PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

FACULTAD DE EDUCACIÓN

ASSESING THE INSTRUCTIONAL LEADERSHIP'S AND ACADEMIC HETEROGENEITY'S INFLUENCE OVER TEACHERS' EFFICACY BELIEFS

POR

GERMÁN RODOLFO FROMM RIHM

Tesis presentada a la Facultad de Educación de la Pontificia Universidad Católica de Chile

para optar al grado académico de Doctor en Ciencias de la Educación

PROFESOR GUÍA: DR. PAULO LUIS VOLANTE BEACH

EVALUADORES:

DR. ALEJANDRO CARRASCO ROZAS

DR. JUAN CARLOS CASTILLO VALENZUELA

DR. PHILIP HALLINGER

Diciembre de 2016

Santiago de Chile

A mis padres, maestros de la vida y

los mejores profesores que he tenido

Alida, por profesión

Günter, por vocación

AGRADECIMIENTOS

Estos agradecimientos van dirigidos a todos quienes hicieron posible por contribución directa o indirecta con su generosidad que esta tesis se haya llevado a cabo satisfactoriamente.

En primer y mayor lugar, quiero agradecer a los establecimientos educacionales que me recibieron en sus aulas y regalaron su tiempo para la recolección de datos. A sus directores que decidieron confiar en este trabajo y el valor que pueda llegar a tener.

A Paulo por la guía inteligente y focalizada desde principio a fin, con admiración.

A Alejandro y Juan Carlos por su orientación dedicada y fantásticas clases, que me hacen sentir el privilegio de haber sido su estudiante.

To Dr. Philip Hallinger for his generosity and trust to give me a place among the newcomers in a field that he has shaped so significantly.

Agradezco el apoyo de CONICYT en el financiamiento de aranceles y manutención entregada.

A mis compañeros del Programa y del Laboratorio cuya presencia me ha mostrado lo que es sentir eficacia colectiva. Especialmente a Fabián Derby y Natascha Roth, por las críticas amistosas y la cercanía intelectual.

A Myriam Navarrete por su preocupación, apoyo y socorro.

A Christian Dittmer por la generosidad con su base de operaciones, transporte y alimentos durante cada una de las visitas en el terreno.

A Juan de Dios Dougnac por editar con humor, ojalá que "por lo menos haya sido entretenido".

Finalmente, a Laura Fromm simplemente porque: "A veces las familias se demoran en terminar las tesis".

Abstract

A better understanding of how principals can influence teacher's efficacy beliefs and educational efforts is useful for practice and research. School leaders are recognized of being able to install optimistic 'operative cultures' in their schools that spur teaching among teachers. At the same time, teachers adjust their efforts to the expectations they have about the academic composition of their students. Both processes influence the teachers' efficacy beliefs at collective and individual levels. The purpose of this work is to offer a model that can explain how teachers change their self-efficacy beliefs. A viable causal model between instructional leadership and academic heterogeneity and how both affect the teachers' efficacy beliefs is tested. The cross-sectional design controlled structural social characteristics by sampling in Chile's Araucanía characterized as the poorest region, with highest indigenous students' proportion and rural schools. A sample of 366 teachers answered a questionnaire and secondary available data of the students were combined. A SEM confirmatory analysis showed a mediated effect of the leadership through collective efficacy (0.702) over teacher self-efficacy (0.617) but no other significant influences. This study helps to understand how principals can enforce socio-cognitive consensual scripts in the faculty, that prescribes shared efforts and expectations.

Index

Figures

Tables

I. INTRODUCTION

All over the world, educational systems make efforts to understand what their leaders should do in order to improve their schools and achieve better learning results with their students. Considering that different leaders in different schools achieve such improvement and results in varying manners, it becomes paramount to understand how these differences are taking place. This issue has been addressed by identifying 'school effectiveness factors' and relations between them that are found consistently in research.

For example, school leadership as one of this factor can have significant positive effects on students learning and in more specific areas like teacher and community satisfaction, change initiation, organizational improvement, and teachers' efficacy beliefs. In school contexts, efficacy beliefs are a teacher's projected ideas about their future performance. These beliefs can be about their own performance -called self-efficacy- or about the whole faculty's performance -called collective efficacy- (Goddard & Goddard, 2001). A considerable amount of studies has addressed the influence leaders can exert over efficacy beliefs both outside school contexts (e.g. Jung & Sosik, 2002; Waumba, Wang, Lawler, and Shi, 2004) and inside them (Dussault, Payette, & Leroux, 2008; Calik, Sezgin, Kavgaci, & Kilinc, 2012; Dumay & Galand, 2012; Volante, 2012).

At the same time, student composition differences have been considered to influence the teachers' perception about their job, affecting their efficacy beliefs. For example, Ronald Goddard and Yvonne Goddard (2001) found a high correlation (r=.73; p<0.01) between collective efficacy and the academic composition operationalized through previous learning results differences within groups. Moreover, an HLM school composition analysis was found to be the most powerful academic compositional predictor of perceived collective efficacy differences among schools (Goddard & Skrla, 2006), measuring eight school-

context and five teacher-demographic variables (46% explained variance). The authors concluded, at least for this model, that operative shared beliefs could not be predicted solely through social and academic composition but also through other factors, that they considered likely to be leadership as an all in all shared discursive construction called 'operative culture' (Goddard & Skrla, 2006).

In this thesis, a combined analysis of two empirically grounded relations is going to be attempted in order to understand efficacy beliefs' influencing factors.

This is relevant not only because these relationships have not been studied together before, but also because teacher performance is closely determined by efficacy beliefs, as previous studies found out (Goddard, Sweetland, & Hoy, 2000; Goddard, Hoy, & Hoy, 2004). This rationale can be explained through cognitive theory, stating that beliefs shape people's behavior. This predisposition comes from psychologically multidimensional (perceptual emotional, cognitive, etc.) scripts or schemes that are shared in groups (Augoustinos, Walker, & Donaghue, 2014). The extent to which teacher performance and expectations (another belief about students' success) influence student achievement is widely understood (Hallinger & Heck, 1996; Leithwood & Jantzi, 1999; Witziers, Bosker, & Krüger, 2003).

This goes further from what Bandura (1998) established as the four ways to foster efficacy beliefs, namely previous mastery, vicarious experience, social persuasion and emotional states. If leaders and a contextual variable can be compared by the way they affect teachers' efficacy beliefs, a concrete step towards understanding how much leaders influence their schools will have been made. In deciding where to put their efforts educational leaders have little practical guidance, and efficacy beliefs' influence seems to be a missing link in a chain that can define operative school cultures of optimism, effort and positive emotional influence.

Finally, drawing on conceptual frameworks, an influence "path" explains how different teacher self-efficacy levels can be associated to schools' academic

heterogeneity, leadership, and collectively shared efficacy beliefs. By selecting these variables (a task that requires previous relevant research knowledge), a sound model is delivered. It can also be related to a broader discussion about contextualized sensible leadership exercise and segregation issues that are indirectly having part on student learning via teachers' attitudes.

1.1.   Background and framework

The general paradigm of school effectiveness research (SER) has been in force for about four decades of educational research. The main assumption is the effectiveness of schools depend on series of school intern factors, and an agenda to explore, describe and explain these factors is still active. The importance that school factors gain over contextual factors is disseminated in the public imaginary (Reynolds, Teddlie, Hopkins, & Stringfield,2000) determining policy and practice. This conception of school differences is closely tied to explanations about school effects that are attempted to be measured in a broad tradition of quantitative research (Leithwood, Pattem, & Jantzi, 2010; Robinson, Lloyd, Rowe, 2008; Hallinger & Heck, 1996). The critics of the paradigm can be organized in three main aspects (Carrasco, 2008). The first, points out that these kind of research usually ignore local context variations resulting in 'one size fits all' kind of knowledge and policy designs (MacBeath, Gray, Cullen, Frost, Steward, & Swaffield, 2007). The second, is an instrumental approach to education and pedagogy that lacks of sound conceptual frameworks as in "black box" and "input-output" models. The third is a general misrecognition of the role that social or contextual factors like social disadvantage, poverty, marginalization, and discrimination play, in explaining school effectiveness boundaries and moderations

(Hopkins, 2001). The risk of labeling schools as 'ineffective', because of assumptions about the schools' capacities, is inherent in this paradigm.

The paradigm apologists, on the other hand, focus on a pragmatic perspective to solve school problems (while recognizing or not contextual boundaries), promote good practices based on empirical evidence and avoid the pessimistic perspective that schools are not able to counter deficit conditions.

The present study can be considered as part of the paradigm's agenda, exploring how school intern dynamics take place. It considers the external factor of academic heterogeneity on which schools exert no control on an integrated fashion inside the dynamics. The importance to produce specific knowledge about disadvantaged settings is also addressed, while trying to relate to pragmatic and optimistic perspectives on how schools can counter 'deficit thinking' (Goddard & Skrla, 2006).

In the next segments three research fields are reviewed in order to cover the conceptual and empirical knowledge that is needed to frame the present study.

1.1.1. Efficacy Beliefs

The research history of efficacy beliefs is framed in socio-cognitive theories the focus on exercised agency through peoples' dispositions. The concept was originally coined in the '70 by Albert Bandura to explain behavioral change, as the result of optimistic attitudes towards new behaviors. A common definition states that the efficacy is the perception of the capability to accomplish a future given task. As such it can be understood as combination of intentions, forethoughts, self-regulation, self-reflectiveness, perception of owns capabilities, quality of functioning, and motivation. It is also practical theory for agency and expectations (Bandura, 2001). The concept took root as a solid and consistent explanation, that can be measured through people's own perceptions and useful for many performance areas (Bandura, 2006). The soundness of measuring the beliefs about something to happen, already integrating people's perception about the task itself and the effort they can and are willing to do, offers researchers a predictive way of addressing complex phenomena. For instance, in education alone, efficacy beliefs about parenting, learning, teaching and principalship can be easily distinguished. The main critics that can be discussed about efficacy belief constructs are specific to the field in which they are applied and can be understood as content validity problems (Bandura, 2006). Further, efficacy beliefs should not be confounded with actual performance or obtained effects, even when they are good predictors.

In regard to school effectiveness, teacher's efficacy beliefs are recognized as a relevant factor to predict, their efforts for teaching and achieving better results with the students (Goddard, Sweetland, & Hoy, 2000; Goddard, Hoy, & Hoy, 2004).

As socio-cognitive scripts teacher efficacy beliefs help to explain how they interact with teacher performance in dynamic relations:

"[...] *teachers who believe strongly in their instructional efficacy create mastery experiences for their students. Those beset by self-doubts construct classroom environments that are likely to undermine students' sense of efficacy and cognitive development. As Woolfolk and Hoy (1990) reported, teachers' sense of personal efficacy affects their general orientation toward the educational process as well as their specific instructional practices. Those who have a low sense of instructional efficacy favor a custodial orientation that relies heavily on extrinsic inducements and negative sanctions to get students to study. Teachers who believe strongly in their instructional efficacy support development of students' intrinsic interests and academic self-directedness. Ashton and Webb (1986) documented the cumulative impact of teachers' instructional self-efficacy on students' academic achievement.*" (Bandura, 1993, p. 140)

This corresponds in great extend to the expectations that teachers have about their students, the standard they accept for their own and their students' performance, and what pressure they exercise to achieve it. This constitutes a 'construct' that explains how specific teachers and faculties understand their teaching and can vary among schools if cultural dimensions are considered.

Socio-cognitive constructs like academic optimism (Hoy, Hoy, & Kurz, 2008), collective efficacy (R. D. Goddard, 2001; R. D. Goddard, Hoy, & Hoy, 2000) or individual teacher self-efficacy (Bandura, 1993; R. D. Goddard & Goddard, 2001) have been used to distinguish 'operative cultures' in schools that are very different even in similar contexts. They all have in common that, when people believe themselves capable of reaching given goals, they are more likely to approach them persistently and creatively until they succeed. However, when they lack a sense of efficacy, reflecting unlikely achievement beliefs, those efforts and persistence are considered more futile.

Efficacy beliefs are mainly divided into individual teachers' efficacy beliefs -when teachers report their own expectations about future performance on a specific task- and collective efficacy -if teachers respond such expectations on an aggregated level-. While, teacher's self-efficacy is an individual's belief collective efficacy is a perception of faculty's competence to deal with the specific educational challenge of the school. Self-efficacy constructs and scales therefore measure a supposed teaching performance. Goddard and LoGerfo (2007) recommended for collective efficacy scales that ask for group perceptions, such as "Teachers in this school believe" instead of averaged individual teacher collective efficacy scores, because they represent a collective level rather than the mean of individual efficacy perception. The constructs in fact predict organized activity, agents' actions and their perceived control of the future (Goddard & Skrla, 2006). Efficacy beliefs have been consistently associated with actual performance and student achievement (Goddard, Sweetland, & Hoy; 2000; Goddard, Hoy, & Hoy, 2004; Volante, 2012).

A causal relation between these two levels was addressed by Goddard and Goddard (2001), who associated a one standard deviation increase in collective efficacy with a quarter standard deviation increase (0.248; p<.01) in teacher efficacy, after accounting for the contextual variables. Their rationale explained how social pressure and shared previous successful experiences, as conceptualized by Bandura before (1998), determine individual beliefs of future achievement. Moreover, collective efficacy explained all of the variation among schools in teacher efficacy in the sampled schools (Goddard & Goddard, 2001). On the other hand, Calik, et.al. (2012) and Demir (2008) observed positive and significant effects of self-efficacy over collective efficacy with β=.13 (p<.05) and β=.49 (p<0.001) respectively, but incoherent explanations of reciprocal causalities were offered, leaving unanswered the question as to why one-way causalities were used. Considering both attempts, only the consistently explained collective over individual influence will be used in this study.

A relation between educational leadership and schools that shows the desirable enthusiasm and persistence to teach can be established. Recent empirical evidence from different analyses show correlations of r=.52 (p<.01) (Leithwood, Patten, & Jantzi, 2010) in Canada and r=.54 (p<.01) in Chile (Volante, 2010). Also significant effects of β=.34 (p<.01) in Calik, Sezgin, Kavgaci, and Kilinc (2012) in Turkey were observed. Only recently in Oman and Iran structural equation modeling (SEM) was used and obtained a similar result of β=.60 and β =0.58 respectively (Al-Mahdy, Emam, & Hallinger, 2018; Hallinger, Hosseingholizadeh, Hashemi, & Kouhsari, 2017). All these coefficients are similar in size but not absolutely comparable as they are obtained with different data analysis (correlation, regression, and SEM). It is still discussed how and to what extent a principal can exercise influence through mediating variables (like efficacy) over teachers' performance (Calik et al., 2012; Dumay & Galand, 2012; Goddard, 2001; Goddard et al., 2000; Leithwood & Jantzi, 2008; and Leithwood et al., 2010). For educational policies this is paramount, because of the impact in all the system through its multiplying effect of having better leaders in charge of many teachers and of even more students in turn. The influence of school leaders over teachers' efficacy beliefs at a collective and individual level, and again over teachers' performance, to finally reach student learning is a solid work hypothesis, increasing the relevance of this knowledge.

Conveniently, some research of the relation between compositional factors and collective efficacy has been performed (Goddard & Goddard, 2001; Goddard & Skrla, 2006; Knoblauch & Hoy, 2008). In their model, Ronald and Yvonne Goddard (2001) included contextual information about the amount low-SES students and racial minority proportion and related them (among others) to collective efficacy. Significantly (p<0.01) high (although negative) correlations of r=-0.726 and r=-0.479 were found respectively. These size effects are considered together with prior mathematics achievement that showed a high (but positive) correlation to (r=0731,

p<0.01) making those effects comparable. This leads to the conclusion that collective efficacy is a strong predictor of differences among schools.

Notoriously, the results of the efficacy beliefs research seem to contradict the academic heterogeneity advances, as effects and results should be favorable for heterogeneity. But the results do not contradict but problematize a broader question that has not been thoroughly answered, regarding the ample spectrum of schools and social compositions that determine school differences at a high scale. These differences can be considered structural by nature and usually relate to the families' SES, ethnicity, living in rural areas, school size, curricular tracks, and so on. What efficacy beliefs try to capture are operative cultures that may vary on a much smaller scale. In this sense, local structural differences like the ones elicited by Goddard and Skrla (2006) might be hinges between the high and low scale differences, showing moderate results. As stated before regarding teacher challenges in front of heterogeneous groups, their cognition favors high efforts. However, under structural disadvantages their sense of efficacy will likely decrease. This means that structural contexts seem to moderate the effects of heterogeneity over teacher cognition. The rationale behind this supposition is that a given experience on people's efficacy beliefs is less a function of the actual event than of its 'cognitive processing' (Goddard & Goddard, 2001; Raudenbush, Rowan, & Cheong, 1992; Bandura, 1993). From this perspective, many factors shape efficacy beliefs at the same time and competent leaders exercise their influence under heavy contextual boundaries. To be fair, those structural differences will determine their actions and possibilities, but when compared in similar settings their effectiveness can be held accountable for efficacy beliefs in greater extent.

1.1.2. Instructional Leadership

School leadership has been researched for over 35 years to strengthen the educational agents in the pursue of effectiveness, improvement and achievements. The findings have determined that educational leadership is a school effectiveness factor (Leithwood, Seashore Louis, Anderson, & Wahlstrom, 2004; Witziers, Bosker, & Krüger, 2003). It is also related to many desirable effects as students learning achievements (Hallinger & Heck, 1996; Robinson, Lloyd, & Rowe, 2008) and teachers' and community sense about the school (Robinson & Timperley, 2007; Wahlstrom & Louis, 2008). For educational policies this is paramount, because of the impact overall the system through its multiplying effect of having leaders in charge of many teachers and of even more students each. So a field that includes researchers, policy-makers, school managers, schools, universities, governments and mostly principals is constituted around the phenomena.

Conceptually, it has developed to take distance from an administrative focus proposing interpersonal or transformational (Hallinger & Leithwood, 1994; K. A. Leithwood, 1992), instructional (Hallinger & Murphy, 1985; Hallinger, 2011; Robinson, Lloyd, & Rowe, 2008; Robinson & Timperley, 2007), and distributed (Spillane, Halverson, & Diamond, 2004; Timperley, 2005) among other models and perspectives redefining effective leaders. These models overlap, discuss complementary elements, and vary in some the priorities describing real principals' practices.

Probably, the most divergent conceptual controversies concerning educational leadership follow the same critics that are applicable to SER and are pointed out as decontextualized conclusions (Thrupp, Lauder, & Robinson, 2002) lack of attention to complex dynamics (Carrasco, & Fromm, 2016) and invariant perspectives even in front of social disadvantage (MacBeath, et. al., 2007).

Instructional leadership focuses on the principal's capability to improve the teaching and learning process. Predominantly, the model urges school leaders to be "instructional leaders" and to take responsibility for improving teachers' knowledge and skills (Leithwood & Mascall, 2008). It is broadly supported that instructional leadership practices require concrete definitions to provide sound measures that can produce and/or observe relevant effects (Robinson, Lloyd, & Rowe, 2008).

The term instructional leadership was coined in the early '80 in the English speaking research community (Hallinger, 2013). In Spanish the wording "*liderazgo instruccional*" has an educationally speaking wrong meaning as in 'giving instructions' being interpreted as anti-constructivist among educators. Therefore, the first translations of instructional leadership were "*liderazgo pedagógico*". The problem is that 'pedagogical leadership' is a critical proposition of the early XX century of John Dewey about the social meaning of education (Smyth, 1989). So crossed translations mix explicit school effectiveness practices with a reflection about values of the highest purposes in education. In the Spanish field the broadly used Pedagogical Leadership "*liderazgo pedagógico*" conceptualizes often imprecisely both meaning in the same term. In a Spanish SER the wording "*liderazgo instruccional*" should be used unambiguously as indicated in previous Chilean research (Volante, 2010, Fromm, Hallinger, Volante, & Wang, 2016).

In Chile instructional leadership has been studied and measured in secondary education in relation to organizational development and students' performance improvement (Volante, 2012; Fromm, Volante, & Olbrich, 2015). These findings support a positive association between instructional leadership of directive teams and teachers' efficacy beliefs. Moreover, the effects of instructional leadership on efficacy belief are related positive and significantly to academic results of students in university admission tests (Volante, 2010).

1.1.3. Academic heterogeneity

The concept of academic heterogeneity has been examined as such for 25 years now (Cohen, 1991). It refers to cultural, economic, and linguistic differences within classrooms that have a measurable impact on student performance. Research hypotheses frequently assumed that student outcomes are higher when a "good" peer group is influencing the individuals. Therefore, placing students in homogeneous groups (schools or classes) allows teachers to, in theory, adjust better to the pedagogical challenges to support and meet the needs of individual students. However, placing students into heterogeneous groups (schools and classes) would allow teachers to capitalize on diversity and benefit from peers (like tutors, but also participative dialogic interactions) as resources to support learning. This question has driven many researchers to look for a definitive answer to settle the debate (Wilkinson & Fung, 2002; Hattie, 2002; Thrupp, Lauder & Robinson, 2002; McEwan, 2003; and Treviño, Valenzuela, & Villalobos, 2014). However, a broad spectrum of methods considers often in the same approach social and academic dimensions of the students, making it hard to pinpoint precise rationales of what heterogeneity stands for.

A state of the art after a series of meta-analyses that originated a controversial debate (Slavin, 1987; Kulik & Kulik, 1987; Kulik, 1992; Lou et al., 1996; Rui, 2009), seems to agree that grouping has negative effects on students' performance (Dupriez, 2010; Venkatakrishnan & Wiliam, 2003). In a normal context, ability grouping produces negative consequences among lower-achieving students. High-achieving students seem not to benefit from homogeneity, nor to be affected negatively by heterogeneity either (Dupriez, 2010; Resh & Dar, 2012).

Consistently, conceptualizations of academic heterogeneity have been within-effects, of variance nested in schools or classes (depending on the studied segregation level) and usually neglected alternative constructions (e.g. Reynolds &

Teddlie, 2000) based on individual aggregation (e.g. Scharenberg, 2012). The common use of regression coefficients in regression models as heterogeneity definitions are always determined (and biased) by the dependent variables selection. Alternatives like considering 'raw heterogeneity indicators' like the standard deviation of a group or 'relative standard deviation' (the average divided by the standard deviation) also called 'coefficient of variation' have been explored in few cases (e.g. Scharenberg, 2012; Valenzuela, Bellei, & Allende, 2016). For this reason, this work will rely on the schools' students' standard deviations in the national performance assessment.

Specifically, in Chile, Academic Heterogeneity studies have relied regression analysis to answer for structural differences as SES, geographical location, educational tracking, student's selection done by the schools and segregation, finding high homogeneity inside schools (Valenzuela). As far as known, other social composition dimensions, like indigenous students' percentage in a classroom or the parent's average income in a class show low magnitude effects in language increases (McEwan, 2003). For instance, a 1σ of indigenous percentage's percentage increase (equal to 9%) lowers 0.03σ, the language increase, all else equal (McEwan, 2003).

At a general level, consistent with international literature, students with higher diversity inside schools show higher learning results, even more among low performers in low SES contexts (Vandenberghe, 2002). As far, it can be assumed that higher diversity diminishes the low SES situation effects on students (Valenzuela, Bellei, & Allende, 2016).

However, all studies assume `black box' explanations of how the effects take place inside schools. Sociocultural perspectives of the matter (e.g. Wilkinson & Fung, 2002) are essential to understand academic heterogeneity. Cohen (1991) and Persell (1977) explained how teachers shape their expectations and are pedagogically challenged when confronted with a heterogeneous classroom, spurring greater

efforts. On the contrary, tracking by ability grouping often increases the gap between students and was found unjustifiable as a school policy (Gamoran, Nystrand, Berends, & Lepore, 1995; Burris & Garrity, 2008; Rui, 2009; Nomi & Raudenbush, 2016). These explanations are still unrelated to the regression analysis that give statements about the effects, without paying attention to more complex explanations that have been offered.

The way of how this happens remains an unanswered question and to explore mediated effects of inside school factors becomes relevant. As schools under the previous antecedents' perspective can be considered as 'ineffective', because of factor the they do not exercise any control, the intern school moderator might be crucial to know (Carrasco, 2008). As 'operative cultures' are contextually sensible the efficacy beliefs constitute a sound choice to attempt this exploration. As stated before, schools can be considered differently by the SES or vulnerable students proportion among other factor, but are afterwards simply assumed to be populated be similar kind of students. This assumption does not help to differentiate the imbricated complexity the generates specific practices or sets boundaries for leaders and educators Carrasco, 2008).

1.2. Specification of the Chilean Setting

This study took place in Chile, which means that some particularities must be discussed to understand why the results can be generalized. Chile has been discussed thoroughly as particular case of market driven educational policies and increased standardization. Since the first students' social movement in 2006 the political discussion has modified substantially the local educational system. This year can be considered the end of one cycle of reforms developed since the '90 focused in the finances, provision and access that the education required, and the beginning of another cycle of reforms aimed to change its quality by design (Martinic & Elacqua, 2010). To achieve this, a standards based reform articulates the whole system setting task and achievements for school principals. This tenses the principals to answer to school external requirements for planning, administration and management and at the same time achieve pedagogical and learning standards in their schools. Under the Ministry's centralized control, principals perceive the capacities diminished in order to address their contextualized, communities' and local problems. In the public system principals cannot choose freely over firing or hiring teachers, have now control over the schools' budget or development plan and need to put great effort in lobbying with the Municipality's stakeholder. The stakeholders however must make market-driven decisions about the school management, pressing leaders towards non-academic foci (Carrasco & Fromm, 2016).

Nevertheless, in Chile, similarly to other countries like the USA, the UK or New Zealand, principalship is striving towards an instructional role in schools (MINEDUC, 2016). For example, a "Good Leadership and Principalship Framework" was established in 2005 and revised thoroughly in 2015 (MINEDUC, 2015). Principals are trained under policy design programs (Arancibia, 2011), selected in the public system through an open process (García, Peña, & Parra, 2016)

and kept accountable through a "Quality Management System" since 2006 (Donoso, Benavides, Cancino, Castro, & López, 2012). All these policy designs presume that educational leaders exercise influence over their organizations (MINEDUC, 2016). Also, distinctions between schools have been made, to tell learning-enforcing cultures apart from each other, measuring teacher efficacy beliefs through a yearly country-wide assessment called SIMCE since 1988 (Bravo, 2011). Other attempts of measuring teacher self-efficacy (Covarrubias & Mendoza, 2013; 2015) and collective efficacy (Volante, 2010; López & Gallegos, 2014) have been undertaken. In the same manner, the main issue of school leadership in specific contexts is present. However, the diversity inside the school system means that the common leadership criteria must be treated only referentially and not prescriptively (Donoso et al., 2012). For example, the variations of managing schools ranging in size from a dozen to thousands of students, the poverty levels, the private or public stakeholders, indigenous attendance proportion, or rural placement's consequences for educational leaders and the organizational self-perception must be considered.

The student heterogeneity under segregation lens is a critical research focus in the Chilean school system. Students are segregated in schools according to their socioeconomic status conforming factual tracks. This reduces heterogeneity in socioeconomic terms, with structural determination of the academic diversity. Also, schools with students from different SES tend to segregate more through differentiated class grouping (Treviño, Valenzuela, & Villalobos, 2014). In a competing market with voucher funding through parents' school choice, schools struggle to attract and retain students with higher abilities (Carrasco & Fromm, 2016). As, the effects of structural differences at the large scale over efficacy beliefs have not been tested internationally either, a local replication will be useful to attempt a confirmation of the literature so far.

Despite the policy development in principalship and educational leadership, Chile is also facing a structural educational reform, called Law of Inclusion (Ley N°

20845, 2015). This law stops student selection done by the schools. Before the law students' admission was a choice of schools as well as parents. After the law the choice remains solely on the parents implying increasingly more diverse student populations inside schools. Knowledge about all types of diversity will be more relevant for school administration in the future.

Under the actual Chilean reform processes, it becomes relevant to understand the complexity of the principal's influence over school's culture. Specifically, in regions or locations were historically the standard based reform has failed to achieve (or at least perceive) learning improvements this is even more relevant. The Araucanía is generally the poorest, with most indigenous students, with the highest rural proportion and the lowest achievement results. These contextual factors stress a standardization logic most of all by the high diversity that this implies. This work will also contribute to set priorities in policy design that modify or rely on school intern practices and can spur positive processes.

To understand better the relation between instructional leadership and the assumed professional development of teachers, through their efficacy beliefs is of worth for practice. Under the current principals' regulations, to count with studies that can inform how to do things in schools is useful.

Finally, to transfer the kind of studies that in SER have a long lasting tradition to the Chilean context becomes also a scholar issue. To check for empirical validation of advanced knowledge in the fields in this context makes sense in the global and local discussion. To some extend Chile is in need of confirmatory studies that help to answer the question if English-speaking knowledge production stand valid here. At the same time a contribution with Chilean data helps to close gaps, as the interrelated analysis of independently treated fields, and can provide findings in regard to cross-cultural research.

1.3.   Research Problem

Based on a review of the literature, this is the first study to examine all variables in the same analysis. Consistently, students for each part of this study have provided underlying notions of how school leadership and academic heterogeneity work as predictors for collective efficacy and self-efficacy. But, as the relation between these variables has not been consistently established, this study overcomes the relevant knowledge's fragmentation. As an integrated and simultaneous research of the three aspects, this study aims to perform a confirmatory analysis of the empirical evidence that could support a theoretical understanding of the relational model.

If this study gets evidence to support the proposed relations, the implications for practice will increase through the multiplying impact of leaders over the whole system. As thought by Goddard and Skrla (2006), school leaders can work to create conditions that counter institutionalized deficit thinking by changing perceptions of a school about itself.

However, this is attempted under very specific boundaries. A locally low scale scenario designed to observe and maximize 'operative cultures' minimizing the high scale structural influences is prone to produce overstatements. If by design the contextual conditions are controlled no statistical analysis can produce estimates that can be hold accountable for big scale variations. This unobserved structural effect, offers -like a microscope observing a cell- a deeper organizational understanding, neglecting relevant aspect of the overall system –like losing track of the tissue or the organism to which the cell belongs-.

1.4. Hypothesis

The research problem is a prototypical case of observed variables in a path analysis as defined by Kline (2011). Such models allow to explicitly consider measurement errors in the observed variables. Traditional regression analysis ignores these errors in the predictor variables and can lead to misleading substantive conclusions. Further, this method allows the inclusion of complex multivariate mediators that in traditional regressions need to be performed in independent steps. The results yield direct and indirect effects depending on the mediating variables. Finally, such designs can be used to calculate total effects that stand for the explanatory power of those variables (instructional leadership and academic heterogeneity) on dependent variables (collective efficacy and teacher self-efficacy).

This approach is part of the structural equation modeling (SEM) that requires the theoretically grounded proposition of a model that is compared to empirical sample data. The hypothesis is presented through a schematic display with the usual conventions. The numbers on the arrows are there to comment specific hypothesis.

Figure 1: Hypothesized Model
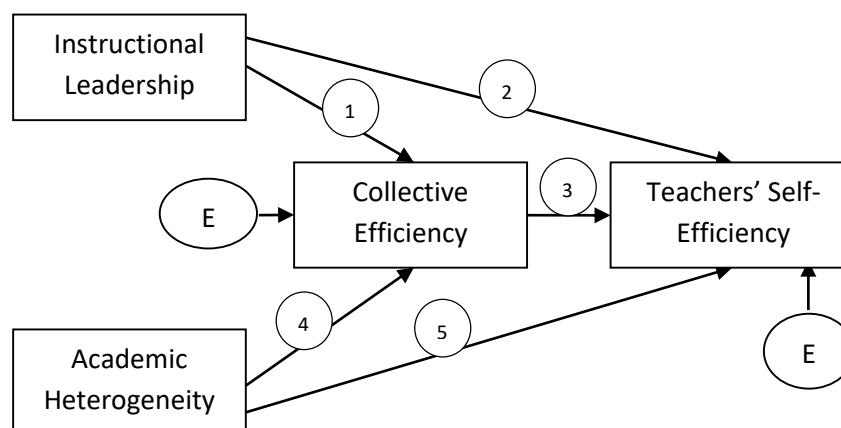
Source: Own design

As depicted, five specific hypothesis rise from this conceptualization, corresponding to the arrows in the figure.

Hypothesis 1: The instructional leadership has a strong association over collective efficacy. As previous finding point out this relation should be relatively high (≈0.5-0.6).

Hypothesis 2: The relations of instructional leadership will have a considerable effect on teacher self-efficacy but lower than the indirect effect, through collective efficacy.

Hypothesis 3: The collective efficacy has a causal effect over teacher self-efficacy (≈0.5) following the rationale presented in the theoretical background information.

Hypothesis 4: Academic heterogeneity might be associated with collective efficacy, but this an exploratory analysis. If social and not academic composition correlations are taken into account an effect size of ≈0.7 could be expected, however social composition is measured along the wholes system. This study relies on the structural equalization so magnitudes have to be smaller.

Hypothesis 5: Academic heterogeneity is associated with teacher self-efficacy as well, however the effect size is smaller than the indirect effect through collective efficacy.

The unit of analysis will be the individual teacher, as for once, the dependent variable is a self-report of an individuals' own self-efficacy. It is assumed that belonging to a faculty will exercise a certain peer-pressure and make some shared impressions on all who belong to the same group, but the experience is still individual. Also, the collective efficacy is usually measured through individual perceptions of the construct, and not through aggregated scores of self-efficacy scores (. Instructional leadership is also a social or group influence inside schools, but the leaders' impact differs greatly between persons in the same school. Therefore, it is interesting to explore in this case the individual effects to. Again, the instructional leadership construct is usually measured by responses given by the lead individuals, because it has been proven to be a more valid (less biased) strategy (Hallinger, 2013). Academic heterogeneity on the other hand is difficult to grasp as an individual construct if not by impressions, that might be more influenced by other thing, like teacher's previous experiences in other schools or other times and a cognitive 'correction' of their own perceived geo-social space. Failing to grasp the 'raw' academic heterogeneity would mean a biased interpretation of the context, that drove this research question. The choice to use comparable, standardized, student achievement as collective scores seems to be more fit.

Teacher's responses in most cases integrate significant analysis of leadership dynamics inside schools, affecting in complex manners the way how they shape their expectations and prepare for work.

1.5.   Research Objectives

The main objective of this work is defined as:

- **Assess empirical validity for the specified model of causal relations between educational leadership and school's contexts over efficacy beliefs in a sample of teachers in challenged schools.**

The empirical validity is obtained through the SEM analysis, that offer goodness of fit indexes that when close to recommended thresholds indicate adjustment between data and model specifications or in other words empirical sources and theory.

From this main objective specific work objectives are derived, that can be understood as follows:

1.7.1. Specific Objective 1: Translate, validate, adapt, and apply the used instruments.

This objective required a piloting study that was performed during 2014 and described specifically in a publication of the results for the scale (Principals Instructional Management Rating Scale – PIMRS) (Fromm, et. al., 2016). The results will be commented in the methodology section that describes the used instruments. The other used instruments have previous validations in Chile that will be commented in the same section.

1.7.2. Specific Objective 2: Achieve a sample that matches the studies requirements.

This second objective required an approach for 3 issues. The first is the general discussion about statistical significance of findings, the second is about sample sizes in SEM and third the representativeness of achieved samples. All these aspects will be described in the methodology section under the samples description in order to qualify the achieved number of cases as sufficient for the validations purposes.

1.7.3. Specific Objective 3: Discuss and determine how the variable school context is defined for this specific sample and measured in this study.

As can be understood form the previous sections this objective defied the academic heterogeneity as a viable specification for school contexts, but its discussion will be presented in the methodology's segment about variables choice and operationalization. The presented arguments will settle academic heterogeneity as the proxy for school contexts.

1.7.4. Specific Objective 4: Specify and overcome the structural equation modeling limitation of sample size, error estimation, reliability, and magnitude prescriptions of the specified correlations.

This specific objective is treated continuously during the presentation of the tested models results, in the corresponding section. In SEM the results need discussion that deepens the findings, discuss the limitations and provide post-hoc analysis solutions in order of appearance. The fulfillment of this objective can be seen during the discussions and finally will be decided in the conclusions, together with the main objective.

1.7.5. Specific Objective 5: Collect and analyze data for the confirmatory analysis.

This specific objective is treated continuously during the presentation of the tested models results, in the corresponding section. In SEM the results need discussion that deepen the findings, discuss the limitations and provide post-hoc analysis solutions in order of appearance. Its fulfillment can be considered together with the main objective and concerning discussions about the process quality.

II.   Methodology

This section goes through the following topics. It starts with a description about the achieved sample and the population it represents. This is done by comparing previously available data of the sample and the universe for schools on which the teachers work and collected data of the study. Also, a discussion about sample size, error, representations and literatures recommendations is presented.

Then it describes the variables starting with the ones considered in the model, both, measured and obtained through secondary data. After that, comes de data collection procedure. This section ends with the analysis plan that describes both, the measurement model assessments and the model testing with the follow-up steps.

2.1.   Sample

One of the problems that this study addresses is the effect of the socio-cultural background of individual students and its heterogeneity inside groups. Since families have made similar choices regarding their children's education, the potential peer group is already adjusted to those choices. One might expect that families are more similar inside a school when compared with other schools. This is, generally speaking, the social composition of the students who attend the school and depends on how it is measured (for example, by racial or ethnic composition, socio-economic status or other cultural variables). Also, the institutional structure of a nation's education system that differentiates public or private schools with fundamental differences in funding and autonomy evidences other dynamics that tells schools apart. This is consistently exemplified in the UK, USA and Chile, in regard to increased academic performance in private schools that, however,

disappears once student SES is controlled (Goddard & Skrla, 2006; Lubienski & Lubienski, 2005). This moderation problem is reduced if large-scale structural differences (advantages or disadvantages) are equalized through sampling, enhancing the perceived school-based differences.

The sample selection responded to a specific setting of large-scale contextual variables. For instance, municipal schools in the Araucanía Region of Chile were selected. As such they are in Chile's poorest region, with the highest rural and indigenous representation (MINEDUC, 2013). Six municipalities (Temuco, Padre las Casas, Freire, Gorbea, Loncoche, and Villarrica) gathered a universe of 119 schools that was reduced to 67 because some had 10 or less teachers in staff and were considered inappropriate for aggregated evaluation. The sample was not strictly randomized, as some schools were in ethnic and politically conflicted areas, lacked the disposition to participate, or access to the location was difficult. In the end, 20 schools were sampled, representing a sample error of 18.5%. However, in the schools' universe worked 2175 teachers, from which exactly 600 were considered in the sample. Considering that individual level data were paramount for the research, at this level the sample error is reduced to 3.4%. The final sample of 366 responses represents 61% of the sample and would signify a raw sample error of 4.7% of the universe.

In the next table we can see a comparison between the universe and the sample in regard to some structural characteristics.

Table 1: Universe and Sample Comparison

	All Schools	Public Schools	Sample
Schools	387	119	20
Percentage of male teachers	27%	30%	33%
School Size (in teachers)	17 (18.2)	18 (17.0)	30 (11.9)
School Vulnerability Index	82% (16.4%)	86% (14.2%)	85% (11.1%)
Teacher's work experience years	14 (6.9)	14 (6.5)	15 (11.2)
Principal's years at school	17 (13.4)	14 (13.5)	13 (10.2)
Rural schools percentage	47% (49%)	53% (50%)	55% (51%)

The standard deviations of each value are in parentheses.

Source: Own design with data from MINEDUC

 The only important difference between the sample and the universe is that bigger schools were selected, which is explained by the exclusion of 52 schools with less than 10 teachers, which lower the universe's average. Other differences (school vulnerability index, principals' years at the school and rurality) are comparably similar to the schools of the universe. In fact, the samples values are closer to their public school universe, than that universe's scores with the scores obtained of all schools including private and subsidized schools. Therefore, the aim to equalize structural characteristics seems to be fulfilled.

 Four other descriptive statistics gathered during the study can be observed in table 2. In this table, first the teachers' age is depicted. Second comes the teachers' salary (in five thresholds, in Chilean pesos: 1=less than $400.000; 2=between $400.000 and $550.000; 3=between $550.000 y $700.000; 4=between $700.000 and

$900.000, and 5=more than $900.000). In third place the answer to the question 'What percentage of socially vulnerable students do you think attend the school?' is registered and named perceived vulnerability index. The fourth and last is the answer to the question 'What percentage of Mapuche students attend the school?'. Mapuches are the region's predominant indigenous people.

Table 2: Sample's Demographics

	N	Min	Max	Range	Mean	S.D.
Age	311	22	71	49	42.63	12.244
Salary	314	1	5	4	3.10	1.209
Perceived School Vulnerability Index	310	0	100	100	77.56	19.714
Perceived Mapuche Attendance	310	5	99	94	53.64	24.930

Source: Own design with collected data

The results describe the sampled teachers as close to the national standards or to previous studies. The perceived vulnerability index correlates with the vulnerability indexes obtained from the national database, higher (r=0.881) at school level than at individual level (r=0.631). This particular correlation is to be expected as individuals vary their own perceptions of their school's index more than their average perception.

2.1.1. Sample size and sampling error recommendations

Much has been discussed in methodology specialized articles about how statistically recommended thresholds for sample sizes and findings reliability should be considered as a continuum and not a dichotomy (e.g. Concato & Hartigan, 2016). In other words, the usual threshold of 5% statistical error should not be applied as a criterion to consider sample sizes enough or not enough. Failing to achieve the 'according' error for determined sample sizes does not produce false results, but only makes them less reliable. Moreover, less reliable does not mean useless results either. As long as its interpretation limitations are presented it is advisable to report 'failing' significance sizes. The standard of 5% is debatable and should be adjusted to the concepts that are being researched in order to evaluate if the findings are or are not contributions to the field (Schmidt, 1996). Conceptually significant data should be considered by their own meaning regardless of associated significance values. Together with the interpretation of the results, possibilities of different circumstances -mostly bigger sample sizes, but other contexts or less population-sample differences- might change the findings should be presented in order to satisfy research purposes even when the 5% standard is not met.

As significance levels depend mathematically on sample size, effect sizes and data dispersion, the results can always be informative. Specifically, in SEM, similarities to asymptotic statistics have been established -meaning that no corrections for increasing sample sizes are done-, reporting that exaggerated sample sizes could reject models and estimations due to minimal magnitude differences (Tanaka, 1987). Despite the discussion, no exact recommendations for SEM analysis have been established, but sample sizes above 200 cases are advisable (Lei y Wu, 2007; Kline, 2011).

Concluding this part, SEM analysis does rely on adequate sample sizes, but no exact requirement can be established. Above 200 cases samples, can be considered to

meet minimal criteria. Moreover, if the models fit indexes (that again, are sensitive for to small but also to large sample sizes) show sustainable results, sample size could post-hoc be considered adequate. The interpretation of the results (including the models effect sizes), again need to be interpreted with their conceptual significances and not under the 'blind' application of a statistical rule.

2.2. Procedure and Data Collection

The sampled schools were visited during faculty meetings by the same researcher to apply the questionnaires. The final sample yield of 366 teachers was obtained after discarding eight cases (2.1%) who didn't sign the informed consent form and 10 cases (2.7%) that had more than 10% missing data among their responses. The functional sample had 0.4% missing data among the 3 measured variables. All loss can be attributed to random causes.

The treatment of missing data should be discussed from methodological and conceptual perspectives (Schreiber, Nora, Stage, Barlow, & King, 2006). The basic handling alternatives are dropping missing data 'pair wise', 'list wise', or replacing them by 'estimation'. In short, list wise elimination of the data excludes a whole case from the analysis, reducing the sample size and risking bias in the data because of the underlying traits of people that produces de missing data. Pair wise elimination tries to minimize de sample sizes shrinkage by keeping the available 'complete pairs' in the analysis excluding only the pairs with missing parts on every step. This decreases less the estimation power, but still risk bias because of the kind of people that is not providing answers. The third way is the estimation –those again through different strategies- and replacement of missing data with the most probable result, in the sample (by averages) or the most similar cases (expectation-maximization EM algorithm). This effectively reduces the risk biases and keeps the

sample size as data are imputated (or assigned to empty slots in the database). The used form of treating the missing data is by default of the Software list wise deletion. Kline (2011) recommends that less than 5% missing data in a sample can be safely assumed to be because of random reasons, so the bias risk is not problematic. MCAR tests that compare the actual missing data with the chi-square probability of missing them by chance confirmed this assumption for the used scales.

The overall high response rate allowed proceeding with the analysis without performing data imputation of missing data, what is a common practice that is seldom reported.

Outliers were found in the scales assessments only for lower scores, but will not be considered or reported because Likert scales' ranges that conceptually should not admit outliers. The closed maximum and minimum scores allow by definition the offered responses, without risk of imposing an outlier effect over the sample.

The outlier analysis of the not closed ranged academic heterogeneity variable resulted in no case beyond two standard deviations above or below the samples mean. This is considered acceptable if not beyond three standard deviations from the average by literatures recommendations (Kline, 2011).

Demographic data had missing data in 31 cases, but those do not affect the SEM model. All responses had a double data entry that avoided 1.5% of typing errors.

2.3. Variables

The models variables are divided in measured variables (instructional leadership, collective efficacy, and teacher's self-efficacy) and secondary constructed variables (academic heterogeneity), presented in this order.

2.3.1. Teacher's self-efficacy

This variable was assessed with the short version of the Teacher Self Efficacy Scale (TSES) developed by Tschannen-Moran and Woolfolk Hoy (2001). Responses are given on a five-point scale concerning the agreement with statements about teachers' own practice, beliefs and perceptions (1 stands for "nothing" and 5 for "a lot"). This questionnaire is also presented in the appendix under part 3.4. The scale has been applied in a Sample of more than 544 teachers in Chile reporting a reliability score of 0.92 (Covarrubias & Mendoza Lira, 2015). The short version consists of 12 items grouped in three dimensions with 4 items each named "efficacy for instructional strategies", "efficacy for classroom management", and "efficacy for student engagement". A measurement error is included in the analysis as this is an endogenous variable. The scale was validated before in another sample of 262 teachers in Valparaíso and Viña del Mar during Covarrubias' (2014) own doctoral thesis, applying an expert panel internal validation strategy and an evaluation about the questionnaire to the same respondents.

2.3.2. Collective efficacy

This variable was measured through Goddard, Hoy & Hoy's collective teacher efficacy scale (Goddard et al., 2000), validated with a Chilean sample of more than 500 participants by Volante (2010). It comprises 21 items assessing an individual's perception of collectively worded efficacy beliefs (example: "Teachers in this

school are able to get through to difficult students"). It is worth of notice, that the perception are individual responses (self-reports) about a collective phenomenon. The scale is presented in appendix 3.3. The items are grouped in two dimensions with some negative items. The dimensions are "group competence" (13 items, 7 of them positive and 6 negative) and "task analysis" (8 items, 4 positive and 4 negative). Negative item scores are reversed before analysis. More specifically, participants report on a six-point Likert scale the degree of agreement with statements about faculty (1 stands for "totally disagree" and 6 for "totally agree"). The reported reliability in that sample is 0,802 (Volante, 2010). As an endogenous variable measurement, errors are specified in the model even if it is a predictor of self-efficacy. The validation in Chilean settings, even if it lacks of more sophisticated internal validation strategies, can be considered achieved through external validation as the findings were consistent with the exposed theory (Volante, 2010) and the international literature (Leithwood, Patten, & Jantzi, 2010).

2.3.3. Instructional Leadership

This first variable is defined by the short form of the Principal Instructional Management Rating Scale (PIMRS) of its Spanish validated form (Fromm, et al., 2016). It consists of 22 items grouped in 3 dimensions and applied to teachers who report on their principal's leadership performance. The three dimensions are "defines the school mission" (5 items), "manages the instructional program" (7 items), and "develops a positive learning climate" (8 items). The instrument is presented in the appendix 3.2. It uses a five-point Likert scale asking about the frequency of specific work tasks in each item (1 stands for "never or almost never" and 5 for "always or almost always"). Its reliability rating tested through Gen Theory (fit for nested responses) is 0.99 in both Chilean and international samples (Fromm, et al., 2016). As one of the exogenous variables, no measurement error is

assumed and therefore sound psychometric properties are required. Besides optimal reliability, validity has been assessed through confirmatory factor analysis (CFA) resulting in good fit indexes and all factor loadings of items ranging from 0.589 to 0.935 (Fromm, et al., 2016). These data derive from an intentional validation study performed during 2014 in a sample of 595 teachers in Santiago under the work for this Thesis.

2.3.4. Constructed with secondary data: Academic heterogeneity

One of the specific objectives requires to make a precision about what is to be considered the variable school context and how it is measured. A definition of what can be considered school's context can be determined by broad ranged aspects like system's structure, economy, society, organizational characteristics, or local culture. All could be considered school contexts, but following this studies pretension to examine closely local variations, broad structural clustering will not be addressed discarding variables at educational system (like curricular track, or public and private management), economy and social stratifications (poverty, SES). To find results about how influence over operative school culture varies a decision was made for academic characteristics. However, isolating an academic aspect of schools, from socially structured aspects is not easy. For instance, social heterogeneity, peer group composition, national school test scores, rural emplacement, and a vulnerability index are highly related to social differences. The choice to use students' performance indicators, but consider them locally for each school grounds the choice for academic heterogeneity for three reasons.

First, it allegedly varies independently from social structural characteristics. In relatively similar schools, where choice and segregation dynamics operate similarly academic heterogeneity should appear as independent from social stratifications. For instance, the correlations with some available variables like national school test

scores (SIMCE) are r= -0.295 and with the perceived vulnerability percentage at the school r= 0,289. This means that in higher performing schools and in less vulnerable schools' heterogeneity is lower. However, only about 9% with correlations close to 0.3 of the academic heterogeneity can be explained by variations of school scores or vulnerability. Also, the correlations with an applied social justice scale, reported discriminated students' percentage, reported Mapuche students' percentage, and teachers' salaries present low correlations (between -0,1 and 0,1). It is safe to assume for academic heterogeneity that it varies independently enough from structural social variables.

Second, as shown in the theory framework, it is supported by a solid research body. This previous research has addressed methodological issues, presented consistent findings (at least during the last ten years) and have related the heterogeneity to a couple of school relevant variables like student's performance (Rui, 2009), citizenship development (Eidhof, 2016) and social skills learning (Gottfried, 2015). It seems a correct choice to think of academic heterogeneity as a possible contextual influence over other school's characteristics like efficacy beliefs and the operative culture that it implies.

Third, the academic heterogeneity is probably a variable that is outside of schools control range. Public schools in Chile do have and will have (Ley 20845, 2015) low decision power over what school intake they receive, however socio-location dynamics and parent's choice still apply. Particularly heterogeneity is a local scale phenomenon as said before, minimizing those dynamics for the observation. Therefore, the notion of a contextual variable for schools holds. Moreover, it is conceptually independent from the principalship, even stronger evidenced by the low correlation that it has with instructional leadership (r=0.128).

All this supposes and advantage in defining the school context by the academic heterogeneity.

A sound proposal for academic heterogeneity requires considering the level of grouping that is addressed. It must not be confused with other heterogeneities like social conditions, curricular tracking, specific interventions or projects, and specific instructional designs. These would be perceived as academic heterogeneity without necessarily being truly academic. Therefore, academic heterogeneity will be defined as the standard deviation of the language performance of the student groups (e.g. Scharenberg, 2012).

To make those performances comparable, the secondary data of the yearly censal language test SIMCE (applied in 4th grade for primary schools and in 10th grade for secondary schools) will be used. The standard deviation is therefore a proxy of the academic abilities dispersion of the students for each school. This variable has been used in previous studies to understand segregation of the school system (Mizala & Torche, 2012; Treviño, et. al., 2014). It is important to note that the school staff does not know individual test scores, meaning that teachers do not necessarily relate directly to this dispersion, as their experience of the classroom might be different. Also, the proxy of the particularly assessed class will stand for the whole school, introducing an important assumption. However, this design is preferred to marks (e.g. Scharenberg, 2012), because it neither is a variable that is perceived by the teachers (reducing social desirability biases) nor controlled by the schools (providing an independently comparable dataset).

A thought about causality between academic heterogeneity as a contextual variable and school intern variables (like efficacy beliefs) is needed. As long as teaching affects academic performance the schools have influence over academic heterogeneity in the some extend. First, heterogeneity is not equal and does not even correlate well with results (r=-0.130; p<0.000 in primary education and r=-0.018 in secondary education, calculated both with national language data). So, if teachers were able to increase the lowest achieving students, without increasing as much the best achieving ones, heterogeneity could be reduced by teaching efforts. On the

other way around, teachers could be neglecting the best achieving students due to unsuccessful practice but without having detrimental effects on the low achieving ones. As this does not seem to be the possible, it is safe to assume that schools do not influence academic heterogeneity (except by expelling and choosing students). At least, teaching and teachers mind frames have no causal explanation that convincingly could affect academic heterogeneity. The other way around the students' diversity is recognized as a relevant factor to shape teacher attitudes and expectations (Cohen, 1991). Finally, the collected data for the academic heterogeneity date from 2014, one year before the data collection of the other variables, giving a 'time' contingency argument for causality between both variables, at least for a cross-sectional perspective.

2.4.   Data Analysis

The dataset was prepared for the path analysis assessing normality and running descriptive analysis, assuring sound reliability and validity values and revising the correlation matrix as the literature recommends (Kline, 2011). Each of the measured variables was analyzed for reliability through Cronbach's alpha and Generalization Theory if required, and for internal validity through a confirmatory factor analysis (CFA). This last analysis specifies every measured variable (teacher's self-efficacy, collective efficacy, and instructional leadership) as a latent variable composed by their respective dimensions and items structures. The calculations were performed using the Lavaan package of R software specifying the WLSMV estimation (Rosseel, 2012).

After having every measure validated the structural model was specified and the analysis was run with the R program using the Lavaan package to (Rosseel, 2012). The literature recommends weighted least squares (WLS or WSLMV in Lavaan)

estimation methods for ordinal data like these and samples near 400 or greater (Muthén, du Toit, & Spisic, 1997, Li, 2016). The goodness-of-fit indexes determined the overall adjustment of the model and the factor loadings were used to understand particular effects. The literature recommends the use some of CFI (>0.8), GFI (>0.8), TLI (>0.8), RMSEA (<0.06), and SRMR (<0.08) goodness-of-fit indexes and the thresholds in parenthesis (Kline, 2011; Lei & Wu, 2007).

III.   Results

The results are organized in descriptive statistics of every variable (including reliability), the explicit measurement model behind it, correlation analyses and finally the model's adjustment.

3.1.   Descriptive Statistics

This section comprises a table with the main statistics of the three measured variables (self-efficacy, collective efficacy, and instructional leadership). Every scale is presented with the corresponding sub-dimensions to expose scale confidence. Every measurement was performed with an exact sample size of 322.

Table 3: Descriptive Statistics of the measurement variables

Variable	Dimensions	Min	Max	Range	Mean	S.D.	Skew	Kurt.	α	Excluded cases in α
Teacher's Self-efficacy					4.09	.460	.162	.855	.854	1.2%
	Efficacy for instructional strategies	2	5	3	4.10	.504	-.140	2.659	.705	0.9%
	Efficacy for classroom management	2	5	4	4.06	.646	-.537	1.092	.867	0.3%
	Efficacy for student engagement	3	5	3	4.10	.576	-.107	.292	.737	0.3%
Collective efficacy					4.16	.632	-.054	.159	.836	4.0%
	Group competence	2	6	4	4.66	.757	-.297	.012	.850	2.8%
	Task analysis	2	6	4	3.66	.744	.085	.127	.516	2.2%
Instructional leadership					3.73	.807	-.551	.521	.947	9.9%
	Define the school mission	1	5	4	4.09	.753	-.882	1.063	.892	2.4%
	Manage the instructional program	1	5	4	3.67	.815	-.463	.100	.877	5.3%
	Develop a positive learning climate	1	5	4	3.61	.862	-.460	-.122	.908	3.4%

Source: Own design with collected data.

As observed in table 3, internal consistency and normality indexes of the variables are all good (α>0.8; skew indexes between ±1; kurtosis indexes between ±1). The dimensions' reliabilities present similar ranges (α>0.7), with the exception of the "task analysis" in the collective efficacy scale, which showed a somewhat lower internal consistency of (α≈0.52). The normality assessment of the dimensions was between ±1 for the skew indexes and between ±3 for the kurtosis indexes, signifying optimal values for further SEM. Even at item level (not presented in the table) normality assessments were unproblematic (skew and kurtosis indexes between ±2).

3.1.1. Descriptive Statistics of Academic heterogeneity

The last variable included in the model requires special attention, as it is obtained through the secondary dataset. Therefore, a comparison between the populations' statistics and the sample's parameters is possible and required. However, some specifications to differentiate the primary and secondary schooling must be considered. The next table shows the sample and population statistics for each case.

Table 4: Descriptive Statistics for Academic Heterogeneity

Academic heterogeneity	Level	N	Mean	Delta	Populations S.D.	Skew	Kurt.
Primary Language	Sample	12	45.1	0.8		.178	-.176
	Population	5656	44.3		7.4	-.216	1.291
Secondary Language	Sample	8	42.1	-0.9		.068	-1.775
	Population	2724	43.0		6.5	.181	.143

Source: Own design with data from MINEDUC.

As can be perceived, the samples' means in all cases are similar to the populations'. Both primary school scores are higher than the national averages, while secondary school scores are lower. This difference means that in primary school more differently achieving kids attend the school, while in secondary schools the students achieve more similar that the nations standard deviations average. The differences between the samples and the populations' means from where they were obtained all have small magnitudes, which is an argument to think that the sampling error are still in "normal" ranges for the given population. Moreover, skewness and kurtosis indexes for the language scores lie within acceptable ranges for further analysis with SEM methodology (Kline, 2011).

3.2. Analysis of the Measured Variables

A central task in a theoretical proposal like this is to define the variables that will be integrated in the framework. The empirically assessed variables of teacher's self-efficacy, collective efficacy and instructional leadership will be presented. The results are reported with the following structure:

a) Structure and relations: Correlations (internal and external)
b) Reliability: Cronbach's Alpha and Gen Theory
c) Internal validity: Confirmatory Factor Analysis

3.2.1. Teacher's Self-Efficacy

The first construct is the teachers' self-efficacy, operationalized in this study by the Ohio State teacher efficacy scale (OSTES) (Tschannen-Moran, & Woolfolk Hoy, 2001). The construct is determined by three dimensions, named 'instructional strategies', 'classroom management', and 'student engagement'. The instructional strategies are composed by items about pedagogical strategies and dialogic engagement with students. The classroom management is based on indicators that ask for disciplinary strategies to avoid students' disruptions. And last, the student engagement asks for the belief about increasing interest to learn among the students.

 Each dimension covers what a teacher believes he or she can achieve with his or her students. Altogether the dimensions are part of a whole efficacy belief. The short form assigns four items to each dimension. The items presented in appendix 3.4 are kept in Spanish to remind the reader of the empirical reactive that produced the data.

3.2.1.a Self-Efficacy: Internal correlations of teachers' self-efficacy

The correlation matrix that is presented in table 10 shows the internal correlations among all items of the OSTES. The significance of each correlation can be read from the table depending on the alpha thresholds. The four-item groups that constitute dimensions are framed in grey areas to facilitate the analysis of item correlations inside versus outside the dimensions. The last row and column depicts the item-test correlations between each item and the tests average.

Table 5: Internal correlation matrix of teacher's self-efficacy

Items	1	2	3	4	5	6	7	8	9	10	11	12	Mean
1	1	.407**	.358**	.361**	.147**	.236**	.186**	.332**	.308**	.249**	.217**	.257**	.528**
2	.407**	1	.453**	.284**	.278**	.318**	.280**	.360**	.297**	.247**	.233**	.092	.543**
3	.358**	.453**	1	.427**	.246**	.275**	.210**	.321**	.237**	.220**	.156**	.118*	.515**
4	.361**	.284**	.427**	1	.357**	.335**	.347**	.399**	.268**	.358**	.205**	.231**	.597**
5	.147**	.278**	.246**	.357**	1	.614**	.593**	.501**	.380**	.347**	.296**	.173**	.652**
6	.236**	.318**	.275**	.335**	.614**	1	.725**	.605**	.454**	.395**	.351**	.230**	.732**
7	.186**	.280**	.210**	.347**	.593**	.725**	1	.608**	.443**	.391**	.395**	.264**	.723**
8	.332**	.360**	.321**	.399**	.501**	.605**	.608**	1	.535**	.477**	.420**	.237**	.759**
9	.308**	.297**	.237**	.268**	.380**	.454**	.443**	.535**	1	.644**	.460**	.316**	.695**
10	.249**	.247**	.220**	.358**	.347**	.395**	.391**	.477**	.644**	1	.556**	.342**	.680**
11	.217**	.233**	.156**	.205**	.296**	.351**	.395**	.420**	.460**	.556**	1	.497**	.637**
12	.257**	.092	.118*	.231**	.173**	.230**	.264**	.237**	.316**	.342**	.497**	1	.529**
Mean	.528**	.543**	.515**	.597**	.652**	.732**	.723**	.759**	.695**	.680**	.637**	.529**	1

Source: Own design with collected data

The shown correlations vary from small $r_{2-12}=0.09$ to large $r_{6-7}=0.73$ effect sizes. No correlation is negative as expected inside a concurrent scale, but inside dimensions the correlations are not always higher than outside them (for example $r_{2-8}=0,36$ compared to $r_{2-4}=0,284$). Only correlations r_{2-12} and r_{3-12} were statistically not significant for $p<0.01$.

Despite somewhat low concurrent correlations among items inside dimension D1 'instructional strategies', the correlations in D3 'student engagement' are middle and in D2 'classroom management' are considerably high. Overall, the constructs' structure could probably hold in this particular sample, because of the consistently sound item-test correlations (all r>0.5).

The following correlation matrix includes the main construct 'teacher self-efficacy, its three dimensions, and the other assessed variables. All scores are obtained as averages. The significance can be observed with their thresholds.

Table 6: External correlations of the teacher's self-efficacy

	Teacher Self-Efficacy	D1: Instructional Strategies	D2: Classroom Management	D3: Student Engagement
Instructional Strategies	,746**	1	,468**	,398**
Classroom Management	,851**	,468**	1	,536**
Student Engagement	,809**	,398**	,536**	1
Instructional Leadership	,293**	,240**	,230**	,237**
Academic Heterogeneity	.050	.096	.039	-.007
Collective Efficacy	,401**	,365**	,336**	,274**

** p<0.01 * p<0.05

Source: Own design with collected data and from MINEDUC

The correlation index between dimensions' ranges from $r_{D1-D3}=0.4$ to $r_{D2-D3}=0.536$. Those imply a good discrimination capacity between dimensions. A concurrent analysis shows fit correlations between the construct and its dimensions $r_{SE-D1}=0.746$ and r>0,8 the other two. Ideally those indexes should have reached values ≈0.85, but still are measuring together a construct. The results show that the items and dimensions are approaching complementary but independent aspects, something that altogether makes sense.

The self-efficacy construct correlates moderately (r=0.29) with instructional leadership and relatively strong with Collective Efficacy (r=0.4). The correlation with academic heterogeneity can be considered inexistent (r<0.1). As could have

been expected the dimensions correlate with the other variables lower than the construct, except for the instructional strategies with academic heterogeneity, but still in considerably the same size range. This implies a close representation of the hypothesis with the exception of the lack of a relation with the heterogeneity.

3.2.1.b. Self-Efficacy: Reliability Assessment

The Cronbach's Alpha through which reliability can be assessed is 0.860 (12 items; 98.4% responses). The individualized dimensions (D1) instructional strategies present a 0.711 (with 4 items and 99.2% responses), the (D2) Classroom Management a 0.861 (4 items and 99.5% responses), and (D3) Student Engagement a 0.758 (also 4 items and 99.7% responses). These can be considered good parameters and are close enough to the original scores in the English version (see Table 11).

Table 7: Teacher's self-efficacy scales reliability comparison

Scale	Present Findings	Tschannen-Moran, & Woolfolk Hoy, 2001
Self-Efficacy	0.860	0.90
Instructional Strategies	0.711	0.86
Classroom Management	0.861	0.86
Student Egagement	0.758	0.81

Source: Own design with data from Tschannen-Moran and Woolfolk Hoy (2001).

Despite the fact that the reliability scores are somewhat lower in this sample than in the original one and similar in the dimension 2, the differences just imply a greater variance that is not explained by the questionnaire. The scores are still in fine ranges for psychometric reliability.

3.2.1.c. Self-Efficacy: CFA

A CFA and its goodness of fit indexes are employed to assess the scale's validity. The obtained indexes are CFI=0.940 and RMSEA=0.073. Those can be considered acceptable fits (and CFI even a high fit). Therefore, the structure can hold a validity analysis for the construct.

Going further, the standardized factor loadings of the items can be observed in the following figure number 2.

Figure 2: CFA of the self-efficacy
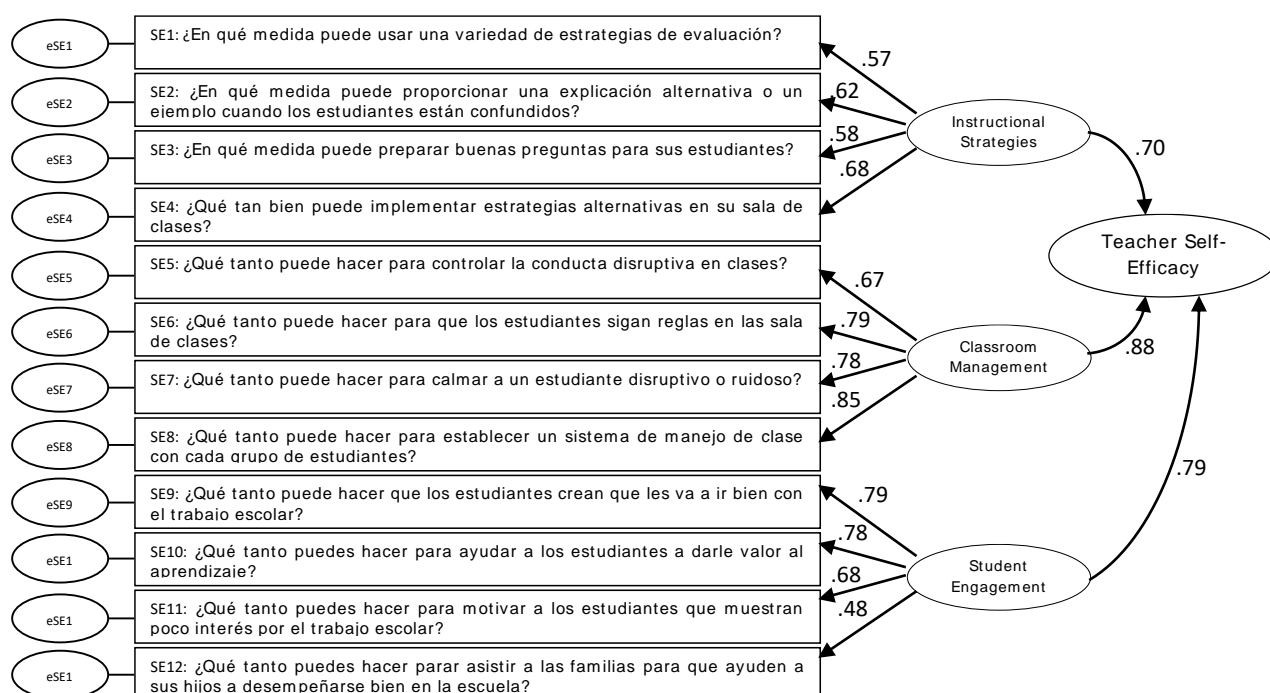
N	Chi²	DF	CFI	GFI	TLI	RMSEA	SRMR
366 (85%)	55.593; p=0.306	51	0.998	1.000	0.998	0.016	0.047
*p<0.01 **p<0.001 ***p<0.0001 WLSMV (DWSL) used							

Source: Own design with collected data.

All of the factor loadings score above 0.6, with items 1, 3, and 12 as exceptions with loadings of 0.569, 0.577, and 0.477 respectively. However, these loadings are insufficient to suggest doubts of the underlying structure and allow to state that the scale is performing a sound assessment of a latent construct as was proposed before.

No re-specifications were necessary as all goodness of fit indexes are in optimal thresholds.

3.2.2. Collective Efficacy

The operationalization of Collective Efficacy in this study was originally the Collective Teacher Efficacy Inventory (CTEI) (Goddard, Hoy & Hoy, 2000) that consists of 21 items grouped in two dimensions named group competence and a task analysis with 13 and eight items each.

While group competence can be considered the 'raw' efficacy belief, the task analysis could wrongly be interpreted as something independent of the group. Theoretically, a competent group could consider a given task as more easy while a less competent group likewise more difficult. However, such an abstract approach to efficacy beliefs is rare (if not impossible). People and groups form their opinions about their own competence in regard to the actual given task. An easy task will not require a great effort, thus lowering the efficacy of the ones performing it. A more difficult task raises expectations to accomplish it; therefore, it will engage people more, increasing their disposition to perform. The combination of both aspects can be understood statistically as a covariance, and constitutes a construct.

It is important to stress out that while it is in fact a 'collective' construct, the responses in the questionnaire were given individually. So the most precise operationalization is an individual perception of the apparent faculty's collective efficacy belief. However, these perceptions are nested in schools, assuming a shared variance because of belonging to a given school and an intra-school variance given

by individual traits. A fit indicator to estimate the amount of collective/individual variance is the intra-class correlation (ICC), that measures how the data relate to each other inside a group. In the case of the scale in this sample the ICC=0.22 meaning that about 22% of the collective efficacy's answers variance that teacher gave depended on the school they belonged to. The rest of the variance is explained by individuality and/or chance.

Coherently the ICC for the two dimensions are 0.17 for group competence and 0.32 for the task analysis. This indicates that the task analysis is, and it should be, much more influenced by the schools' conditions. The group competence's variance is explained by the faculty teachers belong to by 17%, which means that it can be considered a considerable individual appreciation, even if the phenomena is collective. For instance, and to compare the effect sizes, the previous scale of teacher' self-efficacy had a 10% explained variance to, even if there is no evident explanation for a grouping effect.

Of the 22 items, ten are asked negatively -six are part of the group competence and four are part of task analysis-. For interpretation and analysis purposes those items scores are reversed in the database representing a positive appreciation, even when the sentence was negative. The items names are kept in Spanish to remind the reader of the empirical reactive that produced the data.

However, the psychometric analysis resulted less optimal than expected. It was recommended in the literature that the instruments used in further SEM analysis had to be proven valid and reliable (Kline, 2011; Lei & Wu, 2007), but for the present case no evidence against the Collective Teacher Efficacy Inventory (CTEI) was given. Therefore, the best Items that can be theoretically and empirically considered optimal will be kept and the others, eliminated as noise inside the latent construct. Still, in comparison to alternative instruments, this reduction of the CTEI constitutes the better choice for further analysis.

It seems convenient to evaluate stepwise the distinct aspects that are used to determine the best items to be kept. First the item-test correlations will be considered. Then CFA are preformed to confirm the resulting structural soundness of the model.

3.2.2.a Collective Efficacy: Internal correlations of collective-efficacy

The following correlation matrix in table 7 presents the internal correlations among all items of the CTEI. The plus sign after the item number means a reversed score, so a positive (or negative) correlation conceptualizes a direct (or indirect if negative) association. Correlations' significance can be read from the table depending on the alpha thresholds. The items numbers are not kept as in the applied form that can be seen in appendix 3.3, to group them in their dimensions that are framed grey. The last row and column depict the item-test correlations between each item ant the tests average.

Table 8: Internal correlation matrix of collective efficacy

	1	2	3	4	5	6+	7+	8+	9+	10+	17	18	21+	11	12	13+	14+	15	16	19+	20+	Mean
1	1	.394**	.267**	.296**	.298**	.078	.094	.122*	.078	.090	.214**	.242**	-.046	.244**	.117*	.244**	.160**	.449**	.332**	.138**	.126*	.488**
2	.394**	1	.540**	.511**	.505**	.167**	.143**	.225**	.162**	.262**	.395**	.471**	.138**	.210**	.198**	.060	.128*	.204**	.340**	.128*	.030	.517**
3	.267**	.540**	1	.706**	.652**	.308**	.169**	.319**	.297**	.401**	.541**	.601**	.211**	.215**	.147**	.004	.124*	.182**	.260**	.090	.038	.566**
4	.296**	.511**	.706**	1	.775**	.313**	.162**	.399**	.317**	.383**	.552**	.616**	.240**	.273**	.210**	-.007	.207**	.189**	.260**	.117*	.168**	.635**
5	.298**	.505**	.652**	.775**	1	.358**	.193**	.442**	.243**	.378**	.525**	.592**	.223**	.293**	.218**	-.030	.227**	.160**	.295**	.114*	.147**	.630**
6+	.078	.167**	.308**	.313**	.358**	1	.379**	.570**	.417**	.427**	.238**	.278**	.296**	.026	-.152**	.102	.218**	-.037	.104*	.257**	.221**	.481**
7+	.094	.143**	.169**	.162**	.193**	.379**	1	.376**	.226**	.371**	.218**	.253**	.255**	.079	-.010	.052	.221**	-.081	-.023	.186**	.219**	.397**
8+	.122*	.225**	.319**	.399**	.442**	.570**	.376**	1	.459**	.514**	.328**	.319**	.328**	.139**	-.031	-.044	.225**	-.025	.074	.254**	.235**	.530**
9+	.078	.162**	.297**	.317**	.243**	.417**	.226**	.459**	1	.592**	.254**	.248**	.356**	.022	-.082	.085	.232**	-.007	.073	.236**	.161**	.463**
10+	.090	.262**	.401**	.383**	.378**	.427**	.371**	.514**	.592**	1	.392**	.374**	.415**	.138**	-.015	.043	.291**	.030	.120*	.346**	.212**	.590**
17	.214**	.395**	.541**	.552**	.525**	.238**	.218**	.328**	.254**	.392**	1	.735**	.237**	.287**	.219**	-.114*	.192**	.219**	.287**	.149**	.113*	.579**
18	.242**	.471**	.601**	.616**	.592**	.278**	.253**	.319**	.248**	.374**	.735**	1	.243**	.316**	.267**	-.080	.150**	.250**	.306**	.084	.131*	.614**
21+	-.046	.138**	.211**	.240**	.223**	.296**	.255**	.328**	.356**	.415**	.237**	.243**	1	.116*	.008	.039	.244**	-.049	.004	.176**	.227**	.422**
11	.244**	.210**	.215**	.273**	.293**	.026	.079	.139**	.022	.138**	.287**	.316**	.116*	1	.577**	.007	.355**	.202**	.377**	-.007	.203**	.538**
12	.117*	.198**	.147**	.210**	.218**	-.152**	-.010	-.031	-.082	-.015	.219**	.267**	.008	.577**	1	-.088	.279**	.142**	.326**	-.132*	.091	.361**
13+	.244**	0.060	.004	-.007	-.030	.102	.052	-.044	.085	.043	-.114*	-.080	.039	.007	-.088	1	.082	.162**	.083	.128*	.187**	.275**
14+	.160**	.128*	.124*	.207**	.227**	.218**	.221**	.225**	.232**	.291**	.192**	.150**	.244**	.355**	.279**	.082	1	.030	.141**	.237**	.349**	.547**
15	.449**	.204**	.182**	.189**	.160**	-.037	-.081	-.025	-.007	.030	.219**	.250**	-.049	.202**	.142**	.162**	.030	1	.395**	.007	.045	.383**
16	.332**	.340**	.260**	.260**	.295**	.104*	-.023	.074	.073	.120*	.287**	.306**	.004	.377**	.326**	.083	.141**	.395**	1	.019	.130*	.503**
19+	.138**	.128*	.090	.117*	.114*	.257**	.186**	.254**	.236**	.346**	.149**	.085	.176**	-.007	-.132*	.128*	.237**	.007	.019	1	.318**	.391**
20+	.126*	.030	.0377	.168**	.147**	.221**	.219**	.235**	.161**	.212**	.113*	.131*	.227**	.203**	.091	.187**	.349**	.045	.130*	.318**	1	.498**
Mean	.488**	.517**	.566**	.635**	.630**	.481**	.397**	.530**	.463**	.590**	.579**	.614**	.422**	.538**	.361**	.275**	.547**	.383**	.503**	.391**	.498**	1

** p<0,01   * p<0,05

Source: Own design with collected data

As can be seen, the correlation indexes vary between the lowest $r_{12\text{-}19}$=-0.132 and the highest $r_{3\text{-}4}$=0.706 among items. The internal consistency of the items is not as sound as it should. Some correlations are negative (after taking reverse scoring into account), meaning that the scale is registering opposing tendencies in the samples answers (with high significance levels $p<0.01$). Some correlations are close to null like $r_{3\text{-}13}$=0.046 even asking about similar issues ("Teachers in this school have what it takes to get the children to learn" and –reversed– "Teachers in this school do not have the skills to deal with students' disciplinary problems"). These exemplified correlation results are difficult to interpret because many particular cases where strong relations should be found do not show them. This speaks of an inconsistent scale with the disordered dimension structures. However, the item-test correlations support the two latent constructs as eleven of them are in adequate ranges near 0,5 or above. The other ten items have to be considered for elimination, because their item level validity is low.

3.2.1.b. Collective Efficacy: Reliability Assessment

The reliability assessment performed through Cronbach's Alpha's for the remaining 11 items and each of the two dimensions, group competence and task analysis, is 0.845 (11 items), 0.872 (8 items), and 0.550 (3 items) respectively. These are good parameters for the general scales and the group competence, but only acceptable for the task analysis. Moreover, when compared to previous national studies (Volante, 2010) these results are higher for the general scale and group competence and somewhat lower for task analysis, as depicted in table 9.

Table 9: Reliability scores comparison among studies for collective efficacy

Scale	Present Findings	Volante, 2010
Collective Efficacy	0.845	0.806
Group Competence	0.872	0.721
Task Analysis	0.550	0.691

Source: Own Design with data from Volante, 2010.

The former study settled the most equal conditions for the assessed schools in order to find some variance that can be explained solely by the constructs variance among teachers. The same case is occurring in the present study with similar results, making a point in favor of the calculated reliability in this sample.

However, Cronbach alpha as a reliability assessment makes an assumption of randomly distributed respondents in a sample, but this is not the case. In the case of collective efficacy, there is an implicit relationship between a given teacher that is evaluating a specific faculty's belief at the same time as his or her other colleagues in the same school. This distribution is not random, but nested. The reliability therefore should conceptually not be applied to a trait directly distributed in a sample, but to a trait of perceived objects –the collective efficacy belief- organized by groups assigned to each object. This implies a second level between variance that needs to be hold accountable for.

A proposed solution to this problem (Hallinger & Wang, 2015) is the use of the generalization theory (also Gen Theory). This approach relates de reliability concept to the objects characteristics (collective efficacy's mean) rather than the individual's reported traits. The Gen Theory result in this case is 0,754 meaning

that once considered the between variance the reliability of the scale fits in a good enough threshold.

3.2.2.c. Collective Efficacy: CFA

A CFA and its goodness of fit indexes are employed to assess the inventory's validity. The obtained indexes are all well fit (CFI=0.985, RMSEA=0.042) as CFI>0.9 and RMSEA<0.05 are recommended. As no previous CFA report for this scale had been found, the items choice for this analysis followed a more flexible approach. As such a first attempt the empirically yielded scores have no precedents so it counts almost as an exploration of the scales validity. Statistically speaking the confirmatory comparison between the theoretical model and the empirical data fit well for the eleven selected items.
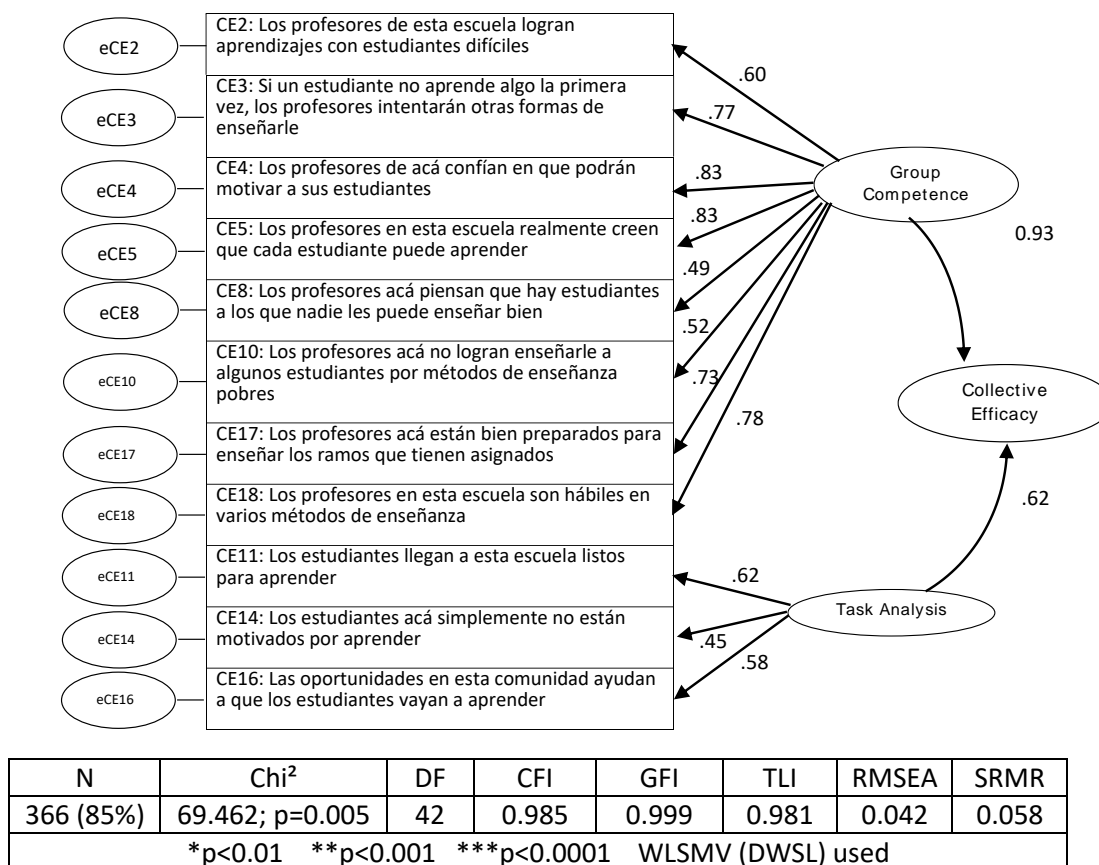
Moreover, the standardized factor loadings of the items are presented in figure 3. Two of those effects sizes are borderline (less than 0,5) for a structural model. As, in the original version (Goddard, Hoy, & Hoy, 2000) all factor loadings were reported to be between 0.47 and 0.87 this can be considered as in an acceptable range, although they were obtained with an exploratory factor analysis. These are methodologically very different calculations, so loadings derived from each one are not directly comparable, but some degree of consistency is to be expected. In the present case the other nine items had loadings above 0.5. The designers of the scale implied solid recommendation to exclude the problematic items from the questionnaire and keep on with the general hypothesized model through the remaining items (Goddard, Hoy & Hoy, 2000). These kind of modifications need first to make theoretical sense, which does. Excluding from the analysis items that ask for general statements that people can answer independently from their future success beliefs seems to be a right approach. However, a contrary argument disregards post hoc analysis with SEM techniques because it increases the

likelihood of type 1 -false positive- errors. This kind of approach has been referred "exploratory journeys" (Schermelleh-Engel, et. al., 2003), "fishing for statistically significant results" (Schreiber, et. al., 2006), or are implied to be unsuited (e.g. Kline, 2011; Kaplan, 2008).

The scale, presented good fits only after carefully selecting the remaining items. As it has been pointed out before, the overall tested model will be calculated with all scale items as observed variables, and the dimensional structures as intermediate latent variables, composing the presented latent variables (e.g. Kaplan, 2008). This would mean that a bad fit of a particular part of the model would be masked in the overall evaluation by its better fitted parts. This could happen the other way around too, as a poorer part like the unsound items could decrease the better fits of its relations.

One of the important points in SEM analysis (and probably in any statistical research) is to keep the course that is theoretically set and not attempt post-hoc trials until satisfactory results appear. In this case, the complex model was chosen due to the theoretical underpinning like developed operationalizations through the instruments and the particular dimensions that have been proposed. However, in the case that a part of the scale results in not useful noise that also carries a risk of mislead correct parts of the analysis the decision to eliminate them seems to be clear. Moreover, if the remaining part of the scale works so well in a conceptual and empirical fashion. The fact that this particular instrument for collective efficacy risked to introduce unclear results into the model is just part of the limitations that have to be addressed in a research process.

Figure 3: CFA of the collective efficacy

N	Chi²	DF	CFI	GFI	TLI	RMSEA	SRMR
366 (85%)	69.462; p=0.005	42	0.985	0.999	0.981	0.042	0.058
*p<0.01 **p<0.001 ***p<0.0001 WLSMV (DWSL) used							

Source: Own design with collected data.

3.2.3. Instructional Leadership

Finally, the instructional leadership's measurement model is shown in the Figure 1, consisting in 22 items grouped in three dimensions: 'defining school mission', 'managing the instructional program', and 'developing a positive school learning climate'. Defining the school mission is represented by tasks like goal setting, structuring the schools' efforts, talking with teachers about them and considering learning results in order to take decisions. The instructional program management is defined through items that ask for the principals' engagement in things like teaching material design, results analysis, and professional conversations. The final dimension is structured through items that broadly explore instructional time protections, principals' visibility, and conversations to enhance teachers and students' development among other.

These dimensions are constituted by five, seven and ten items respectively. Together, the three dimensions conform the Instructional Leadership construct. The names of the items depicted in the appendix 3.2 are kept in the written form in Spanish on purpose, in figure 1, to determine the empirical reactive that produced the data (Fromm, et all., 2016).

3.2.3.a. Instructional Leadership: Correlations

Table 5 show the correlation matrix of the internal correlations among items. The grey zones identify the correlations that are part of the same dimension. The last column and row depict the item-test correlations. Significance levels are not shown because all were low (p<0,001) without exception.

Table 10: Internal Correlations of instructional leadership

Items	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Mean
1	1	,644	,684	,581	,621	,560	,372	,585	,599	,456	,359	,430	,543	,422	,387	,381	,374	,343	,382	,408	,487	,457	,704
2	,644	1	,666	,473	,576	,584	,403	,457	,564	,508	,408	,422	,502	,368	,343	,334	,340	,306	,368	,367	,484	,432	,675
3	,684	,666	1	,601	,638	,616	,435	,557	,583	,513	,384	,431	,579	,424	,363	,370	,332	,372	,394	,445	,425	,397	,715
4	,581	,473	,601	1	,658	,583	,335	,500	,534	,386	,313	,321	,479	,370	,391	,396	,363	,346	,365	,401	,428	,406	,652
5	,621	,576	,638	,658	1	,659	,454	,561	,586	,485	,397	,435	,524	,372	,362	,389	,332	,334	,356	,397	,453	,402	,702
6	,560	,584	,616	,583	,659	1	,400	,560	,552	,497	,347	,409	,542	,364	,304	,301	,341	,313	,384	,351	,377	,396	,664
7	,372	,403	,435	,335	,454	,400	1	,372	,494	,630	,607	,598	,405	,348	,307	,386	,410	,418	,397	,409	,514	,544	,683
8	,585	,457	,557	,500	,561	,560	,372	1	,583	,470	,344	,340	,454	,345	,409	,346	,317	,293	,389	,373	,375	,360	,634
9	,599	,564	,583	,534	,586	,552	,494	,583	1	,594	,469	,490	,581	,367	,408	,411	,453	,471	,543	,514	,540	,480	,769
10	,456	,508	,513	,386	,485	,497	,630	,470	,594	1	,624	,619	,538	,308	,368	,393	,423	,434	,435	,423	,515	,578	,741
11	,359	,408	,384	,313	,397	,347	,607	,344	,469	,624	1	,606	,475	,442	,350	,511	,496	,479	,358	,481	,554	,577	,715
12	,430	,422	,431	,321	,435	,409	,588	,340	,490	,619	,606	1	,527	,325	,352	,334	,403	,409	,425	,469	,515	,516	,696
13	,543	,502	,579	,479	,524	,542	,405	,454	,581	,538	,475	,527	1	,461	,354	,438	,431	,434	,460	,453	,486	,493	,728
14	,422	,368	,424	,370	,372	,364	,348	,345	,367	,308	,442	,325	,461	1	,447	,525	,368	,450	,328	,483	,506	,472	,629
15	,387	,343	,363	,391	,362	,304	,307	,409	,408	,368	,350	,352	,354	,447	1	,414	,426	,362	,422	,390	,367	,369	,585
16	,381	,334	,370	,396	,389	,301	,386	,346	,411	,393	,511	,334	,438	,525	,414	1	,618	,527	,403	,379	,500	,493	,659
17	,374	,340	,332	,363	,332	,341	,410	,317	,453	,423	,496	,403	,431	,368	,426	,618	1	,591	,543	,490	,534	,523	,681
18	,343	,306	,372	,346	,334	,313	,418	,293	,471	,434	,479	,409	,434	,450	,362	,527	,591	1	,621	,581	,573	,538	,689
19	,382	,368	,394	,365	,356	,384	,397	,389	,543	,435	,358	,425	,460	,328	,422	,403	,543	,621	1	,596	,481	,458	,674
20	,408	,367	,445	,401	,397	,351	,409	,373	,514	,423	,481	,469	,453	,483	,390	,379	,490	,581	,596	1	,611	,511	,704
21	,487	,484	,425	,428	,453	,377	,514	,375	,540	,515	,554	,515	,486	,506	,367	,500	,534	,573	,481	,611	1	,718	,763
22	,457	,432	,397	,406	,402	,396	,544	,360	,480	,578	,577	,516	,493	,472	,369	,493	,523	,538	,458	,511	,718	1	,745
Mean	,704	,675	,715	,652	,702	,664	,683	,634	,769	,741	,715	,696	,728	,629	,585	,659	,681	,689	,674	,704	,763	,745	1

Source: Own design with collected data.

As can be seen the correlation indexes vary between slightly below 0.3 (r_{8-18}) to slightly above 0.7 (r_{21-22}). The lower values are found between items from different dimensions giving an argument for good discriminant validity and the higher ones between items of the same dimension supporting convergent validity.

The item-test correlations are all above 0,6, except for item $r_{15-mean}$=0,585, which is still good. This signifies a sound validity at item level and does not need to correct by omitting items at any case.

The main construct 'instructional leadership' and its three dimensions' correlations are presented next, along with the external correlations with other variables in table 6. All scores were obtained as averages.

Table 6: External correlation matrix for instructional leadership

	Instructional Leadership	D1: Defining School Mission	D2: Managing the Instructional Program	D3: Developing a Positive School Learning Climate
Defining the School Mission	.829**	1	.740**	.653**
Managing the Instructional Program	.920**	.740**	1	.765**
Developing a Positive School Learning Climate	.937**	.653**	.765**	1
Academic Heterogeneity	.128*	.183**	.210**	.022
Collective Efficacy	.389**	.361**	.376**	.336**
Teacher's Self-Efficacy	.293**	.308**	.303**	.222**

** p<0.01 * p<0.05

Source: Own design with collected data and from MINEDUC

The correlation indexes between dimensions range from $r_{D1\text{-}D3}=0.653$ to $r_{D2\text{-}D3}=0.765$. Therefore, all three scores evaluate different aspects of the construct. At the same time, the three dimensions are strongly correlated to the instructional leadership (r>0.8), underpinning a concurrent scoring of the different aspects of the construct. The general construct correlates differently with the other variables ranging from a weak correlation with academic heterogeneity (r≈0.13), over a considerable magnitude with teachers' self-efficacy (r≈0.3) to an almost strong one with collective efficacy (r≈0.39). Among the dimensions the third (D3) –managing the instructional program- correlates more strongly than the general construct with every other variable (or the same as D1 with self-efficacy). As this dimension considers the daily work engaged in teaching processes, this association can be related to spurring the teachers. Particularly, academic heterogeneity seems to require more instructional engagement as it correlates with the dimension even higher than with the whole construct. The other dimensions (D1 and D3) have lower correlations with –defining the school vision- the closes to D2 correlations. Of particular note is the low correlation (r≈0.02) between D3 -developing a positive school learning climate- and academic heterogeneity, which gives light that principals actions in this matter are much more independent from the students'

characteristics. The learning climate however is correlated higher with efficacy beliefs (r=0.336) probably because it's similar grounding idea of being operative organizational cultures.

3.2.3.b. Instructional Leadership: Reliability Assessment

Following the data description, the Cronbach's Alpha for the scale is 0.947 (22 items; 88.9% responses); for Defining School Mission is 0.888 (5 items; 97.6% responses); for Managing the Instructional Program is 0.875 (7 items; 93.2% responses); and for Developing a Positive School Learning Climate is 0.904 (10 items; 95.4% responses).

Also in this case the reliability assessment is assuming a random distribution of the instructional leadership in the sample, while it is nested. The teachers' responses are related to their specific principals. The rational for the solution is similar to the on presented in part 3.2.2.b if needed to recapitulate. The result is a reliability score of 0,985 that is interpreted the same way as Cronbach's alpha. Therefore, also thus statistic yielded a sound reliability indicator.

3.2.3.c. Instructional Leadership: CFA

To test the internal validity of the scale a CFA is employed (results shown in Figure 4) and its goodness of fit is compared to the English version's one with samples of 4370 teachers (Hallinger & Wang, 2014). The reported indexes of the English short form are RMSEA=0.088 and an unspecified overall goodness of fit index of 0.965, that in the Spanish version yielded even better ranges (RMSEA=0.027 and CFI=0.996).

Together, these values can be considered as a sound fit between the data and the proposed framework. Considering that the aim is to identify a pattern of results where dimensions measure related but different conceptual constructs subsumed under instructional leadership, the task can be considered achieved.

The standardized factor loadings of the items are presented in figure 2 and were all significant (p<0.001).

These results are not as similar to the English version as the goodness of fit indexes, since no loading was below 0.7. However, the eight descended loadings are all above 0.5 and seven of them above 0.6 which for a factor analysis can be considered good in any case.

Figure 4: CFA of the instructional Leadership
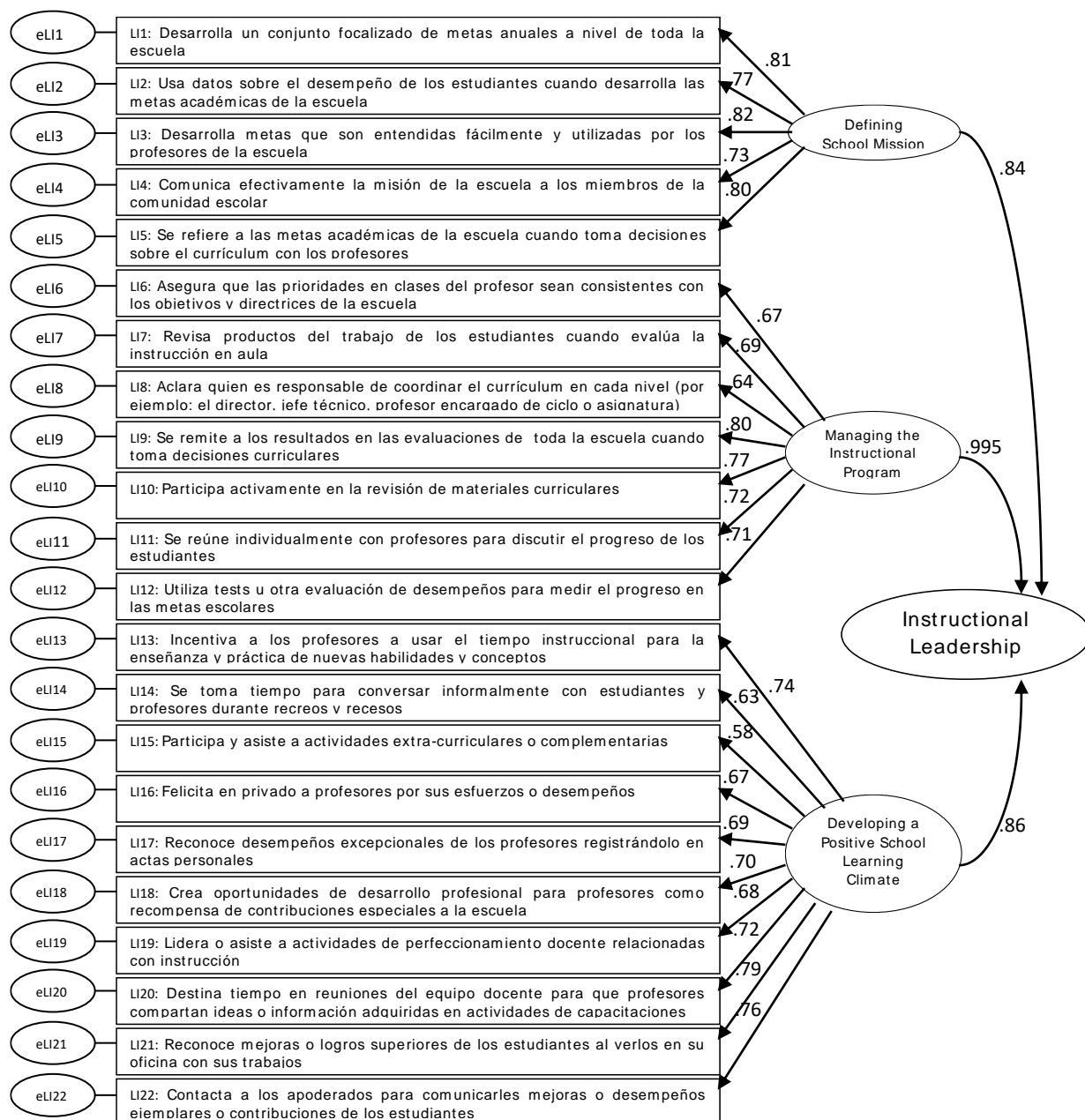
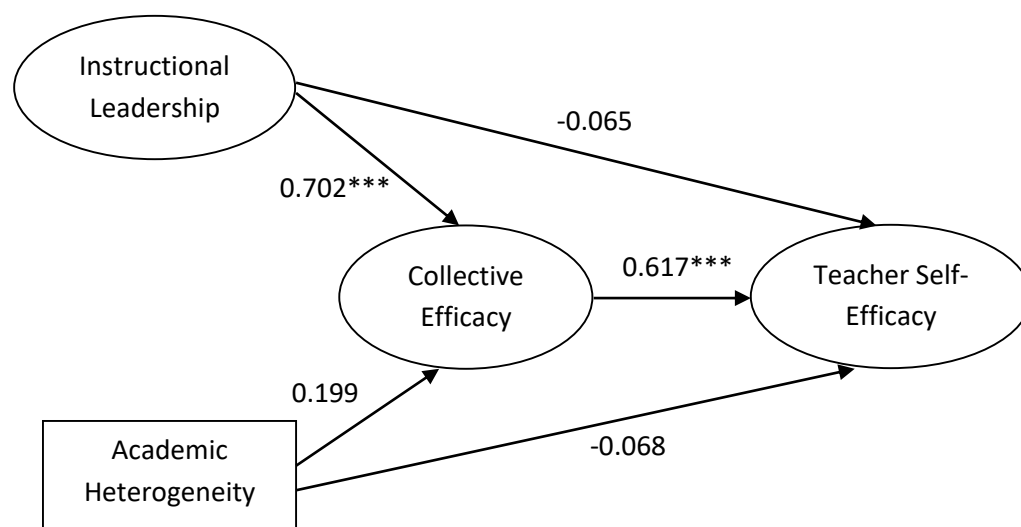| N	Chi²	DF	CFI	GFI	TLI	RMSEA	SRMR
366 (85%)	259.81; p=0.007	206	0.996	0.998	0.995	0.027	0.061
WLSMV (DWSL) used							

Source: Own design with collected data.

3.3. Model testing

The model that has been hypothesized is tested under the following assumptions. The model considers two exogenous variables (instructional leadership and academic heterogeneity) and two endogenous variables (collective and self-efficacy). Three of the four variables, excluding academic heterogeneity, are latent variables underlying the measurement models presented before with 22, 21 and 12 items respectively, constituting a second order model. Therefore, to really grasp the implications of the model, a graphic about each item, loading on its dimension, and each dimension loading on its variables, would be the extensive representation of the following analysis. For interpretability purposes and, because the underlying measurement models do not help to directly answer the research question, the variables are presented as latent variables without the representation of how they are constructed, as can be seen in figure 5.

The calculating method is weighted least squares (WLS or WSLMV in Lavaan) recommended for ordinal scales like the ones used to gather the main body of data (Muthén, du Toit, & Spisic, 1997, Li, 2016). All computations were performed with the Lavaan package for R Software. The following diagram and the table below show the main results, including the goodness of fit indexes of the analysis.

Figure 5: Model testing

N	Chi²	DF	CFI	GFI	TLI	RMSEA	SRMR
366 (85%)	1152.64***	977	0.992	0.997	0.992	0.022	0.058
*p<0.01 **p<0.001 ***p<0.0001 WLSMV (DWSL) used							

Source: Own design

The results of the goodness to fit indexes are excellent, even more considering the lack of previous findings concerning this methodology in the field. The WLSMV estimation adjusts in every index above 0.9 (or below <0.05 for RMSEA and below 0.08 for SRMR). Every recommended range in the literature is optimally met. Even more, the calculated confidence intervals met the recommended criteria for all indexes, considering even the lowest intervals for CFI, GFI and TLI or the highest interval for RMSEA and SRMR. The goodness of fit can therefore be considered reliable, underpinning the whole model's fit.

Concerning the factor loadings between variables, a clear mediation effect of the collective efficacy over the interaction between instructional leadership and self-efficacy is shown. In fact, the direct effect of instructional leadership not reliable

(p=0.555), negative and of low magnitude (-0.087), but indirectly is considerably high (0.433 is the product between 0.702 and 0.617).

At the same time, instructional leadership shows a stronger effect over collective efficacy than academic heterogeneity (0.199). In fact, this variable's supposed effect is not statistically significant (p=0.031). The low result cannot be reliably trusted with current data. The fact that it is a positive result can be considered a hint in regard to reflecting what previous findings have discussed. The almost inexistent relation between academic heterogeneity and self-efficacy (-0.068; p=0.451) requires additional explanations if at the collective level covariances seem to be higher.

3.3.1. Hypothesis testing

Five work hypothesis were established in regard to the loadings found in the model.

Hypothesis 1: The instructional leadership has a strong association over collective efficacy. As previous finding point out this relation should be relatively high (≈0.5-0.6).

Effectively this finding was beyond the hypothesized and reported effect sizes in the literature. The hypothesis nevertheless stands true, and the increased perception might be explained by better evaluation instruments, closer constructs or chance.

Hypothesis 2: The relations of instructional leadership will have a considerable effect on teacher self-efficacy but lower than the indirect effect, through collective efficacy.

This hypothesis is only partially true, as the association between both variables by itself is depreciable, but true in the indirect validity of the relation. Further analysis,

that confirm in different samples the indirect effect's stability and does not yield counterevidence for the direct almost inexistent effect will be useful.

Hypothesis 3: The collective efficacy has a causal effect over teacher self-efficacy (≈0.5) following the rationale presented in the theoretical background information.

The presented study confirms what was already known of the association between both variables and makes the hypothesis valid. The causality of the variables will be treated in the discussions.

Hypothesis 4: Academic heterogeneity might be associated with collective efficacy, but this an exploratory analysis. If social and not academic composition correlations are taken into account an effect size of ≈0.7 could be expected, however social composition is measured along the wholes system. This study relies on the structural equalization so magnitudes have to be smaller.

As an exploration of the issue this hypothesis is true, yielding a result that is not statistically significant but in the supposed way. Understanding that the effect size had to be lower is correct, but the reliability of the finding is unsound. Therefore, the discussions will treat this issue.

Hypothesis 5: Academic heterogeneity is associated with teacher self-efficacy as well, however the effect size is smaller than the indirect effect through collective efficacy.

Again, as in Hypothesis 2, this one is partially true. The exploratory part of finding an association is not, as the result is to small and not statistically significant to make a statement, but the indirect effect size comparison is.

IV.   Discussion

This section is dedicated to analyze the presented evidence. The first section discusses the measurement models and psychometric properties of the used instruments. Then the main issues concerning the tested model are discussed in depth during eight points.

4.1.   Measurement models and instruments

The first issue is to discuss how the actual sample is related to previous research and can hold a claim to provide sound data for the analysis. Through the provided results insight has being gained for each of the variables, which will be approached in order.

4.1.1.   Discussion for teacher's self-efficacy

This construct was operationalized by the OSTES scale. Although its research field is not as developed as the leadership's research notably in Chile is a line of research using the construct and scale. Therefore, the samples results can be compared with the ones yielded by Convarrubias and Mendoza (2015) in Santiago. Only the averages obtained can be compared, depicting the total estimators and those separated by gender in Table 12.

Table 12: Data comparison of the teachers' self-efficacy assessment

	N		Mean		S.D.	
	Sample in Araucanía	Sample in Santiago	Sample in Araucanía	Sample in Santiago	Sample in Araucanía	Sample in Santiago
Teacher's Self Efficacy	353	544	4.10	4.34	.46	.44
Females	240	382	4.15	4.33	.44	.44
Male	113	162	3.98	4.36	.48	.44

Source: Own design with data from Covarrubias and Mendoza (2015).

The results in both samples are highly consistent showing a little bit lower averages but more dispersion than in the sample of Santiago. This is in normal ranges and could be explained as much by implicit traits (like city's inhabitants' characteristics) as measurement differences. The gender seems to work differently in both samples, reducing the effectiveness average but increasing its dispersion for males in Araucanía. It could be worth exploring in the future how gender affects the presented results.

4.1.2.   Discussion for collective efficacy

Also, the adjusted scale for collective efficacy provides a structurally sound construct. The scale, presented psychometric solid operationalization only when a the theoretically and statistically 'better' items were considered. Although, both dimensions were kept intact and confirmed by the factor loading and goodness of fit indexes results. Moreover, in the past literature, the scale was almost only used as a general latent variable. This could be observed in the original paper that describes the scales design and only used concurrent and discriminant correlations with other constructs as the validation strategy (Goddard, Hoy & Hoy, 2000). The same happened in the Chilean application that explored its mediating relation for the influence of instructional leadership over students' results (Volante, 2010). The recommendations derived from this are to ensure the best possible instruments

through careful analysis of the psychometric properties to approach the construct in the most solid fashion.

As previous findings have consistently enough associated collective efficacy to students' performance improvement (Volante, 2010) and organizational development for pedagogically relevant practices (Fromm, Volante, & Olbrich, 2015). Instead of concluding that collective efficacy was not grasped correctly with the original scale, the opposite should be considered. The best part of the scale grasped the construct in a way that previous work finds out to be a relevant variable to understand what happens inside schools and how do leaders influence teaching and learning. In fact, it is so relevant that lacking optimal initial design a new discussion arises pressing for more and better assessments methods, instruments and strategies.

4.1.3. Discussion for instructional leadership

Instructional Leadership has proven to be a solid construct with a mature research field and optimal psychometric operationalization trough the PIMRS. The overall results of this sample can be related to the ones obtained in the instrument validation study (Fromm, et al., 2016) showing somewhat higher averages and lower dispersion than in a sample in the city of Santiago. For the comparison see table 13.

Table 13: Data comparison of the instructional leadership assessment

	N		Mean		S.D.	
	Sample in Araucanía	Sample in Santiago	Sample in Araucanía	Sample in Santiago	Sample in Araucanía	Sample in Santiago
Instructional Leadership	366	595	3.73	3.55	.81	.77
Defines the school mission	366	595	4.09	3.93	.75	.76
Manages the instructional program	366	595	3.67	3.59	.82	.89
Develops a positive learning climate	366	595	3.61	3.43	.86	.89

Source: Own design with data from Fromm, et al (2017)

As it may seem, in Araucanía leaders have tended to be perceived as consistently performing better, the differences are little and in range of normal dispersion for each sample. Moreover, the general structure is consistent, the dimensions 'defines the school mission' always being the highest and 'develops a positive learning climate' always the lowest. This consistency is not a grounding argument but more a pointer that in different (and in this particular) Chilean locations the scale makes similar sense.

The measurement structural model stands for itself considering internal correlations, factor loadings and goodness of fit indexes. It is advisable to keep track of similar SEM studies in order to check for the obtained magnitudes. As presented before, the correlation with collective efficacy of .389 is lower than the expected r=.52 (Leithwood, Patten, & Jantzi, 2010) in Canada, r=.54 (Volante, 2010) in Chile. However, it is quite similar to the effects of β=.34 (p<.01) presented in Turkey (Calik, et al., 2012). It is worth noting that the Turkish finding derives from a simple regression analysis that means that the correlation index (r) and regression coefficient (β) can't be compared as equal in sizes. The difference with the other sources can be explained by the implied construct in the different questionnaires. While the PIMRS ask directly for frequency of principals' practices the instruments used in Canada asks for teachers' opinions about the principal (Leithwood, Patten, Jantzi, 2010) and in Volante's sample (2010) asked rephrased for perceptions about the whole school management teams. The SEM studies in Oman and Iran used the same PIMRS and obtained more similar results (0.58 and 0.60) to the presented ones (Hallinger, Hosseingholizadeh, Hashemi, & Kouhsari, 2017; Al-Mahdy, Emam, & Hallinger, 2018). The collective efficacy scale does the same questioning about opinions, opening the chance to correlate through social desirability.

The Turkish study also reported a relation between instructional leadership and teacher's self-efficacy, again as a regression coefficient β = .32, (p < .01) that is again quite similar to the found correlation of .293 (p < .01). Also, in Turkey a study

related transformational leadership with teacher's self-efficacy with a size r=.44 (Demir, 2008), which is consistent with the proposal that a common factor that could be social desirability augments correlations when no practice-related constructs are measured. The same care about not confounding directly r with β needs to be reminded.

4.2.   Tested model, theory discussion

The accomplishment of the empirical validation's main objective is discussed next. First the informed goodness of fit indexes showed a well fit. As Kline (2011) recommended, a sound structural equation model requires as show earlier that its variables have sound psychometric properties making a consistent analysis possible.

4.2.1.   Instructional leadership's influence on efficacy beliefs

As the weight results show, the tested model can be considered as a valid finding for a relation of instructional leadership over efficacy beliefs, in more or less the same way as the literature suggested (López & Gallegos, 2014; Calik, Sezgin, Kavgaci, & Kilinc, 2012; Demir, 2008; Leithwood & Jantzi, 2008). Even as this can be considered an expected finding due to the presented evidence in the first part, it had not been performed in this fashion. Even as it is not particularly new to consider that school leaders work to influence teachers to change their minds and points of view about their teaching practice, an empirical finding is relevant for two reasons.

The first is conceptual, as it helps leaders find strategies to create a high expectation culture as collective efficacy indicates. If leaders can understand that the perception

of group competence and a realistic but challenging task analysis has to be fostered and made explicit among teachers, they can engage in professional conversations with a clearer aim.

The second relevant factor is scholarly, and helps to understand how effects dynamics take place in schools. For instance, if researchers need to decide what variables should be included in their own observation frameworks, or to relate to particular proxies or effects, these findings provide solid antecedents. In this sense, the measurement of this influence path is the first major finding in this thesis.

This helps understand how secure a confirmatory analysis is when grounded on a solid body of research, with accumulated measurement guidelines, little controversy and preceding pointers.

4.2.2. Academic heterogeneity's influence over efficacy beliefs

On the other side, the way student differences affect how teachers perceive their own future performance with them is less evident. The results show a positive but marginal influence over collective efficacy that is in accordance to the previous literature. The evidence supporting that heterogeneity exercises a positive influence through peer pressure over collective efficacy (Goddard & Skrla, 2006; Raudenbush, et al., 1992) proved to be different in this study.

To test if it would constitute an influential factor over efficacy beliefs seemed to outreach what the literature had been suggesting. In any case, the reliability of the loading (p>0.01) was insufficient to be sure that sample size changes would not affect the findings. However, those changes might or not alter the found effect sizes decreasingly or increasingly, so no sure ground was found in this regard. Further research would be highly recommended as academic heterogeneity is a developing field that enables to attempt alternative measuring and conceptualization strategies.

Finally, the actual design imposes a notorious limitation to the definition of the findings as settling results because, as commented earlier, the way academic heterogeneity is included in the analysis is a proxy for the pursued school phenomena, derived from an external source and not reported by the teachers themselves, and a construct under conceptual development. To be more specific, this variable is a school observation in a standardized test that has been used to approach an individual response with less adjustment to the real variance than the underlying phenomenon might present. As such, covariances between these variables and efficacy beliefs could fall in totally different ranges if different methodological decisions are made. Nevertheless, this is a second major finding in this work, as it explores some interesting findings in a field that still has much to achieve.

Contrary to the previous idea, this finding underpins the need to rely on a solid work to perform confirmatory analysis if the results are to give definitive answers. The controversies that were presented (e.g. Slavin; 1987; Kulik & Kulik, 1987; Kulik, 1992; Lou et al., 1996; and Rui, 2009) even if theoretic tendencies seem to have stronger claims (Wilkinson & Fung, 2002) are part of uncertainties that concepts bear and have proven to be present in this work too.

As pointed put before, attention to the described sociocultural perspectives of schools and teachers in heterogeneous contexts needs to be paid (Cohen, 1991; Persell, 1977). Within-variance explanations are biased by the variables choice and as long as the block box of what principals and teachers need to address in regard to more different student groups, the effects will be prone to simplistic interpretations. This thesis contributes in this matter, attempting to understand how efficacy beliefs are affected in this sense. Despite that the results are statistically unreliable a careful interpretation can be provided. Considering that the reliability indicator ($p=0.031$) was not so different from the recommended threshold ($\alpha=0.01$) and other perspectives about not ruling out 'findings' about effect sizes just because

the reliability does not meet the convention (Schmidt, 1996) a discussion seems in order. The findings need to be considered in its just measure, meaning that the effect size of 0.199 seems to indicate that the association is positive. Therefore, teachers seem to develop a sense of efficacy if the academic differences of the students are higher. Of course, this causal relation cannot rule out inversed or third factor explanations. This is consistent with the international literature about the general students' (Vandenberghe, 2002), and has been addressed in the variables section of the methods. Of course this is not a robust statement but more an orientation for future research about school intern dynamics related to students' academic heterogeneity that in the light of these analysis gains much more relevance.

4.2.3. Comparing instructional leadership and academic heterogeneity's influences

The third finding is the comparably bigger effect of the leadership versus the students' heterogeneity 'paths' over efficacy beliefs. As stated in the theoretical framework, an unsolved problem in the literature and, even deeper, on the research agendas, is the comparable influence that a leader can exercise in specific contexts. More specifically, differently competent leaders are seldom compared in regard to their practices and achievements without interfering school level or structural factors. In this attempt to equalize structural and school level characteristics, individual variances like instructional leadership will be perceived as stronger. Despite the fact that academic heterogeneity was conceptualized to maximize the observation of a between school variance (using standardized test results and a proxy for the school), the literature has not stated how heterogeneity is composed along the system (e.g. Mizala & Torche, 2012), the segregation mechanisms (e.g. Bellei, 2013), inside schools (e.g. Treviño, et al., 2014), and at classroom level (e.g. Scharenberg, 2012). Such an integrated understanding would give strong guidelines

in order to explore effects and their sizes and to hypothesize relations with other variables.

4.2.4. Causality of the collective over individual efficacy beliefs

Another finding with some relevance for the literature is the effect inside efficacy beliefs constructs or levels. Coming back to the discussion about how collective and individual levels of efficacy beliefs influence each other, the results help a lot to settle the matter. The influence of both exogenous variables over the collective level is significantly stronger than over teachers' self-efficacy. Considering that the implicit indirect effects hypothesis of both exogenous variable effects over the teacher self-efficacy stands valid, it gives an "external" argument for the causal influence of the collective over the individual. In other words, if the direct effects of the exogenous variables (instructional leadership and academic heterogeneity) had been positive and not negative, statistically significant or higher than the effect sizes over collective efficacy, the causation would be proven false. The external argument, however, is of empirical nature, and could be misleading if no conceptual framework supported it. In fact, causal rationales have been offered (Goddard & Goddard, 2001; Schmitz & Schwarzer, 2002) explaining how a group-shared belief sets a pragmatic pressure over the individuals, who will in time tend to adjust and benefit from it.

This model includes a notion that leaders, exactly as the literature seems to support (López & Gallegos, 2014; Leithwood & Jantzi, 2008), influence the shared belief, provide professional support to the extent to make the belief true and through this affect the personal beliefs of teachers. The strategy to work as a leader, with groups beliefs is rooted in the very foundations of leadership theory (Hersey & Blanchard, 1969) and helps understand that leaders will not necessarily influence people one by one (through interviews, performance appreciation, or other strategies) but by

having a grasp over the group's peer pressure. On the other hand, contextual conditions, as operationalized in this research by students' heterogeneity, have the same low effects on individuals, suggesting that each teacher has personal reactions to the given student configuration (and that it is, as some literature stated, negative), but also has low incidence over the shared belief. This implies, for educational leaders, that despite having an argument to think that a more challenging heterogeneity in schools activates efficacy in teachers, can rely on their ways to set visions, climate and professional discussions are far more influential than these dispersion of students' capacities. More specifically, the practices of instructional leaders' covariate much more than the covariation of students' capacity diversity, with an optimistic culture in a school.

4.2.5. Indirect influence paths of leaders over teachers

As discussed, the direct effect of instructional leadership over teachers' self-efficacy is first weak, second not significant, and third negative. The particular interpretation of the effect sizes was discussed before and will not be explained again here. However, the indirect influence that can be calculated from this model corresponds to considerable sizes for instructional leadership ($\approx$0.43) and academic heterogeneity ($\approx$0.12). This kind of indirect causal effect measurements have been major findings in many educational research problems (Volante, 2010; Hallinger & Heck, 1996) and help to understand the complexity of school organizations. From this can be understood that collective efficacy seems to be a crucial component for school administrators to take into account. For instance, educational leadership models might benefit from narrowing down their recommendations about having a school vision or plan (Hallinger & Murphy, 1985; Leithwood, Seashore, Louis, Anderson, & Wahlstrom, 2004; Robinson & Timperley, 2007) to specific beliefs of success for their teachers. As efficacy beliefs studies have been developing for

almost 40 years, this narrowing down a construction of a socially shared belief can be tackled through many strategies that will be considered in depth in the next chapter.

4.2.6. Analysis of the integrated model

Nevertheless, this effect size differences gives the solid argument that was searched for with all the considerations that had been made before.

First, to compare a sound and well documented influencing leadership model, with a student specific contextual factor at classroom level that could be influencing teachers seems to combine a 'safe bet' analysis with an exploratory approach for a developing field. This almost caricature initial difference was explored thoroughly enough.

Second, to choose similar schools inside the system that can be considered to operate under comparable conditions is a sound approach to equalize structurally intervenient variables. This works in order to maximize individuals' variances and hope for a deeper understanding of leadership dynamics influencing efficacy beliefs.

4.2.7. Limitations

Some crucial aspects of this research need to be pointed out to avoid misguided interpretations of the findings. To start, the influence of less typical contextual factors (namely academic heterogeneity) should not lead to overstate the principals' capabilities in schools. Many contextual boundaries were plainly omitted by design and therefore this model is not likely to stand representatively all around the

educational system. Moderator and control variable analysis helps to underpin these findings, but as good as they can be, are only an extrapolation of the results to other possible settings. As such, extrapolation projects result into uncertain ranges that might or might not prove to be true.

The inclusion of different schooling levels to get a 'bigger sample` did not give the impression of being wrong, but might bias (by lower sample sizes of both groups or 'natural' tendencies of any of the levels) the present results. As far as it is known now, primary or secondary school principals influence the teachers without significant differences. However, future research, especially if designed to explore this topic, might give different evidence.

Despite the fact that a link between teachers' efficacy and their performance has been established, this study does not include by itself an observation of actual performances. Thus, the relevance can be discussed and the model as valid predictor of better teaching in schools must be considered under this limitation.

Finally, the reached sample size with an associated sampling error below 5%, works without limitations for the individual level as the design supposed. However, as stated before, for an interpretation of this model for school level conclusions serious limitations must be considered. First, the sample size gives a 'bigger than recommended error' (18.6%). This means that almost statistically one of five repetitions of this study will yield false results and the chances for this attempt being the false one are higher than recommended. Second, an ecological fallacy cannot be discarded, which means that the covariances that worked with good fit indexes between empirical data and theoretical design, might not work well in aggregated observations. Part of this problem is solved by a 'fixed effects control', but the best way to solve this pending uncertainty would be a design that includes hierarchical modeling. As long as this is not attempted, transference of individual level conclusions to school level conclusions should not be simply made.

If these considerations are kept in mind, this study can claim that good leaders are more influential to have good teaching in schools than having similar groups of students in them.

4.3. Objectives achievement

This last part revises the achievement of the objectives providing more specific and partially methodological discussions.

4.3.1. Specific objective 1 achievement: Instrument Validity

The specific objective "Translate, validate, adapt, and apply the used instruments" was addressed in the instrument descriptions confirming that every used instrument had a sound and valid application in Chile beforehand. This is one of the recommended requirements in the SEM methodology as the measurement error is assumed to be zero (Kline, 2011). If the instruments show normal distribution problems or cultural bias, an unperceived error is introduced in the empirical dataset that can distort the adjustment values.

Not directly an instrument validation, but related to having sound variables, it is worth of notice as depicted in the descriptive statistics that academic heterogeneity had itself no psychometric properties but operated in the same fashion as the population values. It was also the only variable were such an analysis was possible. The risk of introducing distortion through this inclusion is low. However, it is recommended that alternative conceptualizations for heterogeneity like students' marks (e.g. Scharenberg, 2012) or school's internal regression coefficients (e.g. Treviño, et. al., 2016) should be tried out, in order to discard or to find other observations.

All in all, the objective can be considered achieved, despite the item trim for the collective efficacy scale without post hoc analysis would have shown less optimal fits. It was explained that no previous assumptions of this 'noise' were evident, so the decision to persist with the original design is defendable.

Including the scales CFA's structures into the measurement model as a second order design means that particular low fit contributes to lower the general model's fit. Linking path analysis with factor analysis carries a consequence explained by David Kaplan as:

> "*These added restrictions [of a more complex model] make it all more likely that a reasonably well fitting structural part of the model will be rejected due to problems within the measurement model. Moreover, the potential for misspecification in the measurement part of the model owing to these restrictions can, in some circumstances, propagate into the structural part of the model (Kaplan, 1988; Kaplan & Wenger, 1993).*" (Kaplan, 2008, p. 61).

The adequate fit of the general model without re-specifications and modifications is an achievement by itself. The remaining recommendation would consist in gathering more evidence about the functioning of the collective efficacy e scale in national settings (for example differentiation rural from urban) and/or developing alternative measurement models and corresponding scales.

4.3.2. Specific objective 2 achievement: Matching sample

This second objective, 'Achieve a sample that matches the studies requirements', required attentions to some specific issues. The discussion about statistical

significance of findings was given in the samples description and can be considered settled in a satisfactory way, assuming that recommended statistical significances are part but not determinant of relevant knowledge construction (Schmidt, 1996). The recommendations for sample size when performing SEM analysis have been addressed in the samples description as well, noting that there are no clear directions about it. For instance, the revision of 41 reported studies showed that only five yielded the conventionally non-significant chi-square scores that should be held accountable for adequate sample sizes, with participant numbers equal to 70, 165, 193, 330, and 461 (McDonald & Ho, 2002). This is consistent with recommendations to not use sample size adequacy indexes like chi-square too strictly (Schermelleh-Engel, et. al., 2003). In any case, the yielded sample of this work can be considered among the higher standards comparing sample size in other SEM studies (McDonald & Ho, 2002). This tends to end with a pragmatic recommendation on the lines of "if it worked, it was enough". Considering this a final criterion, the sample size also evidenced enough to estimate the model.

In any case, the sample error in this research allows estimating correctly (with 4.7% error even after omitting missing cases and inconsistent answering of the scales) the teachers that work in public schools in the six municipalities of the Araucanía region.

4.3.3. Specific Objective 3 achievement: Academic heterogeneity as school context

In the section regarding variables selection, the explanation about academic heterogeneity addressed thoroughly the choice for the objective: 'Discuss and determine how the variable school context is defined for this specific sample and measured in this study'. Considering the structural characteristics equalization in the design, trying to take distance from socially explained contextual

characteristics, and a sound research body under development were the main found reasons for the choice. The operationalization has proven not to be wrong, but considering the low effect sizes in the model in relation with the other variables it does requires further analysis. From the correlation results, such low and even negative effects could not have been foreseen and only make relevant the trial and error approach that science uses in order to accumulate knowledge.

New antecedents have been provided for three research fields, namely educational leadership, heterogeneity and segregation studies, and school efficacy beliefs. As such, the relevance of the study remains undoubted, only less visible than big-effect-size kind of findings.

4.3.4. Specific Objective 4 achievement

In order to check if the task 'Specify and overcome the structural equation modeling limitation of sample size, error estimation, reliability, and magnitude prescriptions of the specified correlations' was well met, the presented results during the whole research speak for themselves. In fact, the findings are hopefully both orienting for knowledge construction, and considered with the limitations treated in the last part of the discussion.

There is no evidence that a methodological flaw was not overcome or at least correctly discussed as some reliability indicators were not sufficient and the magnitude prescriptions were not as sound as hoped for. The results yielded a useful model that is of worth for practitioners and scholars.

4.3.5. Specific objective 5 achievement

This specific objective, 'collect and analyze data for the confirmatory analysis' is considered through all the work with some particular specifications in the procedure and analysis plan. The fulfillment of the objective however can only be considered if the work produced the model. As this was discussed thoroughly in the first part of the discussions, it seems redundant to explain how it was achieved.

However, some considerations on the quality of the procedure that this objective implies will be summarized again or expanded here. The data collection was performed under consistent standards, from the contact to the application. This was assured by having every school visited by the same researcher, having the same conversations with the participants and applying the questionnaires in the same way. The high response rate with low missing data are evidence of a sound collection. Only two people did not sign their informed consent, but still answered the questionnaire, raising questions about forgetting to sign, rather than a reason to indispose their participation.

The double data entry eliminated 1.5% of typing errors that compared with studies for procedures in clinics is somewhat higher, ranging from 0.54% and 0.72% (Büchele, Och, Bolte, & Weiland, 2005) or 0.19% (Reynolds-Haertle & McBride, 1992). The digitations of clinical information instead of questionnaires answering imply different stakes so could be performed with other amounts of attention. Moreover, it is different to entry 100 data of one case after another than a handful of results during an attention. These differences can help explain the difference, but statistically speaking can alter the results without even noticing it as a risk in the discussion in social sciences. It is important that researchers comply with quality standards and that those are expressly defined and discussed.

However, it would be strange to start reporting on double data entry when the relevant issues are others, but somehow quantitative researchers must dimension the extent of the problems that lack of attention can carry in their findings. Such an understanding is only possible with a growing body of available data.

The next step implies performing the analysis under the best possible quality standards. As this is highly dependent on the used software, the right way to address this would be to offer a comparison between the alternatives (Narayanan, 2012). The comparison indicated that result estimations among programs are undifferentiated, but accept the limitations in different fashions. Other differences are accessibility (e.g. licenses or open code), output formats and graphic interface but statistically unsubstantial. Therefore, the used R package LAVAAN can be considered as good as any other alternative.

V. Conclusions

The study has approached the main objective, empirical confirmation of the theoretically designed model. The model's structure summarizes a couple of findings that were separately available in different research areas and works. This allowed a relevant analysis for three fields: instructional leadership -and its relations with efficacy beliefs-, efficacy beliefs –and the influence paths off the collective over the individuals-, and segregation policies -academic heterogeneity and the possible influence over school-relevant variables. The results confirmed the hypothesis with good fit indexes, allowing establishing the following conclusions.

Teachers in the sample have their self-efficacy beliefs influenced by those of their peers about the educational task and their own competence. In other words, if a teacher is in a challenging group that feels they can go an extra mile in order to get their students to learn better, he or she will be more willing to actually do so. This can be explained as described in the theory.

A rationale for this teacher belief, is to understand social pressure of the group as a positive engine to spur high expectation beliefs about his or her own practice. This pressure can range from a positive influence -to feel part of the group, rising to meet the implicit quality standards-, or from a less positive fear of being rejected because of underperformance and less peer validation. In any case that might be elucidated in the future, it is a social shared belief that shapes personal disposition towards teaching.

A second –and not exclusive- possibility is that the covariance between both levels (collective and individual) is explained by an unmeasured factor influencing both equally. For instances what Bandura (1998) denominated previous mastery and vicarious experience constitute a shared past success experiences could be one option. To have to been able to succeed and learn together from that success is an interrelated experience for teachers. This means that a school with previous

pedagogical merits will gain a common cognitive label of being effective, increasing the expectations and beliefs of the individuals regarding their groups.

A derivation of this unmeasured factor could also be another shared factor like a social desirability bias of answering any question in the same fashion. However, this would be considered a measurement error and was addressed by the sound psychometric properties of the applications.

Another aspect that this work helps to understand is that teachers do not change their self-efficacy in significant ways because of the leadership practices of the principal. A reported low effect (-0.065), although slightly negative, means that teachers might even feel less competent if they perceive that their leaders engage more frequently in instructional leadership tasks. The fact that teachers who think better about their principals in instructional terms seem to show no changes in their self-efficacy, might mean that more competence of the principal does not activate competence on the teachers' side.

School leaders might notice that a direct influence on the individual could be limited, probably with many variations from case to case, and find more useful addressing specific issues than trying to foster general confidence in the teacher.

The academic heterogeneity of the schools' students also apparently has no relation at all with the teachers' efficacy belief. That means that teachers develop their own perception about future success by other factors than the apparent difficulty of teaching easier and more difficult students at the same time. This helps to understand that there might be no influence and why it has not been reported in previous literature. In fact, Bandura (1998) does not even consider the real characteristics of the work as part of the factors (to remember as previous mastery, vicarious experience, social persuasion and emotional states) that explain efficacy changes. The four factors only consider psychologically processed experiences. For teachers, the previous success experience that was mentioned before, is more

logically related to the success of the whole school or its students, so their variation within the group is kept unaltered. Moreover, the specific choice of the variables' operationalization reduced the individuals' variance inside schools, and little is known now about how teachers could relate to their class heterogeneity and their different subjects or teaching levels. Interesting research options open under this general finding of no association with the internal schools' heterogeneity that under more specific exploration designs could yield different results.

Another thing happens, regarding the opinion of teachers about their group's efficacy beliefs. It seems clear that principals who engage more often in instructional leadership tasks are reported together with groups that are considered to have higher efficacy beliefs. The strength and conceptual consistency of this finding robustly allows thinking of principals as the leading voice on the groups' perception about themselves. Of course, an actual hypothesis can be rejected in future research but as long as no counterevidence is presented this rationale can be considered as valid. It is almost an axiom that leaders are defined as being able to influence teachers' mindframes (Leithwood, et al., 2010) about their capacities, expectations, difficulties and required efforts. So a high relation (0.7) is not a surprise. Other alternatives explaining this covariation through common influencing factors or other inverted causalities can be interesting as scholar issues, but don't seem necessary for the present claims.

It is worth noticing that, while leaders could find less effects in attempting to enhance teacher's self-efficacy individually, they would be highly effective in defining a shared vision. This vision differs somewhat about the strategic goal setting and planning that are recommended in the literature (e.g. Robinson & Timperley, 2007) and standard setting (e.g. MINEDUC, 2015) but is not completely unrelated. While the strategic characteristic is brought into the foreground, the constructed vision of collective efficacy is not only rational but psychologically complex. As such it can be understood from the social cognition perspective.

Teacher would gain cognitive schemas through principals' actions, almost like "*cognitive scripts*" that "*are reliable knowledge structures from which to set goals and anticipate the future precisely because they are consensually based and socially prescriptive*" (Augoustinos, Walker, & Donaghue, 2014, p. 97). If collective efficacy under Roger Goddard and Yvonne Goddard (2001) conceptualization is as a very specific cognitive script for teacher work, the effects of Philip Hallinger's (2013) conceptualization of leaders' work becomes evident. The cognitive script includes the planning process but as an implicit dynamic of social consensus and social prescription that is broader. These are good news for educational leaders that can reframe the recommendations about defining a school-shared vision from a relational and group dynamic point of view and not a narrow rational plan.

Moreover, if leaders are able to promote collective efficacy scripts in the teacher team, they have an indirect way to promote efficacy beliefs in the same direction for the individuals. As discussed earlier, this indirect path of influence, with convergent conceptualizations and empirical confirmation in the model, explains how principals can increase teachers' expectations about their work and put them in the best disposition to work better. This is also true for the opposite, of unengaged leaders missing the chance to organize the collective and having less opportunities to engage teachers. As this second case seems to still have a positive effect on teachers because of the negative effect size and the compensatory hypothesis, a deeper exploration of the relations should be attempted. Some guidelines for it would certainly be the control by schools' success, what happens at specific leadership intervals, and how nested organization of the teachers in groups predicts the outcome variances. Appropriate methods for such an attempt would be hierarchical linear modeling (HLM), or a variant that mixes it with SEM defined as Multilevel Structural Equation Modeling (MSEM) (Kaplan, 2008).

Finally, the collective efficacy was scored higher by the teachers in schools that had a greater dispersion of their students' performance. Despite that the finding is not reliable, the positive effect –however small- gives a clue about what might happen to teachers if broad structural characteristics are equalized. If heterogeneous groups are taught by better disposed teacher teams, claim that can only be made only with severe limitations, then an argument for educational systems design and policies is gained. This argument is not a recommendation because lots of verification needs still to be done, but constitutes an issue of scholarly interest.

Altogether, the main objective of assessing empirical validity for the specified model of causal relations between educational leadership and school's contexts over efficacy beliefs in a sample of teachers in challenged schools can be considered achieved. The findings also gave interesting insights for future research and helped, as intended, to consolidate in one analysis a body of knowledge that was dispersed in the literature.

The unsolved questions are relevant for the national Chilean setting of standards based reform. As long as policy designers introduce 'good practices' tested in exceptional contexts into the whole system holding accountable different actors to implement them, schools will be tensed for action that cannot be or should be performed. This is particularly so in challenged contexts, and does not follow the recommendations for contextualised school improvement (Carrasco, 2008).

5.1. Implications

The implications of a work like this are manifold. But central is the statement about how instructional leadership is or becomes a relevant issue in the studied contexts. The fact that this thesis used what was called a contextual equalization is not only a strategy to deepen in the phenomena itself but teaches something about the specific setting. In this case, poverty, high rurality and Mapuche composition, define the setting and that instructional leadership is also important for these schools can be considered relevant knowledge. As such it serves the SER agenda to produce specific knowledge in the challenged contexts (Carrasco, 2008). While, researchers have stated how leadership practices need to adjust and operate quite differently from one context to another, this study is confirmatory for the known theory about school leadership. Leadership practices seems to operate here under analogous principles as anywhere else, constructing a social cognition or shared vision about what the organization is and what the teachers do. The way a leader can shape this construction with optimism and trust in the professional teams' capacities is strategic. As such it serves the instructional leadership research field (Reynolds, Teddlie, Hopkins, & Stringfield, 2000).

Arguably a confirmatory work like this could mean lack of the problems relevance. However, this seems not to be the case, as it was the first attempt to describe an influence chain with methods (SEM) that adjust particularly fine to this. In doing so, three separated research fields (namely educational leadership, efficacy beliefs, and academic heterogeneity) are linked and finding confirmation for what was already known is in fact more reassuring than unexpected findings. The scientific accumulation of knowledge requires stable empirical observation in order to ground theory and validate methods for these observations. Therefore, this work has established a consistent relations of variables overcoming knowledge fragmentation and provided an understanding of the phenomenon.

5.3. Recommendations for practice

The final words intend to summarize a message of what has been learned that becomes helpful for principal's daily practice. The popular quote "where there's a will, there's a way" synthetizes the main findings of this thesis. The message is that leaders need to understand that direct approaches to school efficacy might not be possible, but persistence and a clear high expectation is useful to achieve it.

First of all, the importance of teacher self-efficacy must be clear. Teaching involves a lot of decision making about the students, the curriculum and the own actions. Teachers who have higher expectations about what to achieve with their decisions are more prone to evaluate, modify and persist on more effective decisions. The way how teachers feel and think about themselves, first does not depend on the students' differences in the school, as can be usually thought. Of course, it does depend on the school (this thesis avoided that issue equalizing structural characteristics), but inside the school students could be academically more different or more the same, teachers seem to have similar expectations about them.

But it does depend highly on the kind of faculty they are inserted in. A high expectation colleague team, that shares a common standard for teaching and learning is important for teachers to believe in themselves. This constitutes a 'operative culture' that again does not depend on the students' composition if compared to what kind of principal are guiding it.

Leaders cannot miss the chance to mark that 'operative culture' with their own seal. Constructing a common narrative with the teachers seems to be the right path to make them believe the additional effort is worth. This is a socio-cognitive 'script'; practical, guiding and motivating. It would seem this reinforces by itself through peer pressure, what leaders try to make happen with great personal effort, but in fact it is one of the main tasks and tools for leaders. Rational vision setting is good,

but not enough, as the consensual 'script' is of a more complex and psychological nature. Being able to shape and make explicit the script for teachers is a daily and subtle task, but once grasped becomes smooth and strong.

Principals should avoid to disregard face to face contact with the teachers, because individual influence seems to be low. Such contact is crucial to reinforce the collective 'script'. Direct conversation is meaningful for the individual to gain insight about what the leaders are trying to build collectively.

Finally, leaders need to believe in their capability to change their schools for the better. The expectations that can be transmitted by the school leaders are real motivators of the whole efficacy process.

5. References

Al-Mahdy, Y. F. H., Emam, M. M., & Hallinger, P. (2018). Assessing the contribution of principal instructional leadership and collective teacher efficacy to teacher commitment in Oman. *Teaching and Teacher Education*, 69, 191-201.

Arancibia, V. (2011). *Plan Formación de Directores de Excelencia.* Paper presented at the Seminario internacional plan de formación de directores de excelencia, Santiago.

Augoustinos, M., Walker, I., & Donaghue, N. (2014). *Social cognition: An integrated introduction*: Sage.

Bandura, A. (1993). Perceived self-efficacy in cognitive-development and functioning. *Educational Psychologist, 28*(2), 117-148. doi: 10.1207/s15326985ep2802_3

Bandura, A. (1998). Personal and collective efficacy in human adaptation and change. *Advances in psychological science, 1*, 51-71.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1), 1-26.

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents*, 5(307-337).

Bellei, C. (2013). El estudio de la segregación socioeconómica y académica de la educación chilena. *Estudios pedagógicos (Valdivia), 39*(1), 325-345.

Bravo, J. (2011). SIMCE: Pasado, presente y futuro del Sistema Nacional de Evaluación. *Estudios Públicos, 123*, 189-211.

Bridges, E. M. (1967). Instructional leadership: A concept re-examined. *Journal of Educational Administration*, 5(2), 136-147.

Büchele, G., Och, B., Bolte, G., & Weiland, S. K. (2005). Single vs. double data entry. *Epidemiology, 16*(1), 130-131.

Burris, C. C., & Garrity, D. T. (2008). *Detracking for excellence and equity*: ASCD.

Calik, T., Sezgin, F., Kavgaci, H., & Kilinc, A. C. (2012). Examination of Relationships between Instructional Leadership of School Principals and Self-Efficacy of Teachers and Collective Teacher Efficacy. *Kuram Ve Uygulamada Egitim Bilimleri, 12*(4), 2498-2504.

Carrasco, A. (2008). Investigación en efectividad y mejora escolar: ¿nueva agenda? REICE. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 6(4).

Carrasco, A., & Fromm, G. (2016). How do local markets pressures shape Leadership practices: evidence from Chile. *Journal of Educational Administration and History, 48*(4), 209-308.

Cohen, E. G. (1991). Teaching in multiculturally heterogeneous classrooms: Findings from a model program. *McGill Journal of Education/Revue des sciences de l'éducation de McGill, 26*(001).

Concato, J., & Hartigan, J. A. (2016). P values: from suggestion to superstition. *Journal of Investigative Medicine, 64*(7), 1166-1171.

Covarrubias, C., & Mendoza, M. (2013). La teoría de autoeficacia y el desempeño docente: El caso de Chile. *Estudios hemisféricos y polares, 4*(2), 107-123.

Covarrubias, C. G. (2014). *El sentimiento de autoeficacia en una muestra de profesores chilenos.* Universidad Complutense de Madrid, Spain.

Covarrubias, C. G., & Mendoza, M. (2015). Sentimiento de autoeficacia en una muestra de profesores chilenos desde las perspectivas de género y experiencia. *Estudios pedagógicos (Valdivia), 41*(1), 63-78.

Donoso, S., Benavides, N., Cancino, V., Castro, M., & López, L. (2012). Análisis crítico de las políticas de formación de directivos escolares en Chile: 1980-2010. *Revista Brasileira de Educação, 17*, 133-158.

Dumay, X., & Galand, B. (2012). The multilevel impact of transformational leadership on teacher commitment: cognitive and motivational pathways. *British Educational Research Journal, 38*(5), 703-729. doi: 10.1080/01411926.2011.577889

Dupriez, V. (2010). *Methods of grouping learners at school*. Paris: UNESCO.

Dussault, M., Payette, D., & Leroux, M. (2008). Principals' transformational leadership and teachers' collective efficacy. *Psychological Reports, 102*(02), 401-410. doi: 10.2466/pr0.102.2.401-410

Eidhof, B. (2016). *Influencing youth citizenship*: Research Institute Child Development and Education (CDE).

Fromm, G., Hallinger, P., Volante, P., & Wang, W.-C. (2017). Validating a Spanish Version of the PIMRS Application in National and Cross-national Research on Instructional Leadership. *Educational Management Administration & Leadership*, 1-26. doi: 1741143215617948

Fromm, G., Volante, P., & Olbrich, I. (2015). "Fidelidad de la implementación de prácticas de liderazgo instruccional", *Revista Magis*, 15 (7)

Gamoran, A., Nystrand, M., Berends, M., & LePore, P. C. (1995). An organizational analysis of the effects of ability grouping. *American Educational Research Journal, 32*(4), 687-715.

García, G. G., Peña, I. S., & Parra, C. S. (2016). Desafíos investigativos en la formación inicial de directivos. *Revista Iberoamericana de educación*(70), 43-54.

Goddard, R., Sweetland, S. R., & Hoy, W. K. (2000). Academic emphasis of urban elementary schools and student achievement in reading and mathematics: A multilevel analysis. *Educational Administration Quarterly, 36*(5), 683-702.

Goddard, R. D. (2001). Collective efficacy: A neglected construct in the study of schools and student achievement. *Journal of Educational Psychology, 93*(3), 467-476. doi: 10.1037/0022-0663.93.3.467

Goddard, R. D., & Goddard, Y. L. (2001). A multilevel analysis of the relationship between teacher and collective efficacy in urban schools. *Teaching and Teacher Education, 17*(7), 807-818. doi: 10.1016/s0742-051x(01)00032-4

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*(2), 479-507. doi: 10.2307/1163531

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational researcher, 33*(3), 3-13.

Goddard, R. D., & LoGerfo, L. F. (2007). Measuring emergent organizational properties - A structural equation modeling test of self- versus group-referent perceptions. *Educational and Psychological Measurement, 67*(5), 845-858. doi: 10.1177/0013164406299103

Goddard, R. D., & Skrla, L. (2006). The influence of school social composition on teachers' collective efficacy beliefs. *Educational Administration Quarterly, 42*(2), 216-235. doi: 10.1177/0013161x05285984

Gottfried, M. A. (2015). Does classmate ability influence students' social skills? *School Effectiveness and School Improvement, 26*(4), 554-585.

Hallinger, P. (2011). A Review of Three Decades of Doctoral Studies Using the Principal Instructional Management Rating Scale: A Lens on Methodological Progress in Educational Leadership. *Educational Administration Quarterly, 47*(2), 271-306. doi: 10.1177/0013161x10383412

Hallinger, P. (2013). *Measurement Properties of the Principal Instructional Management Rating Scale* (5.0 ed., pp. 88): General Research Fund # GRF 841711 - University Grants Commission of Hong Kong.

Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly, 32*(1), 5-44. doi: 10.1177/0013161x96032001002

Hallinger, P., Hosseingholizadeh, R., Hashemi, N., & Kouhsari, M. (2017). Do beliefs make a difference? Exploring how principal self-efficacy and instructional leadership impact teacher efficacy and commitment in Iran. *Educational Management Administration & Leadership*, 1741143217700283.

Hallinger, P., & Leithwood, K. (1994). INTRODUCTION - EXPLORING THE IMPACT OF PRINCIPAL LEADERSHIP. *School Effectiveness and School Improvement, 5*(3), 206-218. doi: 10.1080/0924345940050301

Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *The Elementary School Journal*, 217-247.

Hattie, J. A. (2002). Classroom composition and peer effects. *International Journal of Educational Research, 37*(5), 449-481.

Hersey, P., & Blanchard, K. H. (1969). Life cycle theory of leadership. *Training & Development Journal*.

Hopkins, D. (2001). *School Improvement for Real*. London: RoutledgeFalmer.

Hoy, A. W., Hoy, W. K., & Kurz, N. M. (2008). Teacher's academic optimism: The development and test of a new construct. *Teaching and Teacher Education, 24*(4), 821-835. doi: 10.1016/j.tate.2007.08.004

Jung, D. I., & Sosik, J. J. (2002). Transformational leadership in work groups the role of empowerment, cohesiveness, and collective-efficacy on perceived group performance. *Small group research, 33*(3), 313-336.

Kaplan, D. (2008). *Structural equation modeling: Foundations and extensions* (Vol. 10): Sage Publications.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*: Guilford press.

Knoblauch, D., & Hoy, A. W. (2008). "Maybe I can teach those kids." The influence of contextual factors on student teachers' efficacy beliefs. *Teaching and Teacher Education, 24*(1), 166-179. doi: 10.1016/j.tate.2007.05.005

Kulik, J., & Kulik, C. (1987). Effects of ability grouping on student achievement. *Equity and Excellence in Education, 23*(1-2), 22-30.

Kulik, J. A. (1992). An Analysis of the Research on Ability Grouping: Historical and Contemporary Perspectives. . *Research-Based Decision Making Series*.

Lei, P.-W., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: issues and practice, 26*(3), 33-43.

Leithwood, K. A. (1992). The move toward Transformational Leadership. *Educational Leadership*, 49(5), 8-12.

Leithwood, K., & Jantzi, D. (1999). The relative effects of principal and teacher sources of leadership on student engagement with school. *Educational Administration Quarterly, 35*, 679-706. doi: 10.1177/00131619921968798

Leithwood, K., & Jantzi, D. (2008). Linking leadership to student learning: The contributions of leader efficacy. *Educational Administration Quarterly, 44*(4), 496-528. doi: 10.1177/0013161x08321501

Leithwood, K., & Mascall, B. (2008). Collective leadership effects on student achievement. *Educational Administration Quarterly, 44*(4), 529-561. doi: 10.1177/0013161x08321221

Leithwood, K., Patten, S., & Jantzi, D. (2010). Testing a Conception of How School Leadership Influences Student Learning. *Educational Administration Quarterly, 46*(5), 671-706. doi: 10.1177/0013161x10377347

Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How Leadership Influences Student Learning*. The Wallace Foundation. University of Minnesota, Center for Applied Research and Educational Improvement. Ontario Institute for Studies in Education. University of Toronto, Ontario Institute for Studies in Education.

Ley N° 20845. Diario Oficial de la República de Chile, Santiago, Chile, 8 de junio de 2015.

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.

López, P., & Gallegos, V. (2014). Prácticas de liderazgo y el rol mediador de la eficacia colectiva en la satisfacción laboral de los docentes. *Estudios pedagógicos, 40*(1), 163-178.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423-458.

Lubienski, S. T., & Lubienski, C. (2005). A new look at public and private schools: Student background and mathematics achievement. *Phi Delta Kappan, 86*(9), 696.

MacBeath, J., Gray, J., Cullen, J., Frost, D., Steward, S. y Swaffield, S. (2007). *Schools on the Edge: Responding to Challenging Circumstances*. London: Paul Chapman.

Martinic, S., Elacqua, G. (2010) *¿Fin de un ciclo? Cambios en la gobernanza del sistema educativo*. Santiago, Facultad de Educación PUC – OREALC UNESCO.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological methods, 7*(1), 64.

McEwan, P. J. (2001). The effectiveness of public, catholic, and non-religious private schools in Chile's voucher system. *Education economics, 9*(2), 103-128.

McEwan, P. J. (2003). Peer effects on student achievement: Evidence from Chile. *Economics of education review*, 22(2), 131-141.

MINEDUC. (2013). *Programa de Educación Básica Rural*. Santiago: MINEDUC.

MINEDUC. (2015). *Marco para la buena dirección y el liderazgo escolar*. Chile: Ministerio de Educación

MINEDUC. (2016). *Política de Fortalecimiento del Liderazgo Directivo Escolar*. Santiago: MINEDUC.

Mizala, A., & Torche, F. (2012). Bringing the schools back in: the stratification of educational achievement in the Chilean voucher system. *International Journal of Educational Development, 32*(1), 132-144.

Mojavezi, A., & Tamiz, M. P. (2012). The impact of teacher self-efficacy on the students' motivation and achievement. *Theory and Practice in Language Studies*, 2(3), 483.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika*.

Narayanan, A. (2012). A review of eight software packages for structural equation modeling. *The American Statistician, 66*(2), 129-138.

Nomi, T., & Raudenbush, S. W. (2016). Making a Success of "Algebra for All" The Impact of Extended Instructional Time and Classroom Peer Skill in Chicago. *Educational Evaluation and Policy Analysis, 38*(2), 431-451.

Persell, C. H. (1977). *Education and inequality: A theoretical and empirical synthesis*. New York: Free Press.

Raudenbush, S., Rowan, B., & Cheong, Y. F. (1992). Contextual Effects on the Self-perceived Efficacy of High School Teachers. *Sociology of Education, 65*(2), 150-167.

Resh, N., & Dar, Y. (2012). The rise and fall of school integration in Israel: Research and policy analysis. *British Educational Research Journal, 38*(6), 929-951.

Reynolds-Haertle, R. A., & McBride, R. (1992). Single vs. double data entry in CAST. *Controlled clinical trials, 13*(6), 487-494.

Reynolds, D., Teddlie, C., Hopkins, D., & Stringfield, S. (2000). Linking school effectiveness and school improvement. *The international handbook of school effectiveness research*, 206-231.

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types. *Educational Administration Quarterly, 44*(5), 635-674. doi: 10.1177/0013161x08321509

Robinson, V. M. J., & Timperley, H. S. (2007). The leadership of the improvement of teaching and learning: Lessons from initiatives with positive outcomes for students. *Australian Journal of Education, 51*(3), 247-262.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.

Rui, N. (2009). Four decades of research on the effects of detracking reform: Where do we stand?—A systematic review of the evidence. *Journal of Evidence-Based Medicine, 2*(3), 164-183.

Scharenberg, K. (2012). Do secondary students learn more in homogeneous or heterogeneous classes? The importance of classroom composition for the development of Reading achievement in secondary school. *Online Educational Research Journal, 3*(12), 1-11.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online, 8*(2), 23-74.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological methods, 1*(2), 115.

Schmitz, G. S., & Schwarzer, R. (2002). Individual and collective self-efficacy expectation of teachers. *Zeitschrift Fur Padagogik*, 192-214.

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.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57*(3), 293-336.

Smyth, J. (1989). Developing and sustaining critical reflection in teacher education. *Journal of teacher education*, 40(2), 2-9.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: a distributed perspective. *Journal of Curriculum Studies*, 36(1), 3-34. doi: 10.1080/0022027032000106726

Tanaka, J. S. (1987). " How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child development*, 134-146.

Taut, S., & Escobar, J. (2012). *El efecto de las características de los pares en el aprendizaje de estudiantes chilenos de enseñanza media*. TECHNICAL REPORTS MIDEUC. Measurement Center MIDEUC, Santiago.

Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research, 37*(5), 483-504.

Treviño, E., Valenzuela, J. P., & Villalobos, C. (2014). Segregación académica y socioeconómica al interior de la escuela. Análisis de su magnitud, evolución y principales factores explicativos. *Santiago: CPCE-CIAE*.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805.

Valenzuela, J. P., Bellei, C., & Allende, C. (2016). Measuring systematic long-term trajectories of school effectiveness improvement. *School Effectiveness and School Improvement*, 1-19.

Vandenberghe, V. (2002). Evaluating the magnitude and the stakes of peer effects analysing science and math achievement across OECD. *Applied Economics*, 34(10), 1283-1290.

Vázquez, E. (2015). *Análisis del director escolar en el marco jurídico fundamentado en una reforma basada en estándares; una aproximación desde el movimiento de la Mejora Escolar*. Unpublished Master Dissertation.

Venkatakrishnan, H., & Wiliam, D. (2003). Tracking and Mixed-ability Grouping in Secondary School Mathematics Classrooms: a case study 1. *British Educational Research Journal, 29*(2), 189-204.

Volante, P. (2010). Influencia instruccional de la dirección escolar en los logros académicos. Unpublished PHD Dissertation.

Volante, P. (2012). Liderazgo Instruccional y Logro Académico en la Educación Secundaria en Chile. In J. Weinstein & G. Muñoz (Eds.), *¿Qué sabemos sobre los directores de escuela en Chile?* Santiago: CEPPE and Fundación Chile.

Wahlstrom, K. L., & Louis, K. S. (2008). How teachers experience principal leadership: The roles of professional community, trust, efficacy, and shared responsibility. *Educational administration quarterly*, 44(4), 458-495.

Walumbwa, F. O., Wang, P., Lawler, J. J., & Shi, K. (2004). The role of collective efficacy in the relations between transformational leadership and work outcomes. *Journal of Occupational and Organizational Psychology, 77*(4), 515-530.

Wilkinson, I. A., & Fung, I. Y. (2002). Small-group composition and peer effects. *International Journal of Educational Research, 37*(5), 425-447.

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly, 39*(3), 398-425.

6. Appendixes

Appendix 1.1: Informed Consentient

CONSENTIMIENTO INFORMADO

Usted ha sido invitado a participar en la investigación: "Influencia del Liderazgo Educacional y el Contexto Escolar sobre las Creencias de Eficacia de Profesores". Este estudio se realiza en el Laboratorio de Liderazgo e Influencia Escolar de la Facultad de Educación de la Pontificia Universidad Católica de Chile, por el estudiante doctoral Germán Rodolfo Fromm Rihm y su Director de Tesis Dr. Paulo Luis Volante Beach. El objetivo es validar con datos un modelo teórico sobre relaciones entre liderazgo escolar y contextos escolares sobre creencias de eficacia de los profesores en escuelas municipales de la IX Región.

El propósito de este documento es ayudarle a tomar una decisión informada para participar respondiendo una selección de cuestionarios sobre sus características como Docente, Liderazgo Escolar, Orientación a la Justicia Social y Creencias de Eficacia.

BENEFICIOS Y RIESGOS: Este estudio tiene el beneficio de generar conocimientos sobre liderazgo y su vinculación operativa con los contextos en los que opera. El establecimiento podrá contar con una retroalimentación de los resultados grupales (no individuales). Usted se podrá retirar de esta investigación cuando lo estime y sin dar razones que lo justifiquen. A juicio de los investigadores, su participación en este estudio no conlleva riesgos ni consecuencias para Ud., el director o la escuela. Los cuestionarios no tienen propósitos evaluativos y tampoco involucran respuestas correctas o incorrectas.

ALMACENAMIENTO DE LOS DATOS PARA LA CONFIDENCIALIDAD DEL PROYECTO: Los registros de los cuestionarios no identificarán el nombre de los participantes, ni cualquier otra información que lleve a identificarlos. Su identificación en este documento no permite establecer su identidad pues es relacionada con sus respuestas a través de un código. La información que proporcione además será confidencial y sólo manejada por los investigadores que desarrollan el proyecto. Su establecimiento será registrado a través de una clave por lo que tampoco será identificable.

LUGAR Y TIEMPO INVOLUCRADO: Los cuestionarios serán respondidos en el lugar y tiempo convenidos con usted y autorizadas previamente por la dirección del establecimiento. El tiempo necesario es de aproximadamente 45 minutos.

USO DE LOS RESULTADOS: Los resultados del estudio serán publicados en revistas y actividades de divulgación especializadas. Toda publicación se hará con propósitos académicos. No hay financiamiento asociado a la investigación o a sus resultados.
DERECHOS DE LOS PARTICIPANTES

He leído y tenido la oportunidad de hacer preguntas al investigador responsable acerca del propósito y procedimientos del estudio.

- Mi participación en esta investigación es voluntaria. Puedo negarme a participar o renunciar a participar en cualquier momento sin perjuicio para mi futuro profesional.
- Si durante el transcurso del estudio nueva información significativa llega a estar disponible y se relaciona con mi voluntad de continuar participando, el investigador deberá entregarme esta información.
- Si en algún momento tengo alguna pregunta relacionada con la investigación o con mi participación, puedo contactarme con el investigador responsable, quién responderá mis preguntas. El teléfono de contacto es **8 809 5917** y su correo electrónico es grfromm@uc.cl.
- Si en algún momento tengo comentarios o preocupaciones relacionadas con la conducción de la investigación o preguntas acerca de mis derechos al participar de este estudio, yo puedo contactarme con el Comité de Ética de la Facultad de Educación de la Pontificia Universidad Católica de Chile, a través del Subdirector de Investigación, Profesor Carlos González, al número telefónico 02 2 354 5365, o al correo cgonzalu@uc.cl.

CONSENTIMIENTO INFORMADO

Yo estoy de acuerdo en participar en el estudio titulado: *Influencia del Liderazgo Educacional y el Contexto Escolar sobre las Creencias de Eficacia de Profesores.* El propósito y naturaleza del estudio me ha sido totalmente explicado por el investigador responsable, señor **Germán Fromm**. Yo comprendo lo que se me pide. Sé que puedo contactarme con el investigador responsable o con el Comité de Ética en cualquier momento, para realizar preguntas y resolver dudas. También comprendo que puedo renunciar al estudio en cualquier momento. Además, estoy de acuerdo en que la información que entrego sea registrada de forma confidencial y con usos académicos.

Nombre del Participante (opcional):_____

Código del Participante : _____
(Entregado por el investigador en caso de preferir no usar su nombre)

Nombre del Establecimiento : _____

Nombre del Investigador Responsable: Germán Rodolfo Fromm Rihm

Firma del Investigador Responsable : _____

Appendix 1.2: Presentation Letter to DAEM

PROYECTO DE INVESTIGACIÓN

Estimado Sr. Zerené,

Jefe DAEM Temuco,

<u>presente</u>

Me dirijo a Ud. para presentarle el Proyecto de Investigación de mi Tesis Doctoral "Influencia del Liderazgo Escolar y los Contextos Escolares sobre las creencias de Eficacia de los Profesores". Esta investigación se está realizando en el marco del Laboratorio de Liderazgo Educativo de la Facultad de Educación de la Pontificia Universidad Católica de Chile dirigido por el Dr. Paulo Luis Volante Beach.

El estudio considera como nuestras escuelas son impulsadas a lograr procesos de enseñanza equitativos en calidad a través de prácticas de liderazgo y desarrollo de culturas de aprendizaje que son adaptadas a las realidades locales para ponerlas en acción. Esta investigación busca cuantificar la influencia de las prácticas de liderazgo y las condiciones contextuales para impulsar y fortalecer a los equipos docentes de las escuelas.

Para ello la IX Región es el escenario ideal de investigación, donde las características de contexto clásicas como Nivel Socio-Económico o Dependencia de las Escuelas no están en el foco central, sino lo está la ruralidad, las diferencias culturales y la historia de la comunidad escolar.

Por ello se ha intencionado una muestra que considera como universo a los profesores que trabajan en las escuelas municipales de Temuco. El proceso completo con los profesores de una escuela reunidos toma unos 30-45 minutos para obtener una impresión completa. Los datos entregados serán manejados con total confidencialidad y utilizados exclusivamente con fines académicos. Esta actividad no tiene financiamiento asociado a su realización ni presentación de sus resultados.

Como contraprestación puedo reportar los resultados grupales a cada escuela participante y al DAEM con alcances y formas de interpretarse. Además lo invitamos a revisar nuestra página web www.liderazgoescolar.uc.cl para revisar los demás proyectos que llevamos a cabo.

Agradecería a Ud. facilitar el acercamiento a las escuelas para lograr una fecha y horario de visita para la aplicación de los cuestionarios.

Sin otro particular, quisiera agradecerle su disposición, su tiempo y su posible apoyo para esta actividad. Ojalá podamos compartir con acciones la creencia de que a través de buena investigación con aportes útiles y pertinentes podamos contribuir a la mejor de nuestra educación en Chile.

Germán Rodolfo Fromm Rihm

Dr. (c) en Ciencias de la Educación PUC

Appendix 1.3: Information Letter for School Principals

Pontificia Universidad Católica de Chile
Facultad de Educación
Laboratorio de Liderazgo e Influencia Escolar

Sr. Juan Pérez
Director del Establecimiento Educuacional

Presente

 Estimado Don Juan,

Junto con saludarle, me es grato presentarme. Soy Germán Fromm, investigador de la Facultad de Educación UC y participo del Programa Avanzado en Dirección y Liderazgo escolar UC http://liderazgoescolar.uc.cl/ .

El motivo de este mail, es invitarle a participar de un estudio de seguimiento sobre Prácticas de Liderazgo Escolar e Impacto Social de nuestro programa. Este estudio requiere que los profesores a quienes Ud. dirige, respondan dos cuestionarios en un tiempo que dura menos de 30 minutos. El estudio asegura la confidencialidad de los datos, y los resultados de su colegio le serán reportados una vez concluido el estudio.

Por consiguiente, apreciaríamos mucho que Ud. nos confirme su disponibilidad a participar y nos indique un horario para ir a aplicar este cuestionario en su colegio. Si además nos puede indicar el número de profesores que respondería podemos preparar el material impreso.

Desde ya agradecemos su atención, y esperamos concretar esta actividad y seguir participando en iniciativas de formación e investigación UC.

Quedamos atentos para resolver cualquier duda y saludamos atentamente,

Germán Fromm
Dr. (c) en Ciencias de la Educación
Cel: 88095917 Mail: grfromm@gmail.com

Dr. Paulo Volante
Jefe Programa Avanzado en Dirección y Liderazgo escolar UC
Mail: pvolante@uc.cl

Appendix 2.1: SIMCE Data use permission

Solicitud de Base de Datos SIMCE

---

contacto <contacto@agenciaeducacion.cl>

6 de mayo de 2015,
10:36

Para: Germán Fromm <grfromm@uc.cl>

Estimado Investigador

Junto con saludar, hacemos llegar a ud. los links con los datos solicitados. Ud. encontrará en estas bases de datos la información de Cuestionarios que es factible entregar a investigadores. Asimismo, pedimos a ud. tener presentes las condiciones de uso que se señalan en nuestra web,

http://www.agenciaeducacion.cl/wp-content/uploads/2013/01/Condiciones-de-Uso-BD-Agencia.pdf, donde se indica que:

La Institución y el Investigador Solicitante deben comprometerse a respetar las

siguientes condiciones de uso de las BD:

- Utilizar solo para los fines que el Investigador Solicitante establece en su

Propuesta de Investigación (ver Formulario de Solicitud).

- No usar con fines de lucro.

- En publicaciones y/o documentos de trabajo, no identificar alumnos,

padres, profesores y/o directores. Por ejemplo, en ningún caso se podrá

publicar "el profesor de 4º Básico de la escuela X tiene estudios de...".

- En publicaciones y/o documentos de trabajo, no identificar

establecimientos. Por ejemplo, en ningún caso se podrá publicar "la escuela X

tiene recursos pedagógicos tales como...".

- Solo podrán tener acceso a las BD los integrantes del equipo de investigación

del Investigador Solicitante, los que deberán estar identificados en el

Formulario de Solicitud.

- De producirse cambios en el equipo de investigación, la Institución y el

Investigador Solicitante se comprometen a entregar a la Agencia el listado

actualizado de las personas del equipo que tendrán acceso a las BD. En ningún

caso, una persona que se retira del equipo de investigación podrá llevarse una

parte o la totalidad de las BD.

- Queda prohibido entregar, ceder, comunicar, transferir, o transmitir de cualquier forma, total o parcialmente, las BD a terceras personas que no pertenezcan al equipo de investigación.

Segundo básico 2012	
Cuarto básico 2012	
Segundo medio 2012	
Tercero medio (inglés) 2012	
Segundo básico 2013	
Cuarto básico 2013	
Sexto básico 2013	
Octavo básico 2013	
Segundo medio 2013	

Atte

Agencia de Calidad de la Educación

Appendix 2.2.: Permission to Use Letter

Dr. Philip Hallinger
199/43 Sukhumvit Soi 8
Bangkok, 10110, Thailand
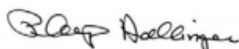hallinger@gmail.com

March 16, 2017

German Fromm

As copyright holder and publisher, you have my permission as publisher to use the *Principal Instructional Management Rating Scale* (*PIMRS*) in your research study. In using the scale, you may make unlimited copies of any of the three forms of the PIMRS.

Please note the following conditions of use:

1. This authorization extends only to the use of the PIMRS for research purposes, not for general school district use of the instrument for evaluation or staff development purposes.

2. This is a single-use purchase for the author's graduate research, thereby requiring purchase of additional rights for use in any future research.

3. *The user agrees to send a soft copy (pdf) of the completed study to the publisher upon completion of the research.*

4. *The user agrees to send a soft copy of the data set and coding instructions to the publisher upon completion of the research in order to enable further instrument development.*

5. The user has permission to make minor adaptations to scale as necessary for the research.

6. If the instrument is translated, the user will supply a copy of the translated version.

Please be advised that a separate *permission to publish* letter, if required by the university, will be sent after the publisher receives a soft copy of the completed study.

Sincerely,

Professor Philip Hallinger

www.philiphallinger.com

Appendix 3.1.: Demographics Questionnaire

Características Generales: Por favor, indique la siguiente información acerca de Ud. Y la Escuela:

(A) Nombre del Establecimiento:
_____RBD:_____

(B) Su Cargo en el Establecimiento: _____

(C) Especialización y/o título profesional:_____

(D) Edad: ☐   (E) Sexo: Femenino ☐   Masculino ☐

(F) Cantidad de años, incluyendo este, de experiencia como Profesor o Profesora: ☐

(G) ¿En cuántos establecimientos (incluyendo este) ha hecho clases? ☐

(H) Cantidad de años, incluyendo este, que lleva trabajando con el director actual: ☐

(I) Salario: Al mes gano en bruto: Menos de $400.000 ☐   Entre $400.000 y 550.000 ☐

Entre $550.000 y $700.000 ☐   Entre $700.000 y $900.000 ☐   Más de $900.000 ☐

(J) ¿Qué porcentaje de estudiantes socialmente vulnerables cree Ud. que atiende este establecimiento? ☐ %

(K) ¿Qué porcentaje de estudiantes Mapuche cree Ud. que atienden? ☐ %

(L) ¿Qué porcentaje de estudiantes cree que son discriminados por alguna condición? (Racial, minoría sexual, discapacidad, trastornos del aprendizaje, etc.) ☐ %

Cuestionarios: Los siguientes cuestionarios son para responder en aproximadamente 25 minutos. No contemplan respuestas correctas o incorrectas. Tampoco se espera de Ud. una tendencia en los resultados (altos o bajos), sino su sincera opinión y experiencia sobre los temas tratados. Sus fines son exclusivamente académicos y no tendrán consecuencias para usted, para el director o el establecimiento. Su información será tratada con total confidencialidad.

En algunos casos las respuestas pueden parecer extrañas; use su criterio para seleccionar la respuesta más apropiada para esas preguntas, SIN DEJAR NNGUNA EN BLANCO. Por favor, encierre sólo un número con un círculo en cada pregunta. Trate de responder cada pregunta.

Escala de apreciación de la gestión instruccional de directores

A continuación se le pide, por favor, responder cada afirmación según lo que mejor describa las conductas y prácticas del director de su escuela o colegio. Considere por igual lo que logre y lo que no logre observar de las acciones de su director a lo largo del año escolar. Lea cada afirmación cuidadosamente. Encierre con un círculo el número que mejor se ajusta a la conducta o práctica específica en el trabajo del director durante el año escolar.

¿En qué medida el Director…?

	CASI NUNCA				CASI SIEMPRE
1. Desarrolla un conjunto focalizado de metas anuales a nivel de toda la escuela	1	2	3	4	5
2. Usa datos sobre el desempeño de los estudiantes cuando desarrolla las metas académicas de la escuela	1	2	3	4	5
3. Desarrolla metas que son entendidos fácilmente y utilizados por los profesores de la escuela	1	2	3	4	5
4. Comunica efectivamente la misión de la escuela a los miembros de la comunidad escolar	1	2	3	4	5
5. Se refiere a las metas académicas de la escuela cuando toma decisiones sobre el currículum con los profesores	1	2	3	4	5
6. Asegura que las prioridades en clases del profesor sean consistentes con los objetivos y directrices de la escuela.	1	2	3	4	5
7. Revisa productos del trabajo de los estudiantes cuando evalúa la instrucción en aula.	1	2	3	4	5
8. Aclara quien es responsable de coordinar el currículum en cada nivel (por ejemplo: el director, jefe técnico, profesor encargado de ciclo o asignatura)	1	2	3	4	5
9. Se remite a los resultados en las evaluaciones de toda la escuela cuando toma decisiones curriculares	1	2	3	4	5
10. Participa activamente en la revisión de materiales curriculares	1	2	3	4	5
11. Se reúne individualmente con profesores para discutir el progreso de los estudiantes	1	2	3	4	5
12. Utiliza tests u otra evaluación de desempeños para medir el progreso en las metas escolares	1	2	3	4	5
13. Incentiva a los profesores a usar el tiempo instruccional para la enseñanza y práctica de nuevas habilidades y conceptos	1	2	3	4	5
14. Se toma tiempo para conversar informalmente con estudiantes y profesores durante recreos y recesos	1	2	3	4	5

15. Participa y asiste a actividades extra-curriculares o complementarias	1	2	3	4	5
16. Felicita en privado a profesores por sus esfuerzos o desempeños	1	2	3	4	5
17. Reconoce desempeños excepcionales de los profesores registrándolo en actas personales	1	2	3	4	5
18. Crea oportunidades de desarrollo profesional para profesores como recompensa de contribuciones especiales a la escuela	1	2	3	4	5
19. Lidera o asiste a actividades de perfeccionamiento docente relacionadas con instrucción	1	2	3	4	5
20. Destina tiempo en reuniones del equipo docente para que profesores compartan ideas o información adquiridas en actividades de capacitaciones	1	2	3	4	5
21. Reconoce mejoras o logros superiores de los estudiantes al verlos en su oficina con sus trabajos	1	2	3	4	5
22. Contacta a los apoderados para comunicarles mejoras o desempeños ejemplares o contribuciones de los estudiantes	1	2	3	4	5

Cuestionario de Eficacia Colectiva de Profesores

Este cuestionario busca obtener su percepción sobre el grado en el que los docentes de esta escuela o colegio comparten creencias y actitudes sobre la enseñanza. Por favor marque en cada una de las siguientes afirmaciones la opción en la escala que mejor represente sus opiniones:

	Totalmente en desacuerdo				Completamente de acuerdo	
1. Los profesores de esta escuela tienen lo necesario para que los estudiantes aprendan	1	2	3	4	5	6
2. Los profesores de esta escuela logran aprendizajes con estudiantes difíciles	1	2	3	4	5	6
3. Si un estudiante no aprende algo la primera vez, los profesores intentarán otras formas de enseñarle	1	2	3	4	5	6
4. Los profesores de acá confían en que podrán motivar a sus estudiantes	1	2	3	4	5	6
5. Los profesores en esta escuela realmente creen que cada estudiante puede aprender	1	2	3	4	5	6
6. Si un estudiante no quiere aprender los profesores de acá no persisten	1	2	3	4	5	6
7. Los profesores acá necesitan más formación para lidiar con estos estudiantes	1	2	3	4	5	6
8. Los profesores acá piensan que hay estudiantes a los que nadie les puede enseñar bien	1	2	3	4	5	6
9. Los profesores acá no tienen las habilidades necesarias para producir aprendizajes significativos en los estudiantes	1	2	3	4	5	6
10. Los profesores acá no logran enseñarle a algunos estudiantes por métodos de enseñanza pobres	1	2	3	4	5	6
11. Los estudiantes llegan a esta escuela listos para aprender	1	2	3	4	5	6
12. El ambiente familiar de los estudiantes proporciona tantas ventajas que los estudiantes están obligados a aprender	1	2	3	4	5	6
13. La falta de materiales de enseñanza y recursos dificultan la enseñanza	1	2	3	4	5	6
14. Los estudiantes acá simplemente no están motivados por aprender	1	2	3	4	5	6
15. La calidad de la infraestructura acá facilita el proceso de enseñanza y de aprendizaje	1	2	3	4	5	6
16. Las oportunidades en esta comunidad ayudan a que los estudiantes vayan a aprender	1	2	3	4	5	6
17. Los profesores acá están bien preparados para enseñar los ramos que tienen asignados	1	2	3	4	5	6

18. Los profesores en esta escuela son hábiles en varios métodos de enseñanza	1	2	3	4	5	6
19. El aprendizaje es más difícil en esta escuela porque los estudiantes están preocupados por su seguridad	1	2	3	4	5	6
20.El abuso de alcohol y drogas en la comunidad circundante dificulta que los estudiantes aprendan	1	2	3	4	5	6
21. Los profesores en esta escuela no tienen las habilidades para lidiar con problemas de disciplina de los estudiantes	1	2	3	4	5	6

Appendix 3.4: Teacher Self-Efficacy Scale

Escala de Autoeficacia Percibida en Profesores

Por favor, indique su opinión acerca de cada uno de los enunciados que siguen:

	Nada	Muy poco	Algo	Bas-tante	Harto
1. ¿En qué medida puede usar una variedad de estrategias de evaluación?	1	2	3	4	5
2. ¿En qué medida puede proporcionar una explicación alternativa o un ejemplo cuando los estudiantes están confundidos?	1	2	3	4	5
3. ¿En qué medida puede preparar buenas preguntas para sus estudiantes?	1	2	3	4	5
4. ¿Qué tan bien puede implementar estrategias alternativas en su sala de clases?	1	2	3	4	5
5. ¿Qué tanto puede hacer para controlar la conducta disruptiva en clases?	1	2	3	4	5
6. ¿Qué tanto puede hacer para que los estudiantes sigan reglas en las sala de clases?	1	2	3	4	5
7. ¿Qué tanto puede hacer para calmar a un estudiante disruptivo o ruidoso?	1	2	3	4	5
8. ¿Qué tanto puede hacer para establecer un sistema de manejo de clase con cada grupo de estudiantes?	1	2	3	4	5
9. ¿Qué tanto puede hacer que los estudiantes crean que les va a ir bien con el trabajo escolar?	1	2	3	4	5
10. ¿Qué tanto puedes hacer para ayudar a los estudiantes a darle valor al aprendizaje?	1	2	3	4	5
11. ¿Qué tanto puedes hacer para motivar a los estudiantes que muestran poco interés por el trabajo escolar?	1	2	3	4	5
12. ¿Qué tanto puedes hacer parar asistir a las familias para que ayuden a sus hijos a desempeñarse bien en la escuela?	1	2	3	4	5