Classroom discipline, classroom environment and student performance in Chile

Carolina Gazmuri, Jorge Manzi and Ricardo D. Paredes

ABSTRACT

This paper assesses the extent to which teachers' actions in the classroom influence the school environment. The assessment is based on a statistical analysis of videotaped classroom observations of 51,329 teachers. The classroom environment was found to have a significant influence on students' performance. More specifically, the teacher's ability to handle the class as a group is consistently more significant than other measures of class environment. It was also founds that the overall school environment is a better predictor of students' test results than the environment in the classrooms of the students whose test results are being reported. This suggests that the most effective course of action would be to improve the overall school environment, although individual teachers have less control over this factor.

AUTHORS

Carolina Gazmuri has a Master of Arts degree in Education at Teachers College, University of Columbia. cgazmurib@gmail.com

Jorge Manzi is a full professor in the School of Psychology of the Pontificia Universidad Católica de Chile. jmanzi@uc.cl

Ricardo D. Paredes is a professor in the School of Engineering of the Pontificia Universidad Católica de Chile. rparedes@ing.puc.cl

I

Introduction

Most of the existing studies on scholastic performance focus on schools' sociodemographic variables but devote little attention to teacher performance. In part, this is because teacher evaluations are few and far between and are indirect measurements. In addition, the number of in-class observations that are carried out is quite limited. Furthermore, to the best of the authors' knowledge, no statistical analyses have been made of the effect exerted by the classroom environment. This is attributable to the cost involved in conducting that kind of study, to the resistance of teachers' unions to such a venture and to the fact that the coding of different forms of conduct is such a new area of research that it has not yet reached the stage where it could provide a basis for accurate metrics in this respect. Be this as it may, parents and educators cite the main school-related problems in Chile as being "students' lack of interest" and "a lack of discipline", while survey results indicate that discipline is one of the key considerations for parents when deciding what school to send their children to (Arancibia, 1994).

This study identifies and quantifies classroom-environment variables that influence academic performance. A distinction is drawn between factors that are basically teacher-determined and those that are more closely associated with the school as a whole. The data for this analysis are drawn from evaluations of over 50,000 public-school teachers in Chile based on video observations of one class per teacher that have been assessed and coded by educational psychologists and other education professionals.

While a number of studies have been done in which an attempt is made to detect factors related to the classroom environment that can influence the learning process, most have taken a more psychological approach and have been based on one-off observations of small groups of students. This study's contribution to the literature is based on the measurement and quantification of the influence exerted by a number of classroom-related factors using a much larger sample (51,329 observations) than those used in any previous study. This method complements the more traditional approach to this subject and makes it possible to differentiate the impacts of various classroom-related factors that are difficult to distinguish from one another on the basis of direct observation and case studies.

The study deals with only one specific aspect of the school environment that can, at least in theory, be manipulated by the teacher in the classroom and can be analysed on the basis of observations of a course module. In addition, the focus is restricted to the relationship between this aspect of the school environment and academic performance, which is evaluated on the basis of standardized test results. It is understood, of course, that the school environment influences other aspects of students' and teachers' experiences within the school setting and is, in turn, influenced by the interaction of other, non-observable factors that affect the observations of that environment used in this study.

The rest of this study is divided into three sections. Section II describes the educational situation in Chile and provides other background information about the factors that led up to the development of a teacher evaluation system. It also reviews the existing literature. Section III deals with the model used to arrive at the estimates presented here, while section IV presents a number of conclusions.

II

Background

1. The education system in Chile

Education has been one of the most important issues for the Chilean government and, while there has been a policy of ongoing reform for quite some time, in the early 1980s major changes were introduced that have strongly influenced subsequent developments and have had a strong bearing on the situation that unfolded during the first decade of this century.

The reforms of the early 1980s decentralized the education system by handing over the administration of the country's public schools to its municipalities. In addition, the historical-cost funding system was replaced with per-student subsidies so that pupils could choose which school to attend. These changes did expand school coverage, but they failed to improve the quality of education, which had been one of the goals of the reform.

This decentralization process has a number of critics (Muñoz and Raczynski, 2007), while Beyer (2009) argues that it has been only partial, since, although it is true that the schools are no longer administered by the central government, the municipalities have not been endowed with the necessary capacity to manage them properly. The associated debate concerning the quality of the education provided by municipal schools has been heated, and there is an ample body of literature on the subject (see Drago and Paredes, 2011).

The economic crisis that broke out in 1981 triggered a steep reduction in funding for public schools. Between 1982 and 1990, government spending on education was cut by 29%. Moreover, Chile had no system in place for assessing the quality of education until 1988, when the Education Quality Measurement System (SIMCE) was introduced. The SIMCE tests are still in use today. At first, these test results were not made public, but they have been in the public domain since 1995.

In 1990, with the promulgation of the Teachers Statute, a wage floor was set for teachers and their rights as members of the teaching profession were codified. This law generated a series of rigidities associated with limitations on teacher mobility and on teacher dismissals. In 1991, schools began to be allowed to supplement government funding with school fees. As of 2008, 49% of the country's 11,905 schools were run by the municipalities, 44% were government-subsidized private

schools and 6% were private educational institutions (Ministry of Education, 2008a).

The main tool used to measure the quality of education in Chile is the SIMCE test, which has yielded comparable results only since 1997. From that year on, test results were stable until 2010, when some improvement began to be seen. The figures attest to sharp inequalities in the quality of education. After 4, 8 or 10 years of schooling, a sizeable portion of the student body does not have the basic skills or knowledge expected of students in the corresponding grade (Muñoz and Weinstein, 2009). As of 2008, 35% of all fourth-grade students were rated as having an initial level of proficiency in reading and 41% scored at that level in mathematics (figures taken from reports on national test results: www.simce.cl).

Based on the test results for fourth-graders (9-year-olds, on average) and eighth-graders (13-year-olds, on average) in 1999 and 2000, Eyzaguirre and Le Foulon (2001) conclude that nearly 40% of all elementary school students cannot understand what they read; for students in the second year of secondary school, i.e. tenth grade (15-year-olds, on average), they put the figure at 33%. The 1999 SIMCE test scores indicate that 32% of the students in fourth grade had not mastered the skills and knowledge that a second-grade student should possess; 25% were at the third-grade level; another 25% had an initial level of proficiency for fourth grade; and only 11% had a satisfactory level of proficiency.

Other standardized tests, such the Trends in International Mathematics and Science Study (TIMSS), yield similar results. One out of every two eighth-grade students is at least four years behind in mathematics. What is more, the average score on this test for Chilean students from households having high levels of educational attainment, who perform better than other students in the country, is below the overall international average and is on a par with the average score of students from households having a low level of educational attainment in the Republic of Korea, Slovenia, the Russian Federation, Belgium and others. In addition to the fact that, on average, the quality of education is low, there is also a marked degree of inequality. Of the students who attend private schools, 1 out of 2 score over 300 points on the SIMCE mathematics tests, whereas only 1 out of 5 students in government-subsidized private schools and only 1 out

of 10 students in municipal schools score at least 300 (Fontaine, 2002; Brunner and Cox, 1995; García and Paredes, 2010).

2. The teachers

The assertion made by Barber and Mourshed (2008) that "the quality of an education system cannot exceed the quality of its teachers" (p. 15) has prompted many governments to focus on their faculty and on making a teaching career more attractive. In Chile, the situation is clear. Most of the country's teachers were not near the top of their class when they were in school, and only 1 out of every 24 of the younger teachers who studied education in universities belonging to the Chilean University Council of Rectors were in the top 10% of their graduating class (Claro, 2009). This is corroborated by Cabezas and others (2013), who draw attention to the need to upgrade teacher qualifications, especially in schools with the most vulnerable students.

In 2008, Chile had 176,472 practising teachers, of whom 46% were working at municipal schools, 43% in government-subsidized private schools and just 11% in private schools. A majority of teachers are women (71%). Teachers' salaries rose by around 200% between 1990 and 2008, but that increase has not been pegged to individual performance (Ministry of Education, 2008a).

The idea that the quality of instruction is the touchstone of learning was what underlay the decision in 2003, in the wake of protracted negotiations with the Teachers' Association, to begin evaluating teachers in municipal schools on an individual basis. The design of the teacher evaluation system was highly politicized and extensively negotiated, which might lead one to suspect that it would not provide an accurate evaluation of teacher performance. Contrary to expectations, however, León, Manzi and Paredes (2008) found that the results of teacher evaluations correlate relatively well with the learning outcomes of their students, which would appear to indicate that the evaluation system has been well-designed.

3. School environment and discipline

The surrounding environment, the way that teachers manage their classrooms and school discipline are generally regarded as crucial factors in students' learning experiences (see, for example, Ritter and Hancock, 2007; Nie and Lau, 2009, and references). While these concepts have been defined in various ways in the literature, generally speaking, all of these definitions encompass the steps taken by teachers to keep order in their classrooms, engage their students and elicit their cooperation (Emmer and Stough, 2001).

Kennedy (2005) suggests that the need to manage students in the classroom often interferes with teachers' efforts to convey ideas to them. She contends that, out of fear of losing their students' interest, teachers sacrifice intellectual content in order to keep the situation under control because, if they present material that is too intellectually challenging, some students will back off or become distracted and disruptive because they find that following the class requires too much intellectual exertion.

The classroom environment and discipline have also been identified as a critical factor in teachers' work satisfaction. Time and again, teachers mention school discipline as one of the greatest challenges that they face (Ritter and Hancock, 2007). Discipline problems are also frequently cited as one of the main reasons why some teachers decide to leave the profession (Morris-Rothschild and Brassard, 2006).

Given how strongly the classroom environment influences students' academic performance and teachers' work satisfaction, a number of studies have been conducted in an attempt to identify different discipline management styles and their effectiveness. Three main styles are identified in the literature (Lewis and others, 2008). The first is associated with the idea that teachers should closely control their classrooms and their students' behaviour and with the "assertive discipline" or "take-control" approach first developed by Lee and Marlene Canter in 1970 (Malmgren, Trezek and Paul, 2005). This approach calls for teachers to set out ground rules at the start of the school year in order to make the students aware of what kind of behaviour is expected of them and what types of consequences they can expect if they fail to comply. During class, teachers are encouraged to reward and recognize good behaviour and punish misbehaviour.

Along these same lines, the "interventionist style" is based on the idea that students learn to behave appropriately in the classroom when good conduct is rewarded and bad conduct is punished and that teachers should therefore maintain strict control over the students' activities in the classroom (Ritter and Hancock, 2007).

A second approach places greater emphasis on students' self-control and less on teachers' authority. This discipline management style is associated with the "teacher effectiveness training" model developed by Thomas Gordon, also in the 1970s. This style is based on the idea that students' self-control is key to their good

behaviour in class and that it should be achieved through negotiation and conversations with them (Malmgren, Trezek and Paul, 2005). In this non-interventionist style, students are expected to play an influential role in the classroom. Teachers are not supposed to worry about bringing students' behaviour into line with what they consider suitable because students will tend to behave more appropriately on their own (Ritter and Hancock, 2007).

The third style places emphasis on participation and group decision-making. In this approach, which is based on the "control theory" model developed by William Glasser, students should take responsibility for the behaviour of their classmates and make sure that they conduct themselves properly. This style of discipline calls for frequent course meetings to discuss various behavioural issues and to build consensus around them (Edwards and Mullis, 2003).

In developing countries and in Chile, in particular, few systematic studies have been conducted, although Eyzaguirre and Fontaine suggest that teachers in high-performing schools devote more classroom time to instruction by planning out classroom activities more thoroughly and by managing disciplinary issues and remediation policies more effectively (Eyzaguirre and Fontaine, 2008).

4. Teacher evaluations in Chile

Teacher evaluation procedures for municipal schools in Chile are set out in the 1991 Teachers Statute. The Teachers Association opposed the evaluation so fervently that it was not actually implemented until 12 years later, when the government and the Teachers Association reached an agreement. The agreement provides for the application of the National Teacher Evaluation System in the country's municipal schools. This system relies on four tools, which are weighted as follows: (i) a self-evaluation (10%); (ii) an interview by a peer evaluator (20%); (iii) a review conducted by the director of the corresponding technical pedagogical unit (10%), and (iv) a portfolio (60%).

The portfolio provides evidence about teachers' instructional practices and is composed of two different modules. The first consists of a description of a pedagogical unit, an evaluation of that unit and an analysis. The second contains a 40-minute video of a teacher in action in the classroom.

Teachers can receive the following performance ratings: (i) Outstanding (a score of between 3.1 and 4), which denotes a professional performance that consistently exceeds the level of expectation for the indicator in question; (ii) Competent (a score of between 2.51 and 3), which corresponds to a satisfactory level of performance as a teaching professional that meets at least the minimum established requirements; (iii) Basic (a score of between 2 and 2.4), which denotes a level of professional performance that sometimes meets the expectations for the indicator in question and sometimes does not, and (iv) Unsatisfactory (a score of between 1 and 1.99), which equates to a clearly weak performance on the part of the teacher concerned.

Teachers who receive ratings of "outstanding" or "competent" can apply for the Variable Individual Performance Allowance (AVDI). To obtain this allowance, they have to pass a disciplinary and pedagogical knowledge test. Those who obtain a rating of "outstanding", "competent" or "sufficient" on the test receive pay raises of between 25% and 5% above the national minimum basic wage (RBMN), which is about US$ 213 per month (figure based on the RBMN for 2010 and the average exchange rate for the dollar for that same year). As of 2010, somewhat fewer than 8,000 teachers in Chile were receiving this allowance.

Teachers who receive a "basic" rating in the teacher evaluation have to take part in a professional improvement programme that includes tutorials, courses, workshops, recommended lectures and class observations conducted by qualified teachers. Teachers who receive a rating of "unsatisfactory" have to participate in the improvement programme and be re-evaluated the following year. If a teacher receives an unsatisfactory rating in the second evaluation, he or she must stop teaching for a year, participate in the professional improvement programme and submit to a third evaluation. If the teacher receives another unsatisfactory rating in the third evaluation, he or she is required to leave the teaching profession.

As of 2009, more than 50,000 municipal-school teachers (67% of the total) had been evaluated. Out of that number, as of 2007, 1,050 teachers had received an unsatisfactory rating at some point; 95 of these teachers received an unsatisfactory rating a second time, and 8 were rated as unsatisfactory three times in a row (Araya and others, 2010).

The evaluation focuses on teachers' performance in the classroom, which is monitored by video recordings that are then analysed by a group of educators who evaluate specified items as measured against established standards. Clearly, the videotaping of teachers' performance in the classroom is open to criticism, since it may provide an inaccurate picture of what usually goes on there. Teachers may be nervous and may make

special preparations for the class that is to be filmed, and students may also act differently. In fact, 20% of the teachers who were evaluated in 2005 and 2006 said that it was a very difficult experience. However, nearly 80% said that students behaved, by and large, as they usually did, and less than 10% said that they behaved worse than usual. On another front, Lock and Strong (2010) suggest that in-class evaluations may overlook important psychological aspects of what goes on in the classroom.

This paper posits that videotaping provides a good metric of teachers' classroom performance insofar as any errors that they commit are not related with other characteristics of the educators, their students or the environment that also influence their performance. In other words, if the departures from usual modes of behaviour are random, then the accuracy of the associated estimates will decline, but the estimates will still not be skewed. This situation should be reflected in the data and in the estimates themselves.

III

Data and results

1. The data

The database includes fourth grade and tenth grade students' scores on the various SIMCE tests in 2008 and the test scores of eighth-graders in 2007 in the country's municipal schools. The information covers various characteristics of each student, including the student's gender, the educational level of the parents and household income, as well as features of the school that he or she attends, such as the teacher evaluation scores for 2005, 2006, 2007 and 2008. The individual scores for teachers' videotaped classroom performances are also available.

There are nine indicators for teachers' classroom performance and in-class interactions: (i) students' focus on proposed activities (INDF1); (ii) teachers' skill in handling the class as a group (INDF2); (iii) teachers' promotion of participation by all the students (INDF3); (iv) the quality of class structure (INDG1); (v) the use of time available for instruction (INDG2); (vi) the activities' contribution to the achievement of learning objectives (INDG3); (vii) the quality of teacher explanations (INDH1); (viii) the quality of teacher-fostered interactions (INDH2), and (ix) the learning-process coaching (INDH3).

Figure 1 depicts the distribution of teachers' scores (from 0 to 4) on these indicators. The areas enclosed in the boxes represent the second and third quartiles for this sample; as may be seen, there is little variance in the results, which could theoretically make it difficult to detect the impact of the different scores.

2. Estimation model

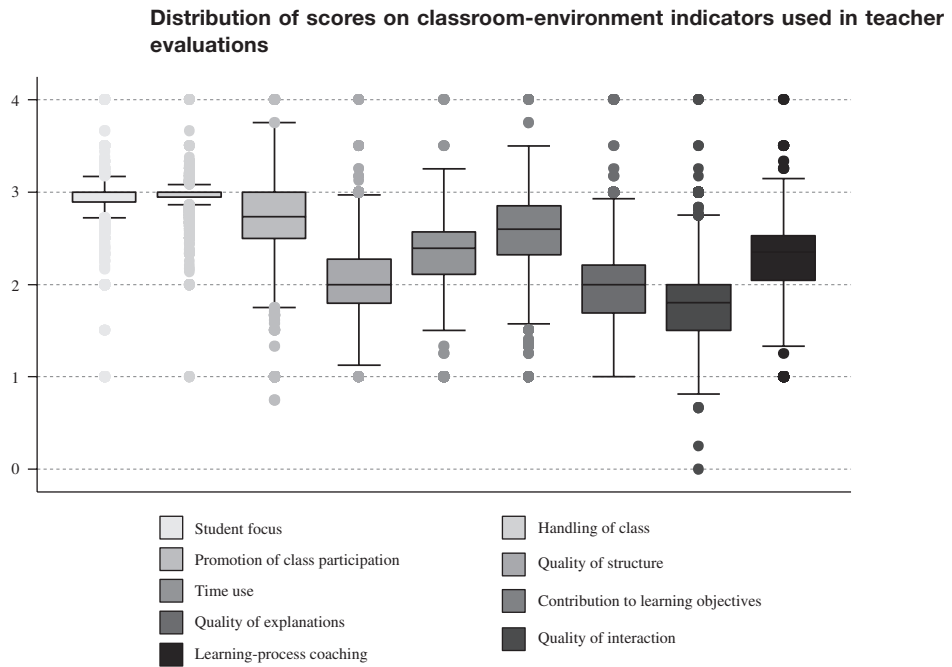To estimate the influence of the classroom environment on student performance, we follow the traditional empirical literature, which considers that a student's academic achievement depends on several factors associated with family situation, the school setting (including teacher-related factors) and environmental factors. Specifically, the model used considers the student academic performance as the dependent variable, measured by the SIMCE test, and as explanatory variables, the student gender, the student socioeconomic situation, a measure of the peer effect, the rural or non-rural status of the school, teacher quality (measured as the overall teacher evaluation, usually a variety of instruments) and, lastly, the classroom climate, the component on which this research focuses. The model is shown in (1).

$$R_{ij} = f\left(A_{ij}, F_{ij}, P_{ij}, E_j, C_j\right) + \varepsilon_{ij} \qquad (1)$$

where R_{ij} is the level of scholastic achievement of student i in school j; A_{ij} denotes student characteristics; F_{ij} stands for the socioeconomic characteristics of the student's family; P_{ij} denotes the characteristics of the student's peers; E_j stands for the characteristics of school j; C_j reflects the environment in school j; and ε_{ij} is the random error.

Since there is no consensus among theorists as to the most appropriate functional form to specify these relationships and since the specification of an unsuitable functional form could generate biased or inconsistent estimates, a range of specifications were tested and then compared using the Bayesian information criterion (BIC). In particular, the criterion developed by Gideon E. Schwarz (1978) was employed to identify the functional form with the lowest BIC; this criterion is closely related to the maximization of the model's likelihood function.

FIGURE 1

Distribution of scores on classroom-environment indicators used in teacher evaluations

Legend:
- Student focus
- Promotion of class participation
- Time use
- Quality of explanations
- Learning-process coaching
- Handling of class
- Quality of structure
- Contribution to learning objectives
- Quality of interaction

Source: prepared by the authors on the basis of the teacher evaluation database of the Ministry of Education.

The results suggest that the various models display a high degree of consistency and robustness. However, based on the BIC, the model employed here includes the following regressors: (i) gender; (ii) father's level of education; (iii) household income (linear and quadratic); (iv) mother's level of education (average); (v) average household income; (vi) rural place of residence; (vii) average score of the school's faculty on the teacher evaluations, and (viii) the average of nine different scores on classroom-environment indicators.

A number of different issues have to be dealt with. First, observations are not independent of one another, since the students are grouped into schools. Therefore, we have two different levels —student and school— and the error term ε_{ij} includes an error at the individual student level, ω_{ij} and an error for the educational institution as a whole, μ_j, which is shared by all the students in that school:

$$\varepsilon_{ij} = \omega_{ij} + \mu_j \qquad (2)$$

Although the aggregations do not yield substantially differing results, ignoring this data structure could lead to biased estimates. If the error term at the school level,

μ_j, is correlated with the regressors, then the ordinary least squares (OLS) estimators will be biased, while if μ_j is independent of the regressors, then the OLS estimators will not be biased, but they will be inefficient.

One possible solution for this problem is to use fixed-effect estimation. The downside of this method when used for our purpose is that the fixed effect is linearly dependent on the variable of interest to us here, school climate, which would make it impossible to identify the effect.

The alternative to using fixed-effect estimators is to use random-effect estimation techniques. However, this method requires the assumption of independence between and the regressors; otherwise, we would obtain biased estimators. If this condition is fulfilled, then the effect estimator yields more efficient estimators (Baum, 2006).

To test the independence of μ_j and the regressors, we used the Hausman test. If μ_j and the regressors are correlated, the fixed-effect estimators will be consistent, but the random-effect estimators will be inconsistent. On the other hand, if μ_j and the regressors are independent, then the fixed-effect estimators will still be consistent (although not efficient) but the random-effect estimators will be both consistent and efficient. Thus, by using the Hausman test to compare fixed-effect and random-effect estimators, we can see whether or not they differ

significantly. If they do, then we can conclude that the assumption of independence is valid (Baum, 2006). The results of the Hausman test indicate that the hypothesis that μ_j and the regressors are independent should be rejected; therefore, the random-effect estimators and the OLS estimators are biased.

To surmount this problem, we follow Mundlak (1978) and use the mean square error estimator (MSEE). The first step in applying this method is to estimate the product solely on the basis of the in-school variables that may change, as in equation (3).

$$R_{ij} = f\left(A_{ij}, F_{ij}, P_{ij}\right) + \varphi_{ij} \qquad (3)$$

The estimated error in equation (3) includes the error associated with a fixed-effect estimate plus the fixed variables effect. The second step is to estimate a regression between the estimated error of (3) ($\widehat{\varphi_{ij}}$) and in-school variables that are not subject to change using a random-effect model, as in (4).

$$\widehat{\varphi_{ij}} = f\left(E_j, C_j\right) + \pi_{ij} \qquad (4)$$

This method yields unbiased estimators for the classroom-environment indicators, provided that there is no endogeneity.

The possible endogeneity of teacher quality, in particular, poses a potential problem. If teachers who have greater classroom skills could choose where they will work, they might prefer the schools that are attended by the best students. In this case, the direction of causality is opposite to the one that we are trying to measure, and our lack of suitable instruments will translate into biased estimators.

The problem of endogeneity is one of the most difficult ones to resolve because, strictly speaking, the only solution is to apply controlled experiments or pseudo-experiments. It can be argued that, because of the complexity of the interacting factors, it is impossible to be sure that no type of endogeneity is present. A weak endogeneity test was used in an attempt to detect any possible endogeneity in the model (Schaffer and Stillman, 2006). This test involves estimating a variable that reflects the effects of factors or characteristics that influence the performance of a school's teachers and have not been included in the model (this is obtained as a residual of the regression of teacher-related factors as

a function of other variables that have been incorporated into the model). The next step is to assess the significance of this estimated variable (factor or characteristics that influence teacher performance that have not been included in the model) in accounting for the error in the original model at the school level. If this variable is significant, then weak endogeneity is found to be present. For our model, the variable is not significant with a p-value close to 1, which suggests that, given the above caveat, there is insufficient evidence to support the endogeneity hypothesis.

Estimates were calculated for the language and mathematics tests for 2007 and 2008 and for fourth and eighth grades. Tables 1 and 2 show the results generated by HLM for mathematics in the fourth and eighth grades.

The results of the Fisher test for all the variables for the school setting indicate that they are part of the model and have a significant effect. They also suggest that the model's predictive power is substantially greater for the higher grades, as the estimates for fourth grade have an R^2 of 7.8%, whereas the estimates for tenth grade have an R^2 of 28%. This differential is primarily due to inter-school differences in R^2: 10% for fourth grade and 49% for tenth grade.

The regressors have the sign and significance typically found in the literature (García and Paredes, 2010) and therefore seem to be satisfactory controls. The teacher-performance and classroom-environment variables are interesting: while the overall teacher evaluation is highly relevant, the only indicator for classroom environment that is consistently significant in the regressions for the different tests and courses is "handling of the class as a group."

This item is closely related to the teacher's objectives in terms of the direction in which he or she is trying to steer the class's learning experience, and although it does not incorporate the amount of time devoted to administrative tasks or to maintaining control (which is depicted in the literature as running counter to the achievement of educational objectives), it seems highly likely that the energy devoted to handling the class and the energy devoted to these other tasks do, in fact, represent trade-offs. The "handling of the class as a group" variable is consistently positive and significant in the different models. This appears to reflect a teacher's skill in handling the class and eliciting objectively desired forms of observable behaviour that will be conducive to learning (Kennedy, 2005).

More specifically, the significance and relevance of the estimated effect of "handling of the class as a group" in the fourth-grade SIMCE mathematics test scores

TABLE 1

Results for the fourth-grade SIMCE mathematics test

	Coefficient (standard error)	Coefficient (standardized)
First model		
Woman	-4.037*** (0.303)	-0.039
Father's education	0.450*** (0.042)	0.039
Mother's education	1.434*** (0.052)	0.103
Income (measured in $10 000 increments)	0.232*** (0.015)	0.128
Income (quadratic)	-.001*** (0.000)	-0.099
Peer effect – mother's education	5.886*** (0.334)	0.186
Peer effect – income	0.757 (0.715)	0.012
Model 1 constant	136.488*** (10.268)	
Second model		
Urban	-12.762*** (0.899)	-0.098
Teacher evaluation score	12.754*** (1.285)	0.072
Student focus on activities	0.725 (2.916)	0.002
Handling of the class as a group	12.191*** (3.674)	0.027
Encouraging all students to participate	3.437** (1.552)	0.015
Quality of class structure	-1.678 (1.266)	-0.009
Use of time available for instruction	1.566 (1.691)	0.006
Activities' contribution to objectives	1.358 (1.234)	0.006
Quality of explanations	0.976 (1.721)	0.005
Quality of teacher-fostered interactions	-3.187* (1.638)	-0.018
Learning-process coaching	0.246 (1.950)	0.001
Model 2 constant	-71.811*** (10.268)	
Number of observations	104 656	
Number of groups	4 281	
In-school R^2	0.036	
Inter-school R^2	0.104	
R^2	0.0787	

Source: prepared by the authors on the basis of Ministry of Education teacher evaluation data.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2

Results for the tenth-grade SIMCE mathematics test

	Coefficient (standard error)	Coefficient (standardized)
First model		
Woman	-9.321*** (0.373)	-0.08
Father's education	0.075* (0.042)	0.007
Mother's education	0.729*** (0.059)	0.047
Income (measured in $10 000 increments)	0.138*** (0.017)	0.069
Income (quadratic)	-0.0004*** (0.0001)	-0.038
Peer effect – mother's education	9.419*** (0.26)	0.287
Peer effect – income	4.925*** (0.581)	0.079
Model 1 constant	102.367*** (2.66)	
Second model		
Urban	-5.316* (3.14)	-0.017
Teacher evaluation score	17.101*** (3.736)	0.124
Student focus on activities	39.699*** (11.98)	0.155
Handling of the class as a group	25.453** (10.79)	0.089
Encouraging all students to participate	3.599 (4.294)	0.019
Quality of class structure	0.065 (3.645)	0.001
Use of time available for instruction	11.485** (5.05)	0.08
Activities' contribution to objectives	-10.019** (4.378)	-0.083
Quality of explanations	3.79 (4.398)	0.031
Quality of teacher-fostered interactions	-2.752 (4.182)	-0.024
Learning-process coaching	-0.034 (6.247)	-0.0002
Model 2 constant	-256.283*** (27.046)	
Number of observations	74 912	
Number of groups	681	
In-school R^2	0.0509	
Inter-school R^2	0.4905	
R^2	0.2834	

Source: prepared by the authors on the basis of Ministry of Education teacher evaluation data.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

is high. The standardized coefficient for this variable —0.27— means that, if teachers' scores on handling the class rise by one standard deviation (0.35 points out of a total of 4), then the SIMCE scores of their students would increase by 0.27 of a standard deviation (about 15 points).

The standardized effect of skill in handling the class on mathematics test scores tends to be three times as great at grade 10 relative to grades 4 and 8, although, because there is so little variance in the responses, we cannot be sure that this difference is statistically significant. While, for some years, the results for variables such as "students' focus on proposed activities," "quality of class structure," "activities' contribution to goal achievement" and "use of time available for instruction" yield the expected signs and significance levels, they do not do so consistently across different tests, courses and years. Finally, the variables "promotion of participation by all the students," "quality of teacher explanations," "quality of teacher-fostered interactions" and "learning-process coaching" are consistently not significant as explanatory variables for student performance.

There are two possible reasons why the variables other than "handling the class as a group" do not appear to be influential or are influential only for some tests. One is that the characteristics represented by these variables actually do not have a significant impact on students' academic performance. The other is that the instrument being used to measure them is not capturing the relevant characteristics, either because of imperfect observations or because teachers are able to modify these aspects for the classroom session that they know is going to be observed.

In order to delve more deeply into the significance of the school environment, one of the questions that needs to be answered is whether the aspects that are key to learning are ones that can be manipulated by the teacher or teachers, or whether they correspond more closely to the overall school environment. In other words, are there schools in which the overall environment is more welcoming and conducive to student commitment, and the actions of the teachers that take place within that framework exert less influence than the environment itself?

In order to test this hypothesis, we replace the variable constructed from the average for all the teachers in a school with a variable constructed only from the observations of the teachers in the relevant cycle and course subject area. Thus, for example, the regressors for the SIMCE mathematics test for fourth grade would now be the average scores obtained by instructors who teach mathematics in the first (elementary) cycle. Clearly, this is a more direct test of the effect of discipline in the classroom in which the relevant subject is taught, since, in addition to focusing on the teachers who are directly involved in the corresponding subject matter, it is also focusing on the relevant cycle. In this case, it is to be expected that the significance of the coefficients in question will be substantially greater.

The results are surprising. In general, the estimated effect of the "handling the class as a group" variable remains significant for nearly all the tests, but the specific estimate is almost halved, even though the estimator is not statistically smaller (the confidence intervals at a 95% significance level overlap).

As in the preceding case, the estimates were calculated for the 2007 and 2008 language and mathematics tests for fourth grade, eighth grade and the second year of secondary school (tenth grade). Table 3 shows the results for the tenth-grade mathematics tests. The estimates for the other grade levels and other tests reflect the same trends.

Finally, table 4 gives the values of this standardized coefficient for the various tests and the two specifications, i.e. using the average for all teachers and using the average for the teachers of the relevant subject in each cycle.

TABLE 3

Results for the tenth-grade SIMCE mathematics test based on secondary-school mathematics teachers

	Coefficient (standard error)	Coefficient (standardized)
First model		
Woman	-9.321*** (0.389)	-0.08
Father's education	0.079 * (0.044)	0.007
Mother's education	0.734*** (0.061)	0.047
Income (measures in $10 000 increments)	0.132*** (0.017)	0.066
Income (quadratic)	-0.0004 *** (0.0001)	-0.037
Peer effect – mother's education	9.692*** (0.269)	0.295
Peer effect – father's education	4.692*** (0.599)	0.075
Model 1 constant	99.816*** (2.772)	
Second model		
Urban	-5.562 (3.709)	-0.017
Teacher evaluation score	13.634*** (2.291)	0.099
Student focus on activities	-2.64 (4.987)	-0.01
Handling of the class as a group	12.523** (5.39)	0.044
Encouraging all students to participate	0.861 (2.862)	0.004
Quality of class structure	2.121 (1.763)	0.021
Use of time available for instruction	0.758 (2.521)	0.005
Activities' contribution to objectives	-4.506** (1.966)	-0.037
Quality of explanations	0.688 (2.031)	0.006
Quality of teacher-fostered interactions	-2.079 (1.938)	-0.018
Learning-process coaching	2.52 (2.819)	0.016
Model 2 constant	-67.412*** (15.159)	
Number of observations	69 265	
Number of groups	602	
In-school R2	0.052	
Inter-school R2	0.4459	
R^2	0.26892	

Source: prepared by the authors on the basis of Ministry of Education teacher evaluation data.

* $p < 0.1$; ** $p < 0,05$; *** $p < 0,01$.

TABLE 4

Comparison of the "handling the class as a group" indicator coefficients for schoolwide averages versus cycle-specific averages

Standardized coefficients for "handling the class as a group"

		School	Cycle
4th	Mathematics	0.027	0.023
	Language	0.034	0.026
8th	Mathematics	0.021	0.024
	Language	0.032	0.022
10th	Mathematics	0.089	0.044
	Language	0.038	0.025

Source: prepared by the authors on the basis of Ministry of Education teacher evaluation data.

IV

Conclusions

In this study we measured and quantified the influence that the school and classroom environment has on academic performance. The fact that the way in which the class is handled proves to be statistically significant and educationally relevant suggests that guiding collective student behaviour in the classroom is key in successful learning. It can also be inferred from the results that the method used to measure the characteristics of interest (videotaping a class as a teacher evaluation tool) makes it possible to distinguish between teachers who are able to handle their class and those who are not. Consequently, evaluators can use this tool to evaluate teachers on their skill in handling their students' group behaviour.

The only facets of the school environment that have been considered, however, are those that can be modified by a teacher inside the classroom and can be perceived by a classroom observer. The atmosphere in the classroom is, of course, influenced by many other factors that do not show up on a video and that could skew the estimates if they correlate with variables used in the model. There is no actual indication of such a correlation, however, and the interpretation arrived at, albeit with some caution, is that the classroom atmosphere and, hence, a teacher's skill in exerting a positive influence on it, are important factors.

The public policy implications of these findings clearly point in the direction of the training provided to teachers-to-be at university, since this is where they develop leadership skills and their personal approach and acquire knowledge about student psychology and classroom management. A corollary may be that reducing class size would help teachers to manage their classes more successfully (Angrist and Lavy, 1999).

Finally, at the level of school policy, it seems clear that the overall school environment, above and beyond what goes on in the classroom, is a factor that has a significant influence on learning outcomes.

Bibliography

Angrist, J. and V. Lavy (1999), "Using Maimonides' rule to estimate the effect of class size on scholastic achievement", *The Quarterly Journal of Economics*, vol. 114, No. 2, Oxford University Press.

Arancibia, V. (1994), "La educación en Chile: Percepciones de la opinión pública y de expertos", *Estudios Públicos*, 54, Santiago, Chile, Centro de Estudios Públicos.

Araya, C. and others (2010), "Theory underlying a national teacher evaluation program", *Evaluation and Program Planning*, vol. 33, No. 4, Amsterdam, Elsevier.

Barber, M. and M. Mourshed (2008), *How the World's Best-performing School Systems Come out on Top*, London, McKinsey and Company.

Baum, C. (2006), *An Introduction to Modern Econometrics Using Stata*, Stata Press.

Beyer, H. (2009), "¿Qué hacer con la educación pública?", *Estudios Públicos*, No. 114, Santiago, Chile, Centro de Estudios Públicos.

Brunner, J.J. and C. Cox (1995), "Dinámicas de transformación en el sistema educacional de Chile", *Education, Equity and Economic Competitiveness in the Americas*, J.M. Puryear and J.J. Brunner (eds.), Washington, D.C., Organization of American States.

Cabezas, V. and others (2013), "Los determinantes del primer trabajo para profesores de Educación Básica de la Región Metropolitana", *Evidencias para políticas públicas en educación*, Sixth cycle of the Research and Development in Education Fund (FONIDE) competitive grant award process, Santiago, Chile.

Claro, F. (2009), "Comentarios a la presentación de Inger Enkvist en el CEP", *Inger Enkvist y la educación actual, Estudios Públicos*, No. 115, Santiago, Chile, Centro de Estudios Públicos.

Drago, J.L. and R. Paredes (2011), "The quality gap in Chile's education system", *CEPAL Review*, No. 104 (LC/G.2498-P), Santiago, Chile, August.

Edwards, D. and F. Mullis (2003), "Classroom meetings: encouraging a climate of cooperation", *Professional School Counseling*, vol. 7, No. 1.

Emmer, E. and L. Stough (2001), "Classroom management: a critical part of educational psychology, with implications for teacher education", *Educational Psychologist*, vol. 36, No. 2, Taylor & Francis.

Eyzaguirre, B. and L. Fontaine (2008), "Las escuelas que tenemos", *Estudios Públicos*, No. 111, Santiago, Chile, Centro de Estudios Públicos.

Eyzaguirre, B. and C. Le Foulon (2001), "La calidad de la educación chilena en cifras", *Estudios Públicos*, No. 84, Santiago, Chile, Centro de Estudios Públicos.

Fontaine, A. (2002), "Equidad y calidad de la educación: Cinco proposiciones interrelacionadas", *Estudios Públicos*, No. 87, Santiago, Chile, Centro de Estudios Públicos.

García, C. and R. Paredes (2010), "Reducing the educational gap: good results in vulnerable groups", *Journal of Development Studies*, vol. 46, No. 3, Taylor & Francis.

Hausman, J. and W. Taylor (1981), "Panel data and unobservable individual effects", *Econometrica*, vol. 49, No. 6, The Econometric Society.

Kennedy, M. (2005), *Inside Teaching, How Classroom Life Undermines Reform*, Harvard University Press.

León, G., J. Manzi and R. Paredes (2008), "Calidad docente y rendimiento escolar en Chile: Evaluando la evaluación", thesis, Santiago, Chile, Pontificia Universidad Católica de Chile.

Lewis, R. and others (2008), "Students' reaction to classroom discipline in Australia, Israel, and China", *Teaching and Teacher Education*, vol. 24, No. 3, Amsterdam, Elsevier.

Lock, A. and T. Strong (2010), *Social Constructionism*, Cambridge, Cambridge University Press.

Malmgren, K., B. Trezek and P. Paul (2005), "Models of classroom management as applied to the secondary classroom", *The Clearing House*, vol. 79, No. 1, Taylor & Francis.

Ministry of Education (2008a), *Estadísticas de la educación 2008* [online] www.mineduc.cl.

———— (2008b), "Resultados nacionales 2008" [online] www.simce.cl.

Morris-Rothschild, B. and M. Brassard (2006), "Teachers' conflict management styles: the role of attachment styles and classroom management efficacy", *Journal of School Psychology*, vol. 44, No. 2, Amsterdam, Elsevier.

Mundlak, Y. (1978), "On the pooling of time series and cross section data", *Econometrica*, vol. 46, No. 1, The Econometric Society.

Muñoz, G. and D. Raczynski (2007), "Reforma educacional chilena: El difícil equilibrio entre la macro y la micro política", *Serie Estudios Socio/Económicos*, No. 31, Santiago, Chile.

Muñoz, G. and J. Weinstein (2009), "Calidad para todos: La reforma educacional en el punto de quiebre" [online] http://www.ceppe.cl/articulos-liderazgo-y-escuelas-efectivas/174-calidad-para-todos-la-reforma-educacional-en-el-punto-de-quiebre-weinstein-munoz.

Nie, F. and S. Lau (2009), "Complementary roles of care and behavioral control in classroom management: the self-determination theory perspective", *Contemporary Educational Psychology*, vol. 34, No. 3, Amsterdam, Elsevier.

OECD (Organisation for Economic Co-operation and Development)/World Bank (2009), *Revisión de políticas nacionales de educación*, Paris.

Opdenakker, M.C. and J.V. Damme (2006), "Teacher characteristics and teaching styles as effectiveness enhancing factors of classroom practice", *Teaching and Teacher Education*, vol. 22, No. 1, Amsterdam, Elsevier.

Paredes, R. and J. Pinto (2009), "¿El fin de la educación pública en Chile?", *Estudios de Economía*, vol. 36, No. 1, Santiago, Chile, Universidad de Chile.

Ritter, J. and D. Hancock (2007), "Exploring the relationship between certification sources, experience levels and classroom management orientations of classroom teachers", *Teaching and Teacher Education*, vol. 23, No. 7, Amsterdam, Elsevier.

Schaffer, M. and S. Stillman (2006), "XTOVERID: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2, xthtaylor", *Statistical Software Components*, Boston College.

Schwarz, G. (1978), "Estimating the dimension of a model", *The Annals of Statistics*, vol. 6, No. 2, Institute of Mathematical Statistics.