger, D. O., McCaffrey, D., Cantrell, S., Archer, J., Buhayar, S., Kerr, K., Kawakita, T., & Parker, D. (2012*). Gathering feedback for teaching: combining high-quality observations with student surveys and achievement gains*. Seattle: Bill & Melinda Gates foundation.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. Journal of human Resources, 46(3), 587-613.

Kim, D. H., Lambert, R. G., & Burts, D. C. (2018). Are young dual language learners homogeneous? Identifying subgroups using latent class analysis. *The Journal of Educational Research*, *111*(1), 43-57.

Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 12, 59-76.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.

Kuhfeld, M. (2017). When Students Grade Their Teachers: A Validity Analysis of the Tripod Student Survey. *Educational Assessment*, *22*(4), 253-274.

Kyriakides, L., Creemers, B. P. M., & Charalambous, E. (2018, July 3). Searching for differential teacher and school effectiveness in terms of student socioeconomic status and gender: implications for promoting equity. *School Effectiveness and School Improvement*. https://doi.org/10.1080/09243453.2018.1511603

Ladd, G. W., & Burgess, K. B. (1999). Charting the relationship trajectories of aggressive, withdrawn, and aggressive/withdrawn children during early grade school. *Child Development, 70*(4), 910–929.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence?. *Child development, 70*(6), 1373-1400.

Ladd, H. F. (2012). Education and poverty: Confronting the evidence. *Journal of Policy Analysis and Management*, 31(2), 203-227.

Langeloo, A., Mascareño Lara, M., Deunk, M. I., Klitzing, N. F., & Strijbos, J.-W. (2019). A systematic review of teacher-child interactions with multilingual young children. *Review of Educational Research, 89*(4), 536-568.

Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly (1982-)*, 377-389.

Levin, B. (2006) Schools in challenging circumstances: A reflection on what we know and what we need to know, *School Effectiveness and School Improvement*, 17:4, 399-407, DOI: 10.1080/09243450600743459

Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher–child interactions in Chile and their associations with prekindergarten outcomes. Child Development, 86(3), 781-799.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936-949.

LoCasale-Crouch, J., Vitiello, G., Hasbrouck, S., Aguayo, Y. C., Schodt, S. C., Hamre, B., & Romo, F. (2016). Cómo medir lo que importa en las aulas de primera infancia: un enfoque sobre las interacciones educadora-niño. *Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana* 53(1), 1-14.

Loeb, S. & Reininger, M. (2004). Public Policy and Teacher Labor Markets: What We Know and Why It Matters. East Lansing, MI: The Education Policy Center at Michigan State University.

Loreman, T. (2014). Measuring inclusive education outcomes in Alberta, Canada. *International Journal of Inclusive Education, 18*(5), 459–483.

Lupton, R. (2005). Social justice and school improvement: improving the quality of schooling in the poorest neighbourhoods. *British Educational Research Journal*, 31(5), 589-604.

Mahatmya, D., Lohman, B. J., Brown, E. L., & Conway-Turner, J. (2016). The role of race and teachers' cultural awareness in predicting low-income, Black and Hispanic students' perceptions of educational attainment. *Social Psychology of Education, 19*(2), 427–449.

Mansfield, R. K. (2015). Teacher quality and student inequality. *Journal of Labor*

Economics, 33(3), 751-788.

Manzi, J., García, M. R., & Godoy, M. I. (2017). *Informe Técnico SEPA, Sistema de Evaluación de Progreso del Aprendizaje* [SEPA Technical Report, Learning Progress Evaluation System]. Retrieved from: http://www.sepauc.cl/wp-content/uploads/2017/05/Informe-T%C3%A9cnico-SEPA-29.05.2017.pdf

Martínez, F., Taut, S., & Schaafa, K. (2016). Classroom observation for evaluating and improving teaching: An International perspective. *Studies in Educational Evaluation, 49*, 15–29.

Martinic, S., & Vergara, C. (2007). Gestión del tiempo e interacción del profesor-alumno en la sala de clases de establecimientos con jornada escolar completa en Chile. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *5*(5e), 3-20.

Martinic, S; Vergara, C; Villalta, M & Huepe, D. (2013). Uso del tiempo e interacciones en la sala de clases. Un estudio de casos en Chile. *Pro-Posicoes*, 24(1), 123-135.

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and classroom characteristics associated with teachers' ratings of prekindergartners' relationships and behaviors. *Journal of Psychoeducational Assessment*, *24*(4), 367–380.

Masyn, K. E. (2013). 25 latent class analysis and finite mixture modeling. In *The Oxford handbook of quantitative methods* (p. 551). Oxford University Press, Oxford.

McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods.* 1996;1(1):30–46.

Meckes, L., & Bascopé, M. (2012). Uneven Distribution of Novice Teachers in the Chilean Primary School System. *Education policy analysis archives*, 20(30), n30.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Journal of Education*, 79(4), 33-53.

MINEDUC (2013a). Equidad en los aprendizajes escolares en Chile en la última década. Serie de Evidencias, Año 2, N17. Santiago, Chile: Centro de Estudios Mineduc, Ministerio de Educación.

MINEDUC (2013b). Estándares Indicativos de Desempeño para los Establecimientos Educacionales y sus Sostenedores. Santiago, Chile: Unidad de Currículum y Evaluación, Ministerio de Educación.

MINEDUC (2013c). Fundamentos Estándares Indicativos de Desempeño para los Establecimientos Educacionales y sus Sostenedores. Santiago, Chile: Unidad de

Currículum y Evaluación, Ministerio de Educación.

MINEDUC (2018) *Política Nacional de Estudiantes Extranjeros 2018-2022.* [National Policy for foreign-born students 2018-2022]. Retrieved from: https://migrantes.mineduc.cl/wp-content/uploads/sites/88/2018/06/POLITICA-NACIONAL-EE-Final-1-1.pdf

Mondaca, C., Muñoz, W., Gajardo, Y., & Gairín, J. (2018). Estrategias y prácticas de inclusión de estudiantes migrantes en las escuelas de Arica y Parinacota, frontera norte de Chile [Strategies and practices for the inclusion of migrant students in the schools of Arica and Parinacota, at the northern border of Chile]. *Estudios atacameños, 57,* 181–201.

Moody, J. (2001). Race, school integration, and friendship segregation in America. *American Journal of Sociology, 107*(3), 679–716.

Mora Olate, M. L. (2018). Política educativa para migrantes en Chile: un silencio elocuente [Educational policy for migrants in Chile: an eloquent silence]. *Polis (Santiago), 17*(49), 231–257.

Moore Johnson, S. (2012). Having it both ways: Building the capacity of individual teachers and their schools. *Harvard Educational Review*, 82, 107-122.

Muijs, D., & Reynolds, D. (2011). *Effective teaching*. Evidence and practice. London: Sage.

Muijs, D., Harris, A., Chapman, C., Stoll, L., & Russ, J. (2004). Improving schools in socioeconomically disadvantaged areas–A review of research evidence. *School effectiveness and school improvement*, 15(2), 149-175.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art: Teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256.

Munoz, M. A., & Dossett, D. (2014). Equity and excellence: The effect of school and sociodemographic variables on student achievement. *Jsl* Vol 11-N2, 11, 120.

Nasiopoulou, P., Williams, P., Sheridan, S., & Yang Hansen, K. (2019). Exploring preschool teachers' professional profiles in Swedish preschool: a latent class analysis. *Early Child Development and Care, 189*(8), 1306-1324.

National Educational Quality Agency (2013). *Metodología de Construcción de Grupos Socioeconómicos. Pruebas SIMCE 2013* [Socioeconomic Group Construction Methodology. SIMCE test 2013]. Santiago, Chile: Performance Achievement Division, National Educational Quality Agency.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). The relation of global first-grade classroom environment to

structural classroom features and teacher and student behaviors. *Elementary School Journal*, 102, 367–387.

National Institute of Child Health and Human Development Early Child Care Research Network. (2004). Classroom Observation System—Fifth Grade. Charlottesville, VA: University of Virginia

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects?. *Educational evaluation and policy analysis*, *26*(3), 237-257.

OECD [The Organization for Economic Cooperation and Development] (2015a). *Can the performance gap between immigrant and non-immigrant students be closed?* (PISA in Focus No. 53). Retrieved from: https://www.oecd-ilibrary.org/education/can-the-performance-gap-between-immigrant-and-non-immigrant-students-be-closed_5jrxqs8mv327-en

OECD [The Organization for Economic Cooperation and Development] (2015b). *Do teacher-student relations affect students´ well-being at school?* (PISA In Focus No. 50). Retrieved from: https://www.oecd-ilibrary.org/education/do-teacher-student-relations-affect-students-well-being-at-school_5js391zxjjf1-en

OECD [The Organization for Economic Cooperation and Development] (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Retrieved from: https://www.oecd.org/education/pisa-2015-results-volume-i-9789264266490-en.htm

OECD [The Organization for Economic Cooperation and Development] (2019). *TALIS 2018 Results (Volume I): Teachers and school leaders as lifelong learners*. Retrieved from: http://www.oecd.org/education/talis-2018-results-volume-i-1d0bc92a-en.htm

OECD [The Organization for Economic Cooperation and Development] . (2013). Do students perform better in schools with orderly classrooms? *PISA in Focus*, (32). https://doi.org/10.1787/5k40d63gcd44-en

Ortega, L., Malmberg, L. E., & Sammons, P. (2017). Teacher effects on Chilean children's achievement growth: A cross-classified multiple membership accelerated growth curve model. *Educational Evaluation and Policy Analysis*, 0162373718781960.

Ortega, L., Treviño, E. & Gelber, D. (in press). La inclusión de las niñas en las aulas de matemáticas chilenas: Sesgo de género en las redes de interacciones profesor-estudiante. *Infancia y Aprendizaje*

Palmer, D. J., Sadiq, H. M., Lynch, P., Parker, D., Viruru, R., Knight, S., ... & Stillisano, J. (2016). A classroom observational study of Qatar's independent schools: Instruction and school reform. *The Journal of Educational Research*, 109(4), 413-423.

Pavez Soto, I. & Chan, C. (2018). The second generation in Chile: Negotiating identities, rights, and public policy. *International Migration*, *56*(2), 82–96.

Pavez Soto, I. (2012). Inmigración y racismo: experiencias de la niñez peruana en Santiago de Chile. *Si Somos Americanos*, *12*(1), 75–99.

Peterson, E. R., Rubie-Davies, C., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*, *42*, 123-140.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational researcher*, 38(2), 109-119.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008*). Classroom Assessment Scoring System (CLASS).* Baltimore, MD: Brookes.

Pianta, R., Hamre, B., & Mintz, S. (2011). *CLASS: Classroom Assessment Scoring System CLASS-Upper Elementary Manual*. Charlottesville, VA: Teachstone.

Praetorius, A. K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: looking back and looking forward. *ZDM*, *50*(3), 535-553.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata. Volume II: Categorical Responses, Counts, and Survival* (Third Edition). Texas: Stata Press.

Ramberg, J., Låftman, S. B., Almquist, Y. B., & Modin, B. (2019). School effectiveness and students' perceptions of teacher caring: A multilevel study. *Improving Schools*, *22*(1), 55–71. https://doi.org/10.1177/1365480218764693

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013a). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of personality assessment*, *95*(2), 129-140.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92, 544–559.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013b). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5-26.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., &

Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement*, 25(2), 197-230.

Riedemann, A., & Stefoni, C. (2015). Sobre el racismo, su negación, y las consecuencias para una educación anti-racista en la enseñanza secundaria chilena [On racism, its denial, and the consequences for an anti-racist education in Chilean secondary education]. *Polis, 42*, 1-21.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*, 137-150.

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, *52*(3), 475-514.

Salas, N., Castillo, D., San Martín, C., Kong, F., Thayer, L. E., & Huepe, D. (2017). Inmigración en la escuela: caracterización del prejuicio hacia escolares migrantes en Chile [Immigration in school: characterization of prejudice towards migrant school children in Chile]. *Universitas Psychologica*. *16*(5), 1-20.

Sanders, W.L. & Rivers, J.C. (1996). Cumulative and residual effects of teachers on future student academic achievement. Knoxville: University of Tennessee Value-Added Research and Assessment Center

Santelices, M. V., Galleguillos, P., González, J., & Taut, S. (2015). A Study About Teacher Quality in Chile: The Role of the Context Where Teachers Work and Value-Added Measures. *Psykhe*, *24*(1).

Scheerens, J. (2000). Improving school effectiveness. UNESCO International Institute for Educational Planning.

Scheerens, J. (2015). *Educational Effectiveness and Ineffectiveness: A Critical Review of the Knowledge Base*. Springer

Scheerens, J., & Blömeke, S. (2016). Integrating teacher education effectiveness research into educational effectiveness models. *Educational research review*, *18*, 70-87.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of educational research*, *99*(6), 323-338.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade:

The role of theory and research design in disentangling meta-analysis results. *Review of educational research*, 77(4), 454-499.

Snijders, T. A. B., & Bosker, R. J. (2012). Discrete dependent variables. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*, 304-307.

Stadtfeld, C., Hollway, J., & Block, P. (2017). Dynamic Network Actor Models: Investigating Coordination Ties through Time. *Sociological Methodology, 47*(1), 1–40.

Stallings, J. (1977). *Learning to look: A handbook on classroom observation and teaching models*. Wadsworth Publishing Company.

Stallings, J. A., & Mohlman, G. G. (1988). Classroom observation techniques. *Educational research, methodology, and measurement: An international handbook*, 469-474.

Stefoni, C., Acosta, E., Gaymer, M., & Casas-Cordero, F. (2008). Niños y niñas inmigrantes en Santiago de Chile. Entre la integración y la exclusión [Immigrant children in Santiago de Chile. Between integration and exclusion]. Santiago: Universidad Alberto Hurtado–OIM [International Organization for Migration].

Stefoni, C., Stang, F., & Riedemann, A. (2016). Educación e interculturalidad en Chile: un marco para el análisis [Education and interculturality in Chile: a framework for analysis]. *Estudios Internacionales*, *185*, 153–182.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S. & Serrano, A. (1999). *The TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States.* Washington, D.C.: U.S. Dept. of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement, 21*(3), 289–314.

Tagliacollo, V.A., Volpato, G.L. & Pereira, A. (2010). Association of student position in classroom and school performance. *Educational Research, 1*(6), 198–201.

Taut, S., Valencia, E., Palacios, D., Santelices, M. V., Jiménez, D., & Manzi, J. (2016). Teacher performance and student learning: linking evidence from two national assessment programmes. *Assessment in Education: Principles, Policy & Practice*, 23(1), 53-74.

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The international system for Teacher Observation and Feedback: Evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation, 12*(6), 561–582.

Teddlie, C., & Stringfield, S. (2007). A history of school effectiveness and improvement research in the USA focusing on the past quarter century. In International handbook of school effectiveness and improvement (pp. 131-166). Springer Netherlands.

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*(2), 253–273.

Teodorović, Jelena (2011) Classroom and school factors related to student achievement: what works for students*? School Effectiveness and School Improvement*, 22:2, 215-236

Thrupp, M., & Lupton, R. (2006). Taking school contexts more seriously: The social justice challenge. *British Journal of Educational Studies*, 54(3), 308-328.

Tijoux, M. E. (2013a). Las escuelas de la inmigración en la ciudad de Santiago: Elementos para una educación contra el racismo [The schools of immigration in the city of Santiago: Elements for an education against racism]. *Polis, 12*(35), 287–307.

Tijoux, M. E. (2013b). Niños(as) marcados por la inmigración peruana: estigma, sufrimientos, resistencias [Children marked by Peruvian immigration: stigma, suffering, resistance]. *Convergencia. Revista de Ciencias Sociales, 20*(61), 83–104.

Torres, R. (2018). Tackling inequality? Teacher effects and the socioeconomic gap in educational achievement. Evidence from Chile. *School Effectiveness and School Improvement*, 1-35.

Treviño, E., Valdés, H., Castro, M., Costilla, R., Pardo, C., & Donoso Rivas, F. (2010). *Factores asociados al logro cognitivo de los estudiantes de América Latina y el Caribe*. SERCE. OREALC/UNESCO, Santiago, Chile.

Treviño, E., Fraser, P., Meyer, A., Morawietz, L., Inostroza, P., & Naranjo, E. (2015b). Factores asociados al aprendizaje. Tercer Estudio Regional Comparativo y Explicativo. Santiago, Chile: Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación, UNESCO-Santiago.

Treviño, E., Varela, C., Romo, F., & Núñez, V. (2015). Presencia de lenguaje académico en las educadoras de párvulos y su relación con el desarrollo del lenguaje de los niños. Calidad en la educación, (43), 137-168.

Treviño, E., Leonor, V., Felipe, G. and María, M. (2016). ¿Qué caracteriza a las interacciones pedagógicas de las escuelas efectivas chilenas?: Una aproximación exploratoria. In: J. Manzi and M. García, ed., Abriendo las puertas del aula: Transformación de las prácticas docentes, 1st ed. Santiago, Chile, pp.185-220.

Treviño, E., Villalobos, C., & Baeza, A. (2016b). Recomendaciones de políticas educativas

en América Latina en base al TERCE

Treviño, E., (2018). Diagnostico del sistema escolar: Las reformas educativas 2014-2017. In: I. Sanchez ed., Ideas en Educación II: Definiciones en tiempos de cambio, 1st ed. Santiago, Chile, pp. 133-166.

Trevino, E., Mintrop, R., Villalobos, C., & Ordenes, M. (2018). What Might Happen if School Vouchers and Privatization of Schools Were to Become" Universal" in the US: Learning from a National Test Case--Chile. *National Education Policy Center*.

Treviño, E., Carrasco, A., Villalobos, C. and María, M. (2019). Financiamiento de la educación escolar en Chile: La reforma estructural pendiente. In: A. Carrasco and Luis M. Flores ed., De la reforma a la transformación: Capacidades, innovaciones y regulación de la educación chilena, 1st ed. Santiago, Chile.

UNESCO [United Nations Educational Scientific and Cultural Organization] (2017). *A Guide for Ensuring Inclusion and Equity in Education*. Retrieved from: https://unesdoc.unesco.org/ark:/48223/pf0000248254

UNESCO [United Nations Educational Scientific and Cultural Organization] (2018). *Global Education Monitoring Report 2019. Migration, displacement and education: Building bridges, not walls*. Retrieved from: https://en.unesco.org/gem-report/report/2019/migration

United Nations General Assembly (2015). *Transforming our world: the 2030 Agenda for Sustainable Development.* A/RES/701, United Nations. Retrieved from: https://www.unfpa.org/sites/default/files/resource-pdf/Resolution_A_RES_70_1_EN.pdf

Valencia, E., & Taut, S. (2011). Evidencia de inequidad en el acceso a los docentes mejor calificados en la Enseñanza Básica Chilena [Evidence of inequity in access to better qualified teachers in Chilean basic education] (MIDE UC Technical Report IT-1108). Santiago: Pontificia Universidad Católica de Chile, Escuela de Psicología, Centro de Medición MIDE UC.

Valenzuela, J. P., Bellei, C. & De los Ríos, D. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy, 29*(2), 217–241.

Van Geel, M., Keuning, T., Frèrejean, J., Dolmans, D., van Merriënboer, J., & Visscher, A. J. (2019). Capturing the complexity of differentiated instruction. *School effectiveness and school improvement*, *30*(1), 51-67.

Vanlaar, G., Kyriakides, L., Panayiotou, A., Vandecandelaere, M., McMahon, L., De Fraine, B., & Van Damme, J. (2016). Do the teacher and school factors of the dynamic model affect high-and low-achieving student groups to the same extent? A cross-country study. *Research Papers in Education*, *31*(2), 183-211.

Wenglinsky, H. (2004). The link between teacher classroom practices and student academic performance. Education policy analysis archives, 10, 12.

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, *17*(1), 33-51.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political analysis*, *18*(4), 450-469.

Vermunt, J. K., & Magidson, J. (2005). Hierarchical mixture models for nested data structures. In *Classification—The ubiquitous challenge* (pp. 240-247). Springer, Berlin, Heidelberg.

Vermunt, J. K., & Magidson, J. (2013). Technical guide for Latent GOLD 5.0: Basic, advanced, and syntax. *Belmont, MA: Statistical Innovations Inc*.

Vitoroulis, I., & Georgiades, K. (2017). Bullying among immigrant and non-immigrant early adolescents: School-and student-level effects. *Journal of adolescence*, *61*, 141–151.

Wubbels, T., Brekelmans, M, & Mainhard, T. (2016). Teacher–student relationships and student achievement. In K. R. Wentzel & G. B. Ramani (Eds.). *Handbook of Social Influences in School Contexts Social-Emotional, Motivation, and Cognitive Outcomes* (pp. 127–142). New York: Routledge.

Youn, M. (2016). Inequality from the first day of school: The role of teachers' academic intensity and sense of responsibility in moderating the learning growth gap. *The Journal of Educational Research*, 109(1), 50-67.

Zhang, Z., & Yuan, K. H. (2016). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: Methods and software. *Educational and Psychological Measurement*, *76*(3), 387-411.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *psychometrika*, *70*(1), 123-133.

7. APPENDICES

7.1. Appendix A

Table A2-1: Values of intraclass correlation for each indicator

Indicator	ICC	Confidence Interval
Objective	0,74	0,663 - 0,796
Clarity	0,88	0,836 - 0,909
Content	0,89	0,839 - 0,92
Dynamism	0,62	0,508 - 0,708
Introduction	0,74	0,646 - 0,814
Presentation	0,74	0,658 - 0,799
Production	0,69	0,586 - 0,764
Practice	0,69	0,596 - 0,771
Review	0,76	0,65 - 0,834
Interest	0,81	0,737 - 0,862
Monitoring	0,82	0,762 - 0,866
Feedback	0,88	0,843 - 0,915
Praise	0,93	0,897 - 0,951
Timing	0,84	0,979 - 0,879
Dedication	0,90	0,864 - 0,922
Independence	0,88	0,838 - 0,916
Support	0,88	0,838 - 0,916
Expectations	0,85	0,81 - 0,888
Courtesy	0,85	0,81 - 0,888
Routines	0,83	0,773 - 0,87
Opinion	0,75	0,681 - 0,804

7.2. Appendix B

Table A2-2: Polychoric Correlation of indicators

Variables	Objective	Clarity	Content	Dynamism	Introduction	Presentation	Production	Practice	Review	Interest	Monitoring	Feedback	Praise	Timing	Dedication	Independence	Support	Expectations	Courtesy	Routines	Opinion
Objective																					
Clarity	.328																				
Content	.414	.759																			
Dynamism	.404	.729	.536																		
Introduction	.351	.486	.411	.451																	
Presentation	.541	.691	.602	.809	.535																
Production	.467	.657	.525	.798	.486	.850															
Practice	.480	.659	.556	.802	.467	.845	.915														
Review	.290	.354	.367	.466	.493	.478	.398	.381													
Interest	.241	.580	.541	.669	.453	.573	.594	.574	.400												
Monitoring	.289	.596	.506	.717	.442	.622	.631	.662	.390	.758											
Feedback	.432	.673	.661	.766	.532	.735	.734	.781	.482	.763	.867										
Praise	.297	.609	.532	.666	.440	.614	.647	.615	.454	.801	.746	.792									
Timing	.316	.623	.484	.806	.424	.644	.609	.650	.499	.622	.613	.662	.565								
Dedication	.321	.562	.420	.843	.381	.634	.622	.652	.417	.598	.620	.681	.571	.885							
Independence	.314	.559	.427	.747	.355	.597	.688	.727	.368	.549	.618	.655	.501	.688	.769						
Support	.252	.389	.359	.518	.326	.525	.532	.578	.342	.630	.600	.666	.623	.466	.471	.512					
Expectations	.251	.534	.395	.672	.338	.575	.546	.589	.397	.675	.564	.658	.640	.712	.723	.552	.531				
Courtesy	.301	.548	.403	.712	.405	.577	.568	.594	.437	.667	.595	.660	.654	.771	.748	.593	.517	.905			
Routines	.315	.507	.411	.599	.373	.517	.478	.501	.355	.462	.475	.528	.424	.648	.614	.503	.388	.655	.695		
Opinion	.397	.556	.455	.665	.441	.682	.644	.616	.480	.593	.558	.694	.602	.569	.567	.562	.491	.629	.592	.503	

7.3. Appendix C

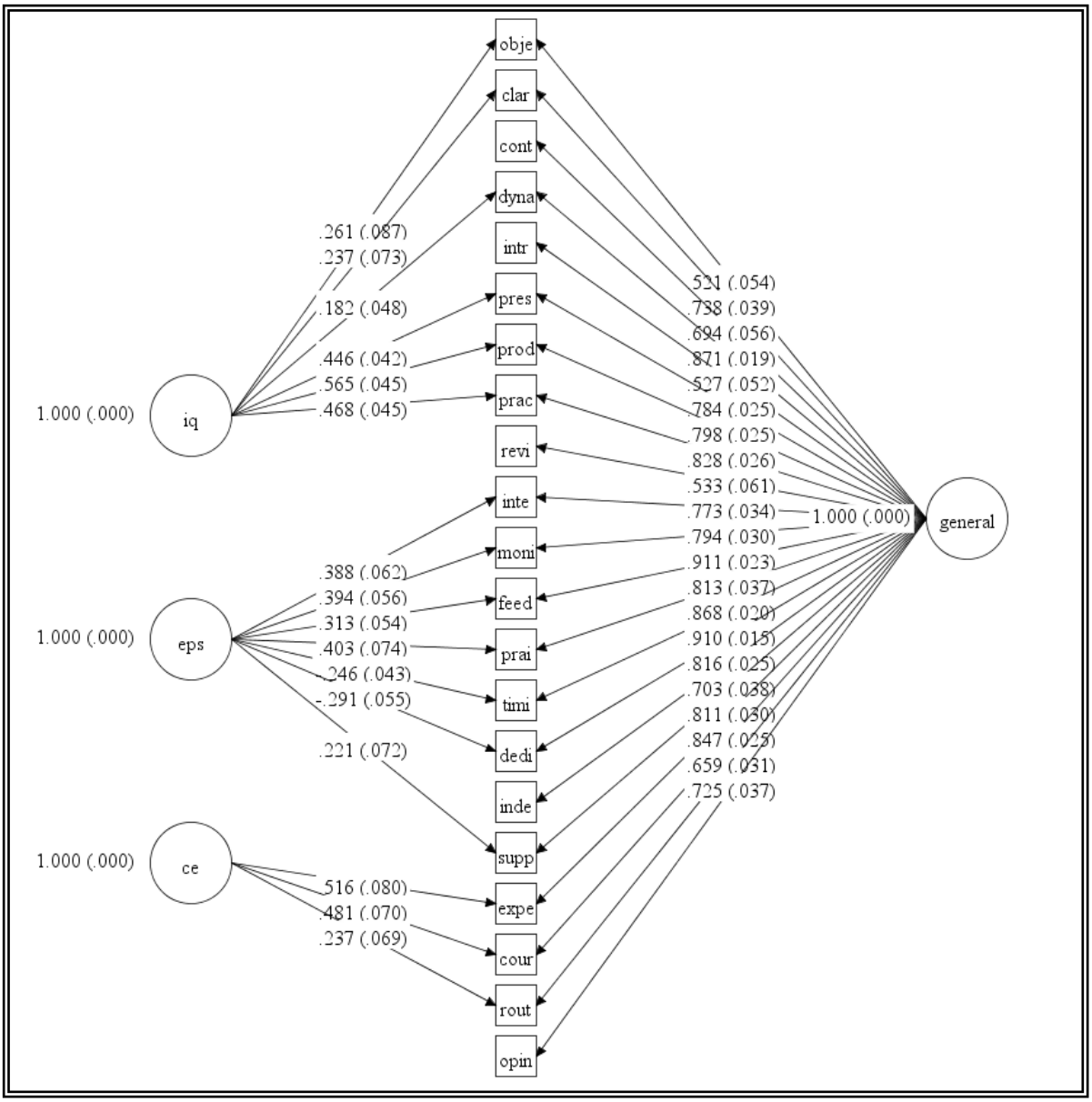

Figure A2-1: Factor loads for the first cycle sample

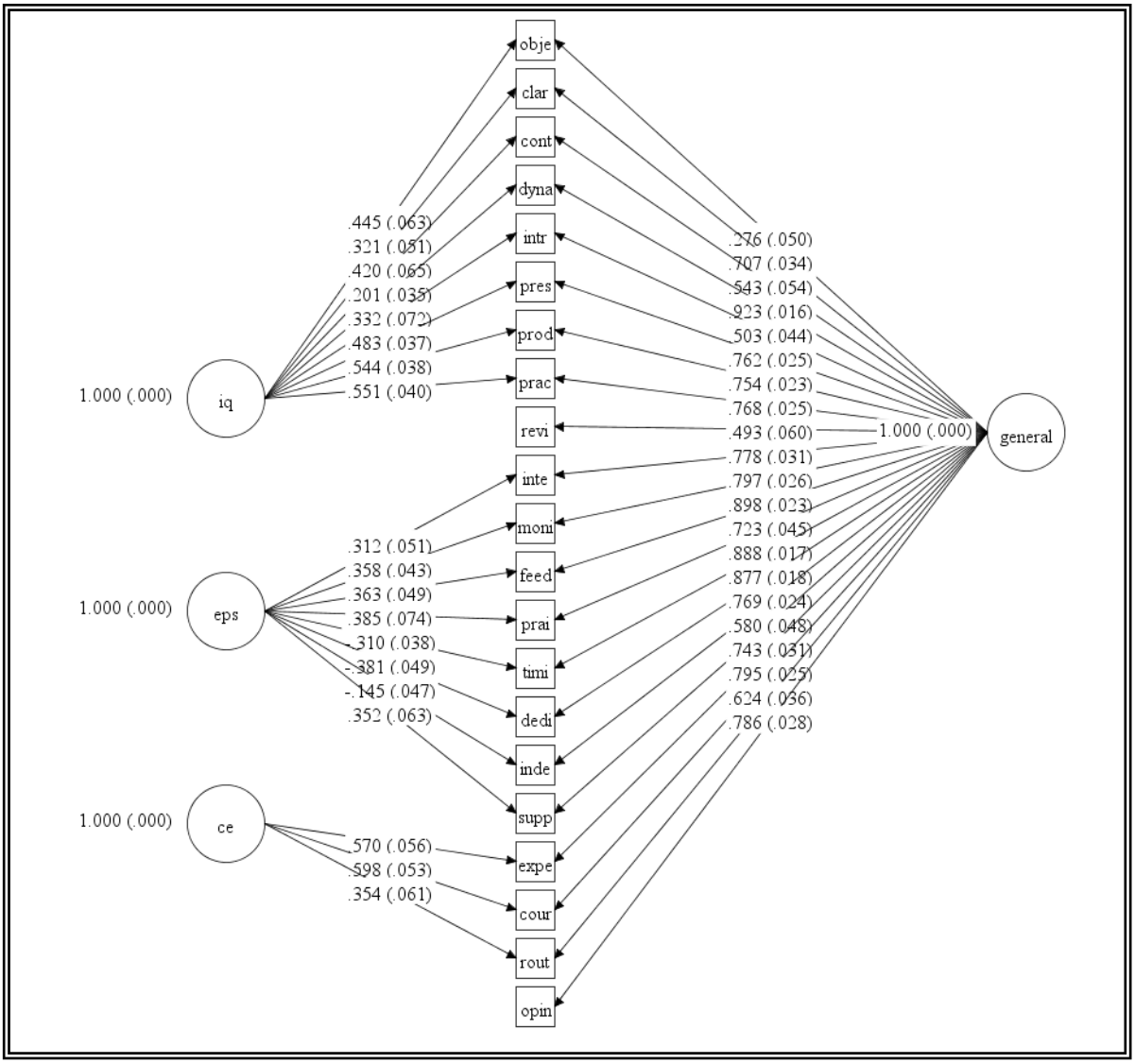

Figure A2-1: Factor loads for the second cycle sample

7.4 Appendix D

Table A2-3 Origin of indicators based on the Ministry of Education's Quality Standards

Indicators in the original document	**Abbreviated name**	**[min-max]**
E5.1 C1	Objective	[1-3]
E5.2 C1	Clarity	[1-3]
E5.2 C2	Content	[1-3]
E5.2 C3	Dynamism	[1-4]
E5.3 C1	Introduction	[1-3]
E5.3 C2	Presentation	[1-3]
E5.3 C3	Production	[1-4]
E5.3 C4	Practice	[1-3]
E5.3 C5	Review	[1-3]
E5.4 C1	Interest	[1-3]
E5.4 C2	Monitoring	[1-3]
E5.4 C3	Feedback	[1-3]
E5.4 C4	Praise	[1-3]
E5.5 C1	Timing	[1-3]
E5.6 C1	Dedication	[1-3]
E5.6 C3	Independence	[1-3]
E6.1 C2	Support	[1-3]
E7.3 C2	Expectations	[1-2]
E8.1 C1	Courtesy	[1-2]
E8.4 C1	Routines	[1-3]
E9.3 C1	Opinion	[1-3]

7.5 Appendix E

Table A2-4: Example of Standard 5.4

Poor Development	Initial Development	Adequate Development	Advanced Development
Does not meet the criteria for adequate development as one or more of the following issues is present:	*Partially meets the criteria for adequate development, though one or more of the following issues is present:*		*Meets all of the criteria for adequate development, while one or more of the following aspects is also present:*
The teachers appear to be indifferent, irritated or aggressive towards their students: they do not know their names, they ignore their concerns, dismiss their contributions, frequently lose their patience, are apathetic towards their difficulties, they threaten or intimidate them, or they are ironic, among others; or, they show favoritism towards certain students and are indifferent towards the others.	The teachers show some interest in and concern for their students, though at times they seem unreceptive or inattentive.	The teachers show interest in and concern for their students: they call them by their name, they listen to their concerns, they know about their interests, they value their contribution in class, they are patient and they encourage and support them when they struggle, among others.	The teachers are outstanding in the concern they show for their students: they help them after class, they listen to them and counsel them when they have personal problems, and they call them at home and ask for them when they're absent, among others.
In general, the teachers do not monitor their students' comprehension or performance during class: they limit themselves to lecturing and setting their students work without checking their comprehension or performance.	The teachers are not very systematic when it comes to monitoring their students' comprehension or performance: they sometimes check and sometimes don't, or they only focus on the students who are struggling the most.	The teachers constantly monitor their students' comprehension and performance during class: they walk around the classroom to observe the work that is being done, they ask students to explain what they have understood, and they listen to their doubts and concerns, among others.	The teachers help their students reflect on their own learning process: they guide them so that they learn to identify whether they are clear on what needs to be achieved (and clarify things if necessary), to define which strategy will be best (considering the effort that is required), and to evaluate and adapt this strategy should it not be effective.
In general, the teachers do not give their students any feedback on their performance or they do so only superficially or leniently, not allowing the students to identify their achievements or any areas that need improvement; or, they only provide feedback to a small number of outstanding students.	The teachers are not very systematic when it comes to giving feedback on their students' performance, or they only focus on the students who are struggling the most.	The teachers constantly give their students feedback on their performance, both individually and as a group: they highlight their achievements, help them identify and be more aware of their mistakes, guide them to correct their errors and explains things again if needed.	The teachers correct assignments, check exercise books and student work, leaving written comments highlighting the students' achievements while also identifying their mistakes and making suggestions on how to improve their work.
The teachers do not highlight or acknowledge their students' progress, effort or perseverance: they constantly criticize them, only acknowledge some and systematically criticize the others, or they compare the students with one another without valuing their individual progress.	The teachers congratulate their students, but this is not directly related to their effort or achievements, or they tend only to highlight their achievements, without acknowledging students who have made a considerable effort but failed to achieve good	The teachers constantly congratulate their students on their progress, effort or perseverance: they value their progress and associate it with the effort that has been made, they help them identify how to overcome a difficulty and encourage them to keep going in order to improve their	The teachers congratulate their students individually on their effort and achievements, leaving personalized written notes or positive comments on their test papers, exercise books or assignments, as well as in out-of-class conversations or notes to parents, among others.

	results.	performance, among others.	

Notes: This standard is then broken down into the following indicators: Interest, Monitoring, Feedback and Praise. Each of these indicators corresponds to a paragraph taken from the standard. The standard was translated into English by the author (p.83, MINEDUC 2013b).

7.6 Appendix F

Table A4-1. Frequency and percentage of student immigrant generation by country of origin.

	Peru	Bolivia	Colombia	Haiti	Ecuador	Venezuela	Dominican Republic	Other Countries	Total
First generation	139	8	47	20	10	63	17	9	340
	40.9%	2.4%	13.8%	5.9%	2.9%	18.5%	5.0%	2.6%	100.0%
Second generation	57	3	4	1	3	1	0	11	80
	71.3%	3.8%	5.0%	1.3%	3.8%	1.3%	0.0%	13.8%	100.0%
Total	196	11	51	21	13	64	17	20	933
	21.0%	1.2%	5.5%	2.3%	1.4%	6.9%	1.8%	2.1%	100.0%

7.7 Appendix G

Table A4-2. Intra-class correlation (ICC) by initiator and type of interaction ($n_{classrooms}$ = 9. $n_{students}$ = 218).

	Teacher-initiated interactions				
	Behaviour Management	*Instructions*	*Administrative*	*Pedagogical*	*Total*
ICC	0.73 (0.66 - 0.78)	0.55 (0.45 - 0.64)	0.71 (0.64 - 0.77)	0.67 (0.58 - 0.73)	0.82 (0.77 - 0.86)

	Student-initiated interactions				
	Behaviour Management	*Instructions*	*Administrative*	*Pedagogical*	*Total*
ICC	-	0.59 (0.50 - 0.67)	0.53 (0.43 - 0.62)	0.79 (0.73-0.83)	0.74 (0.67-0.79)

Notes: Lower and upper limits of 95% confidence interval are shown in parentheses. The ICC for student-initiated interactions of type Behaviour Management was not be estimated due to its low frequency of occurrence.

7.8 Appendix H

Table A4-3. Items and factor loadings of the scale *Teacher's Attitude towards Diversity in the Classroom*.

Scale/Item	**Factor loading**
Teacher's Attitude towards Diversity in the Classroom (Cronbach's α = .88)	
How much do you agree with the following statement? "A class with a high proportion of immigrant students, indigenous students and students with Special Educational Needs (SEN)…	
1. hinders the progression of the class".	.89
2. makes the work of the teacher harder"	.78
3. harms the average academic level".	.89
4. generates discipline problems".	.89

7.9 Appendix I

Table A4-4. Correlations between classroom-level variables ($n_{classrooms}$ = 38).

	Proportion of Immigrant Students	Proportion of Immigrant Students^2	Teacher Attitude towards Diversity in the Classroom	Class Size
Proportion of Immigrant Students	1			
Proportion of Immigrant Students^2	.390*	1		
Teacher Attitude towards Diversity in the Classroom	-.269	-.232	1	
Class Size	.396*	.381*	-.420**	1

Notes: Pearson correlation coefficients. † $p < .10$: * $p < .05$; ** $p < .01$; *** $p < .001$