

ORIGINAL ARTICLE

Open Access



Failing to notice? Uneven teachers' attention to boys and girls in the classroom

Marina Bassi¹, Mercedes Mateo Díaz^{2*} , Rae Lesser Blumberg³ and Ana Reynoso⁴

* Correspondence: mercedesma@iadb.org

²Inter-American Development Bank,
1300 New York Avenue, NW,
Washington, DC, USA
Full list of author information is
available at the end of the article

Abstract: This paper analyzes whether teachers' attention to boys and girls differs in low-performing schools in Chile, where large gender gaps in test scores are also observed. We coded 237 videotaped classes of fourth graders, identifying specific behaviors of teachers toward boys and girls. The results show a general imbalance in teachers' attention and interactions favoring boys. Gender attention gap is correlated with lower scores in math for girls on Chile's national standardized test (SIMCE). The gender attention gap was also greater in general in classrooms in which teachers had overall worse interactions with students, as measured by the Classroom Assessment Scoring System (CLASS). The evidence in this paper contributes to the discussion about whether traditional measures of teacher–student interactions really capture all that matters for learning.

JEL Classification: O12, J16, I2

Keywords: Gender gap, Quality of teacher–student interactions, Student learning

1 Introduction

Teaching quality is critical to fostering student learning. An effective teacher can significantly reduce the learning gap between low- and high-performing students (Kane and Staiger 2012; Araujo et al. 2016). A high-quality teacher can produce a gain in achievement equivalent to one additional school year (Rivkin et al. 2005). Good teachers are also associated with lower school dropout rates and can have long-term impacts on outcomes such as university enrollment and income (Chetty et al. 2011).

Teachers' effectiveness cannot be predicted by observable characteristics such as the level of education, experience (at least after the first few years), age, or the type of contract (Araujo et al. 2016). Effective teaching is associated with the quality of the interactions between teachers and their students, with what happens inside the classroom and how the teacher uses the time in class. Students learn more when teachers spend more class time on academic activities, keep students engaged for longer periods of time, and minimize the time spent on classroom management activities (taking attendance, explaining the schedule for the day, distributing papers, and so forth) (Bruns and Luque 2015; Stallings and Knight 2003).

But what happens when teachers engage more or differently with some students than with others? This paper addresses whether differences between boys and girls in learning outcomes are associated with differences in the quantity and quality of teacher–student interactions. Research on teachers' effectiveness concludes that classroom dynamics and interaction matters, but little work has been done to understand whether

teachers' interactions vary depending on the gender, race, or income level of the student and how these differences affect learning outcomes. The evidence in this paper contributes to the broader discussion of whether traditional measures of teacher–student interactions that focus, in an aggregate manner, more on quality than on quantity capture all that matters for learning.

Using videotapes, we analyze teaching practices in Chile, a country with one of the largest gender gaps in academic achievement. The results show a pattern of differentiated attention favoring boys. Our measures of gender biases (or differentiated attention) in the classroom correlate with the overall quality of teacher–student interactions measured with the Classroom Assessment Scoring System (CLASS). Teachers who perform worse on CLASS also tend to demonstrate greater attention bias by student gender.

We also estimate the correlation between attention bias in the classroom by gender and students' scores on Chile's national test, the Sistema de Medición de la Calidad de la Educación (SIMCE). Girls whose teachers demonstrate greater attention bias favoring boys tend to have lower math scores. The measures of gender bias for student-initiated actions (calling out when not called on) have a statistically significant positive correlation with boys' scores in math and a statistically significant negative correlation for girls. These correlations do not indicate causation.

The paper is organized as follows. Section 2 presents an overview of the literature. Section 3 presents the conceptual model, the sample characteristics, and the setting. Section 4 describes the coding strategy and variables. Section 5 presents the findings and analyzes the correlation between teacher gender bias and student academic performance, as well as between teacher gender bias and measures of the quality of teacher–student interactions. Section 6 summarizes the paper's conclusions and identifies policy implications.

2 Previous literature

Gender differences in educational outcomes have received increasing attention in recent years (Bos et al. 2014; Guiso et al. 2008; OECD 2014; Straus 2015). Many countries have marked gender gaps in enrollment, graduation, or learning achievements, some favoring girls, others favoring boys. On international tests, boys tend to perform better in math and science while girls tend to score higher in reading and language. On the OECD's 2015 Programme for International Student Assessment (PISA),¹ for example, average girls' scores were 30 points higher than boys' in reading, and boys' average scores were 5 points higher than girls' in math (OECD 2015b). Girls outperformed boys on reading in all 72 countries that administered the 2015 PISA; boys outperformed girls in 47 countries.

Math results are especially important, because they are good predictors of future achievement; school test scores and results on university access exams determine, among other factors, career decisions and contribute to differences in employability, occupational segregation in the labor market, and differences in earnings (Bos et al. 2014; Heckman 2011; Mizala 2014; Murnane et al. 1995; Ñopo 2012; Paglin and Rufolo 1990; Lavy and Sand 2015; Terrier 2016).

Using data on Chile, Bharadwaj et al. (2015) show that learning gaps in math (favoring boys) increase with age and over the schooling cycle. They do not find any correlation with parental background or investment in the child, unobserved ability, the

gender of the teacher, or other observable variables of the classroom environment. Research from the USA finds that boys are more sensitive than girls to environmental influences, such as home environment and neighborhood quality (Autor et al. 2016; Bertrand and Pan 2013; Chetty et al. 2016). Carneiro et al. (2017) find the opposite pattern in Ecuador, where daughters of mothers with university education outperform girls whose mothers have only primary education by 0.58 standard deviations; the figure for boys is 0.47 standard deviations.

As teachers are a key component of students' learning, their behavior could also potentially be associated with gender gaps in educational outcomes. Research based on classroom observations in the USA and Latin American countries shows that most effective teachers spend more time using "active" instructional methods (directly engaging with students through explanation, lecture, and questions and answers) than "passive instruction" (walking around the classroom monitoring as students work individually on their assignments). Effective teachers also minimize the time spent on classroom management activities (taking attendance, explaining schedule for the day, distributing papers, and so forth) (Stallings and Knight 2003; Bruns and Luque 2015).

Carneiro et al. (2017) examine whether differences in test scores between boys and girls in Ecuador vary with differences in the quality of the teacher–students interactions (measured by CLASS). Despite the significant gap in math test scores (in favor of boys), they find that, with their data, the quality of interactions cannot predict these gender differences in learning outcomes. As the authors state, however, their data would not reveal whether there was generalized behavior among teachers in Ecuador biased against girls. CLASS measures the general quality of teacher–student interaction; it does not disaggregate teachers' attention and actions by student characteristics.

Related research finds that teachers' gender biases in primary school affect the academic achievement gap during middle school and secondary school and enrollment in advanced level courses in math and science during secondary school (Lavy and Sand 2015; Terrier 2016). Measuring teachers' gender biases by comparing "blind" and "non-blind" (with and without the name of the student) classroom exams in Israel, Lavy and Sand (2015) show that teachers seem to unconsciously discourage female students by underestimating their abilities while overestimating the skills of their male classmates. Using the same approach, Terrier (2016) finds evidence of gender bias against boys among middle school teachers in France.

Only a few studies have used videotapes to look at gender bias in teacher classroom practice (e.g., Davis 2000; Sadker and Sadker 1994). Sadker and Sadker (1994) analyze videotapes to do so in the fourth, sixth, and eighth grades in four states and the District of Columbia in the USA. Their analysis shows the following:

1. Teachers interact more with boys than with girls.
2. Boys receive more praise, criticism, and remediation than girls.
3. During a discussion, boys are eight times more likely than girls to call out (shout out answers even when not called on).
4. Teachers are less likely to reject behavior by boys, even if it violates classroom rules.²
5. Girls receive more "acceptance" (a bare acknowledgment of their work, such as "uh-huh" or "okay") than boys.

6. Girls who receive less attention from their teachers may come to underestimate their abilities and lose motivation.

The literature on developed countries (most of it based on qualitative analysis of small samples) indicates that getting more of a teacher's attention—whether positive (e.g., responding to or working one-on-one with the student) or negative (e.g., disciplining the student)—has consequences for students' performance (Sadker and Sadker 1994). Most research points to a prevalence of gender bias in favor of boys across subject areas and school environments, mostly in the form of teachers giving more attention to boys than girls (AAUW 1992; LaFrance 1991; Sadker and Sadker 1994; Sadker et al. 1993).

3 Setting and methodology

3.1 Study case

Chile presents a paradox with respect to gender. It has full equality in health and survival and parity in literacy and enrollment in education, but it lags behind with respect to women's economic participation and opportunity (WEF 2015).

Gender gaps in academic achievement are also large in Chile. First, more boys than girls are enrolled in primary school (Table 1). Among all fourth graders in Chile, there were 1.2 boys for every girl in 2011 and 2012. In the sample used for this study, the ratios were 1.4 in 2011 and 1.2 in 2012.

Second, Chile's gender gap on the 2015 PISA in math was the third-largest among member countries of the OECD, with boys outperforming girls (OECD 2015a, 2015b).

Third, among students completing tertiary education, the share of women is one of the highest among OECD countries, but female labor force participation is one of the lowest. The completion rate is higher for girls than boys in Chile at almost all education levels (Bassi et al. 2015).

3.2 Conceptual model

International evidence reveals significant gender differences in educational outcomes that increase with age and schooling cycle (Kersey et al. 2018; Banjong 2014; Robinson

Table 1 Ratio of boys to girls in sample schools and all schools in Chile, in 2011 and 2012

Socioeconomic status of school	Sample schools		All schools	
	2011	2012	2011	2012
Low	1.5057	1.0133	1.1603	1.1974
Medium-low	1.3592	1.2551	1.2728	1.2501
Medium	1.2703	1.2281	1.2438	1.1336
Medium-high			1.1593	1.133
High			1.1129	1.278
Average	1.3671	1.1961	1.2164	1.1736

Note: Sample schools refer to schools included in this paper. All schools refer to the universe of Chilean schools that administer the SIMCE standardized national test. The socioeconomic status of the school for the analysis at the school level was adopted as defined in the public databases of SIMCE. The Education Quality Assurance Agency, responsible for the SIMCE, classifies schools into socioeconomic categories based on four variables: mothers' years of education, fathers' years of education, monthly income reported by parents, and a vulnerability index developed by the Ministry of Education

and Lubienski 2011). To explain those differences, we hypothesize that teachers, an important ingredient in the production function of learning, may be doing something different in their interactions with different groups of students. Based on existing evidence, we hypothesize that the amount of attention teachers devote to students is important for learning and that it is different for boys and girls.

Why would teachers behave differently with boys and girls? Teacher biases may reflect a combination of teachers' reactions to students' behaviors and teachers' own expectations of students, beliefs, and stereotypes. It is well documented that boys and girls behave differently in class (Measor and Sykes 1992; Erden and Wolfgang 2004; Sadker and Sadker 1994); teachers may just be responding to these differences. However, evidence also documents teacher biases and the fact that they are often unaware that their interactions are different with different groups of students (Consuegra et al. 2016; Terrier 2016; Lavy and Sand 2015; OECD 2015a; Shumow and Schmidt 2013; Robinson et al. 2011; Van Duzer 2006; Olszewski-Kubilius and Turner 2002; Tiedemann 2002; Li 1999; Fennema et al. 1990).

Traditional measures such as CLASS scores may be overlooking important dimensions of the quality of interactions. A study from Ecuador shows that the quality of interactions between teachers and students affects differences in math scores among children in early elementary school (Carneiro et al. 2017). CLASS looks at teacher class organization, instructional support, and emotional support (see Appendix). Boys tend to be more active than girls; teachers may instinctively pay more attention to boys in order to control the classroom. Within the pool of "good teachers," CLASS will be unable to distinguish between teachers who pay more attention to boys to control and maintain order in the classroom and teachers who try to balance their interactions with boys and girls. The fact that CLASS does not capture attention to different groups of students may explain why Carneiro et al. (2017) find no correlation between CLASS and gender differences in learning outcomes.

The literature on classroom interactions focuses on the nature, quality, and quantity of interactions between teachers and students in different domains in the classroom (Pianta et al. 2008a; Stallings and Knight 2003; Bruns et al. 2016). CLASS does not examine the content of what is being taught or the specific activities carried out. Using a very detailed protocol for each dimension, observers give scores (on a 1–7 scale) per domain (classroom organization, instructional support, emotional support) and overall (see Appendix). Each observation is based on two 15–20-min segments of the class.

A different strand of the literature examines what happens in the classroom. It generates quantitative data on teachers' practice by coding segments of time and observing activities and participation (Bruns et al. 2016). The coding strategy and the variables generated for this study are more closely related to this line of research. One of the advantages of using coding methods based on counting occurrences and time the length of interactions, is that this method can generate quantitative data from qualitative information.

3.3 Data and methodology

The sample includes 237 tapes (almost 590 h of videotaped classes) from the classrooms of 137 academically low-performing schools in Chile (based on 2009 SIMCE

scores). These schools belong to a random sample designed for an impact evaluation of a program implemented in 2011 by Chile's Ministry of Education to improve learning outcomes in math and language of students from pre-kindergarten to fourth grade (Bassi et al. 2016). Eligible schools included public and subsidized private schools in which the average SIMCE scores in 2005–2009 in math and language were below the national average (250) and the number of students per level from pre-kindergarten to fourth grade averaged at least 20. For the year of the tapings (2012), SIMCE scores were available by school, student, and teacher.

SIMCE has been applied to all schools in Chile annually since 1988 to monitor students' learning in fourth, eighth, and tenth grades in math, reading, and social sciences.³ SIMCE also collects information about teachers, students, and parents, through complementary questionnaires. The test is applied in October to early November every year. In this study, we use SIMCE test scores for fourth grade in 2012 for the schools in our sample.

As the sample used for this study is not representative, results cannot be inferred to all Chilean schools. The gender gap in the sample schools is higher than average; it may, therefore, be easier to detect teachers' bias in low-performing schools. The large gender gap observed in both PISA and national exams is representative of all schools, however.

The videos were coded using the CLASS instrument, which provides good measures of specific dimensions of teaching quality (Pianta et al. 2008a). Each of the 237 videotapes shows a single teacher instructing fourth grade students. One teacher in each of the fourth grades of the 137 schools in the sample was videotaped for four pedagogic hours. The tapes cover 69 math classes and 168 classes on other subjects, mainly language arts.⁴

The program implemented in 2011 by the Ministry of Education did not include any component or activity explicitly addressing gender bias in teaching. As the videotapes were produced for another purpose, they proved amenable for studying gender bias in the classroom, as there was no reason to believe that teachers may have altered their behavior in this respect. The taping strictly followed the protocol of the upper elementary version (fourth to sixth grade) of CLASS (Pianta et al. 2008b). Several studies link better student outcomes (in both learning and the development of socioemotional skills) with teacher scores.⁵

Most of the schools in the sample are public schools with students of medium-low socioeconomic status (according to the SIMCE classification) (Table 2). The average 2012 SIMCE score for schools in the sample for combined subjects was 243 (Table 3), 20 points below the national average for that year of 263. The score gap by gender is 10.0 points in language arts (favoring girls) and 4.6 points in social sciences (favoring boys), both similar to the gaps at the national level. In math, the gap is 5.8 points (favoring boys), about 3 points higher than Chile's average gap. All the differences between boys and girls in test scores are statistically significant, with p values for the test of differences in means of 0.

A teacher questionnaire, including questions on educational background, experience, and tenure (number of years in the same school), complemented the videos. Most teachers in the schools studied were women (92.5%) (Table 4). The fact that the sample included very few male teachers made it impossible to differentiate teacher's behavior

Table 2 Descriptive statistics of schools in the sample

Item	Number	Percentage
Type of school		
Municipal (public)	88	64.7
Private, subsidized	48	35.3
Socioeconomic level of school		
Low	25	18.4
Medium-low	88	64.2
Medium	24	17.5
Number of fourth grade basic level classes at school		
1	85	62.5
2	43	31.6
3	8	5.9
Average number of fourth grade students at school		
Less than 25	39	28.7
25–34	71	52.2
More than 34	26	19.1
Number of fourth grade teachers recorded by school		
1	104	76.5
More than 1	32	23.5
Makeup of school and classroom		
Number of basic level students in fourth grade classroom	29.2	
Percentage of boys at school	53.3	
Percentage of boys in fourth grade classroom	51.4	

Source: Schools' administrative data and results of the Sistema de Medición de la Calidad de la Educación (SIMCE)

by their gender. About 90% of teachers had university degrees. Almost the same number of teachers had less than 5 years of experience, 5–10 years, 11–24 years, or more than 25 years.

4 Coding strategy and categories

This study builds on the work of Sadker and Sadker (1994), expanding their coding categories (of remediation, praise, criticism, and acceptance). Following these categories, we “quantify” teachers' interactions (episodes) with girls and boys during these classes.

Table 3 Average 2012 SIMCE scores of fourth grade schools in the sample and in all fourth grades in Chile

Subject area	Sample schools			All schools		
	Average score	Boys – girls	<i>p</i> value	Average score	Boys – girls	<i>p</i> value
Language arts	249.250	– 9.959	0.000	267.928	– 11.029	0.000
Math	242.451	5.803	0.000	262.115	2.700	0.000
Social sciences	237.589	4.605	0.000	259.067	3.136	0.000
Overall	242.745	– 0.216	0.865	262.671	– 1.922	0.000

Source: Schools' administrative data and results of the Sistema de Medición de la Calidad de la Educación (SIMCE)

Note: Sample schools refer to the schools included in this paper. All schools refer to the universe of Chilean schools that administer the SIMCE standardized national test. The column labeled boys – girls reports the difference in means of test scores between boys and girls. The column labeled *p* value reports the *p* value of the test of differences in means between genders

Table 4 Descriptive statistics of teachers in the sample

Item	Frequency	Percentage
Gender		
Female	161	92.5
Male	9	5.2
Not available	4	2.3
Type of institution from which teacher received degree		
Professional institution	15	8.6
University	156	89.7
Regular school	3	1.7
Specialization/post graduate studies		
Yes	102	59.3
No	69	40.7
Teaching experience (years)		
Less than 5	43	24.7
5–10	40	23
11–24	46	26.4
More than 25	45	25.9
Tenure (years)		
Less than 3	43	24.7
3–4	41	23.6
5–13	46	26.4
14 or more	44	25.3

Source: Schools' administrative data

In addition, for each class, we track the amount of time teachers spend with girls versus boys, count the number of interactions by gender, and analyze teachers' responses (distinguishing positive responses from negative ones). We then correlate these data with students' scores on the SIMCE and with teachers' performance (as measured by CLASS).

We refine and complement the conceptual framework of Sadker and Sadker with other dimensions based on the literature on gender bias in the classroom, including time spent with girls versus boys and the level of control a teacher has over the classroom. The coding scheme used in this paper consists of seven variables. The first four are from Sadker and Sadker (1994):

1. *Praise*. Praise can be provided after correct answers are given, unsolicited, or general. Examples include "Good job," "That was an excellent paper," and "I like the way you're thinking."
2. *Criticism*. Criticism includes negative comments or discipline that provides an explicit statement that the work done or a particular behavior is not correct. Examples include "No, you've missed number four" and "This is a terrible report."
3. *Remediation*. Remediation involves helping a student, encouraging him or her to correct a wrong answer, or expanding and enhancing his or her thinking. Examples include "Check your addition" and "Think about what you've just said and try again."

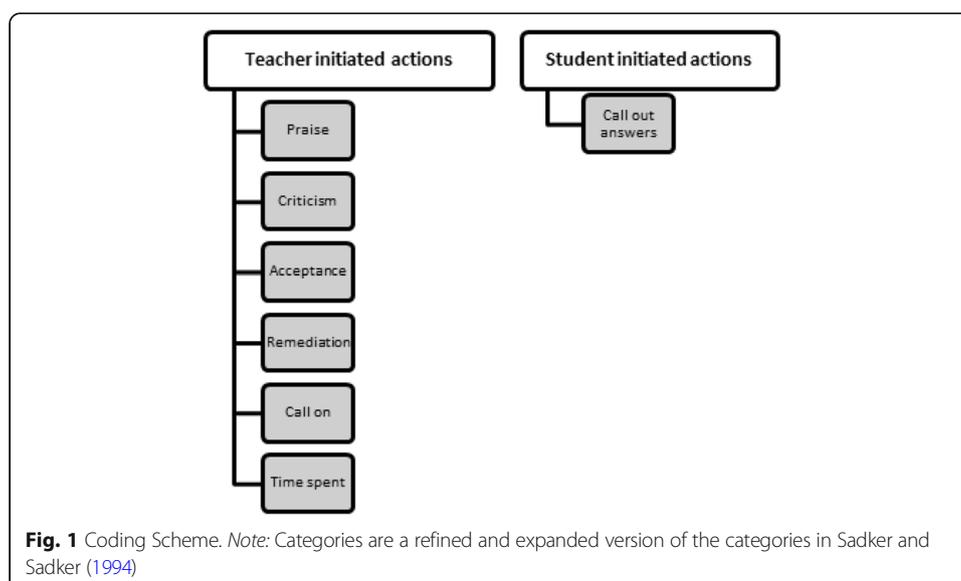
4. *Acceptance*. Acceptance is an acknowledgement of a correct answer given when a student is called on, when the student calls out, or during quiet work. Examples include “Uh-huh” and “Okay.”
5. *Call on*. Call on occurs when a teacher calls a student by name and asks him or her to answer a question, speak to the class, or participate.
6. *Time spent*. Time spent refers to time spent with individual students or groups of students of a single gender (divided into time segments).⁶
7. *Call out*. Call out occurs when a student shouts out an answer when not called on by the teacher.

This study distinguished between classroom behaviors initiated by the teacher and classroom behaviors initiated by the student (Fig. 1). The coding sheets also include a final section where coders can indicate their perception of the teacher’s level of control in the classroom (on a 1–3 scale, from poor control to good control) as well as the teacher’s level of gender bias (on a 1–3 scale, from no obvious gender bias to significant/obvious gender bias).

Except for time spent (the time teachers spend with girls versus boys), coding consists of counting specific events that take place during a video segment and classifying them under a specific category (e.g., praise, criticism, remediation). In this study, coders counted events only when they clearly observed them in the video.

We used a formative method, starting the coding with a random pilot sample of 10% (19 videos). We took this approach not only as a test of the coding variables but also to search for other dimensions that merited coding in the full sample. Based on the results of the pilot sample, we added variables to capture teacher control of the classroom and disaggregated subjects into math and language arts/other subjects, based on the “embedded” nature of some of the language instruction and the similarity of non-math instruction.

The language arts classes were not the only classes that focused on language; grammar and other aspects of language instruction were also incorporated into subjects



such as social studies. Every tape contained some language instruction, and all subjects except math used similar pedagogical techniques, which included reading out loud by the teacher, individual students, and often the whole class in unison. Math was taught in a different way, with teachers using considerable board work and never reading out loud.

Five carefully trained coders (University of Virginia college students, four women and one man) coded 237 videotapes between October 2014 and November 2015. This coder team was different from the CLASS coding team (described in [Appendix](#)). Although coders received thorough training, as well as a template to ensure that they applied the same criteria, video coding involves a subjective component that could result in measurement error. To address this potential problem, we had two coders independently code 41% of the videos. The intercoder reliability rate was 90.7%,⁷ a good result compared with other studies.⁸

5 Results

5.1 Gender-biased variables

The videotapes were not recorded with any intent to study teachers' gender bias in the classroom. It is thus highly unlikely that teachers modified their behavior with respect to this variable because they knew they were being videotaped.

All of the variables studied show a bias in favor of boys (Table 5). In math, differences are statistically significant in criticism, acceptance, call on, and call out. In other subjects (mainly language arts), differences for all variables except praise are statistically significant.

Table 5 Teacher behaviors in interactions with boys and girls, by subject

Subject area/ variable	Obs.	Boys		Girls		T test	
		Mean	Std. dev.	Mean	Std. dev.	Boys – girls	p value
Math							
Praise	69	2.181	2.662	1.558	2.366	0.623	0.148
Criticism	69	6.152	5.731	2.261	2.896	3.891	0.000
Remediation	69	5.319	7.239	3.71	5.047	1.609	0.132
Acceptance	69	12.529	14.294	7.659	5.209	4.87	0.009
Call on	69	11.949	10.19	8.333	6.72	3.616	0.015
Time spent	69	1.833	2.516	1.469	2.176	0.364	0.364
Call out	69	14.442	22.688	5.819	7.197	8.623	0.003
Other							
Praise	168	2.833	3.931	2.414	3.144	0.42	0.281
Criticism	168	5.739	6.368	2.459	4.226	3.28	0.000
Remediation	168	4.988	6.649	3.798	5.964	1.19	0.085
Acceptance	168	14.603	11.629	11.38	8.82	3.223	0.004
Call on	168	17.755	14.112	13.105	11.836	4.65	0.001
Time spent	167	1.702	2.37	1.199	1.982	0.502	0.036
Call out	168	13.846	13.362	7.063	7.487	6.784	0.000

Source: Primary data

Note: T test refers to the result of a test of differences in means of the corresponding variable between boys and girls. The column labeled boys – girls reports the difference in means between boys and girls. The column labeled p value reports the p value of the test of differences in means between genders

Although we did not document which students were actually present in the classroom on the day of the taping, we did document class enrollment by gender. As gaps may partly reflect the fact that there were a few more boys than girls in the classrooms observed, we repeated the measures controlling for the number of students of both genders.

The results show that teachers' bias in favor of boys seems to be robust to the adjustment for the number of boys in the classroom (Table 6). Except for praise in non-math subjects, all gaps remain positive (in favor of boys). Differences in criticism and call out are still statistically significant; call on, acceptance, and time spent are no longer statistically significant. In math, gaps in criticism, acceptance, and call out remain statistically significant.

This analysis tested only for differences in the mean of the measures between boys and girls. We also evaluated the equality of the entire distribution of the measures across gender by graphical inspection and by performing the Kolmogorov-Smirnov test of equality of distributions.

Using the variables described above, we constructed two measures. For all teacher-initiated actions (TIAs) (the four Sadker and Sadker variables, call on, and time spent), we take the first principal component of each factor by subject and plot its distribution for boys and girls. We also look at call out, a student-initiated action (SIA).

The last column in Table 5 shows the correlation of the different factors of the TIA measure with the standardized first principal component. Most factors are strongly correlated with the principal component. The first principal component of teacher

Table 6 Teacher behaviors in interactions with boys and girls after adjusting for number of boys and girls in the classroom, by subject

Subject area/variable	Obs.	Boys		Girls		T test		Correlation with first principal component
		Mean	Std. dev.	Mean	Std. Dev.	Boys – girls	p value	
Math								
Praise	63	0.175	0.236	0.163	0.299	0.013	0.792	0.740
Criticism	63	0.469	0.461	0.191	0.271	0.278	0.000	0.570
Remediation	63	0.467	0.785	0.338	0.557	0.129	0.291	0.749
Acceptance	63	1.031	1.023	0.729	0.642	0.303	0.049	0.637
Call on	63	0.935	0.799	0.776	0.697	0.159	0.236	0.774
Time spent	63	0.161	0.279	0.132	0.210	0.029	0.505	0.682
Call out	63	1.167	1.607	0.560	0.873	0.607	0.010	
Other								
Praise	139	0.213	0.293	0.226	0.330	-0.013	0.729	0.656
Criticism	139	0.444	0.541	0.175	0.276	0.269	0.000	0.597
Remediation	139	0.390	0.568	0.334	0.554	0.057	0.399	0.692
Acceptance	139	1.152	1.067	0.980	0.881	0.172	0.143	0.675
Call on	139	1.349	1.297	1.114	1.122	0.235	0.107	0.832
Time spent	139	0.127	0.206	0.102	0.217	0.025	0.327	0.702
Call out	139	1.128	1.198	0.613	0.733	0.515	0.000	

Source: Primary data

Note: T test refers to the result of a test of differences in means of the corresponding variable between boys and girls. The column labeled boys – girls reports the difference in means between boys and girls. The column labeled p value reports the p value of the test of differences in means between genders

behavior in math classes mainly captures call on, praise, and remediation; in other subjects, it mainly captures call on, time spent, and remediation. For the two types of actions, we plot the distribution for girls and boys separately for math and language arts/other subjects. We report the p value of the test of equality of the principal component across gender.

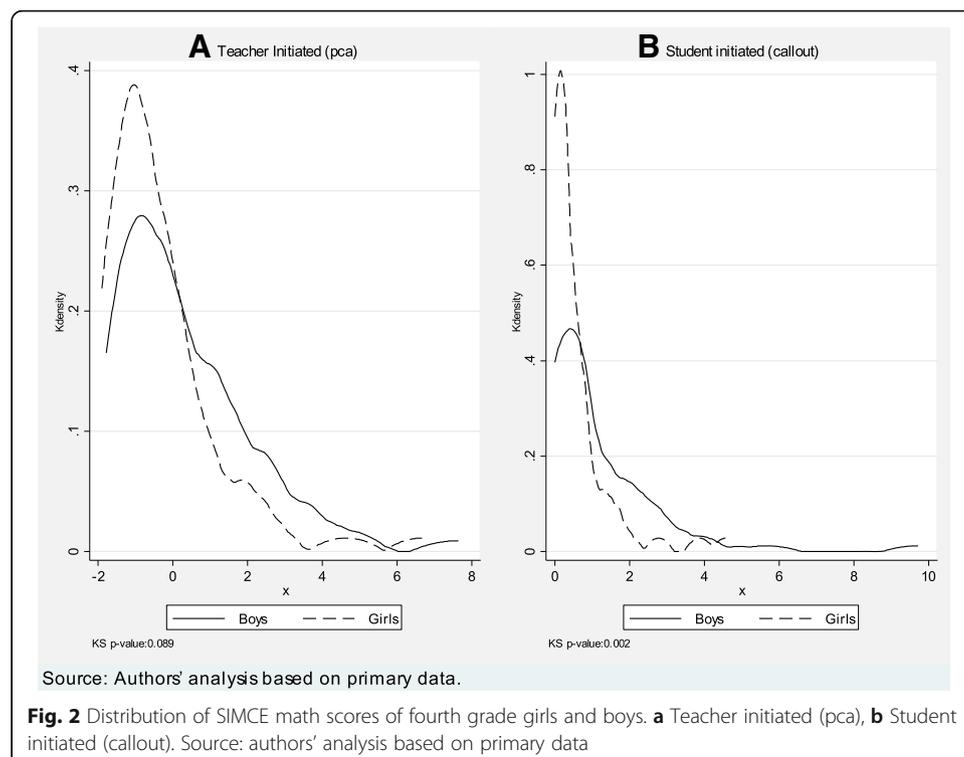
Figure 2 shows the distributions of the TIA first principal component and the SIA for girls and boys in math (both measures control for the number of students in the class). Panel a presents the TIA first principal component; panel b illustrates SIA (call out). Both types of actions show statistically significant differences in the distributions for boys and girls, with the girls' distribution skewed to the left. There are almost no cases of high attention to girls (call ons) and few or no cases of high participation (call outs) of girls.

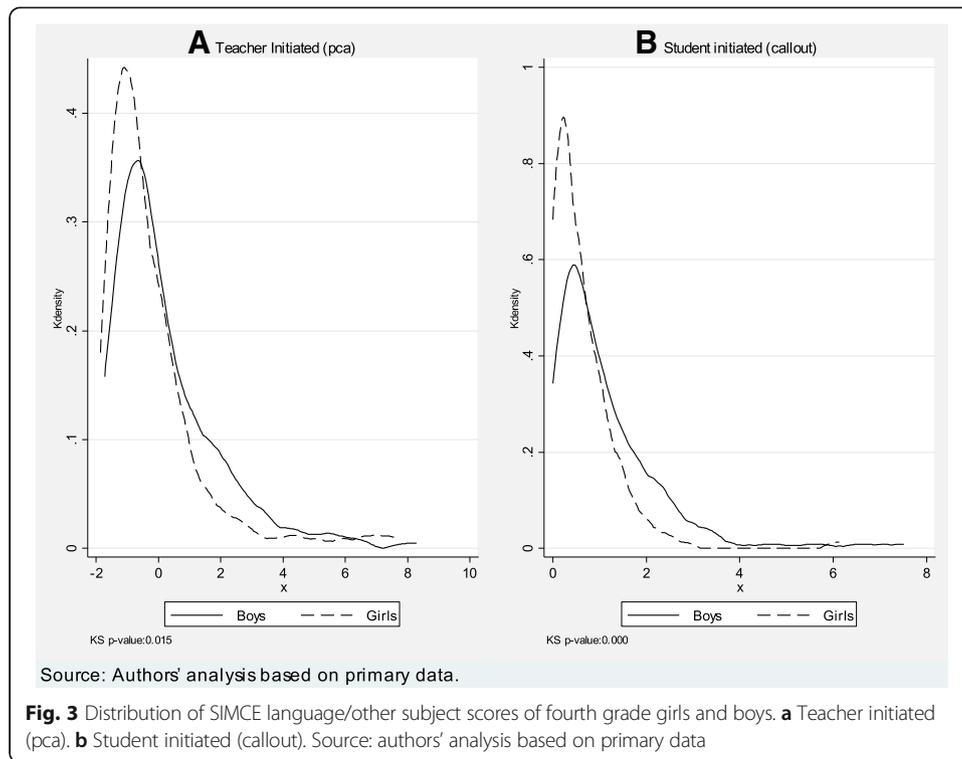
Figure 3 repeats the exercise for other subjects (mainly language arts). It shows consistent differences between the distributions by gender. Gender differences in teachers' negative responses to SIAs are also significant.

Based on tests of equality of distributions of the TIA and SIA measures for boys and girls, we reject the null hypothesis that teachers interact with girls and boys in the same manner.

5.2 Correlation between gender variables and SIMCE and CLASS scores

This section investigates (a) the association between the observed measure of teachers' gender bias and observed measures of other dimensions of teacher–student interactions and (b) how gender bias in the classroom relates to students' performance on the SIMCE. To do so, we create an indicator for gender bias in TIAs and SIAs. The





indicator TIA is the difference between the first principal components of TIAs for boys and girls. The indicator SIA is the differences between boys and girls in calling out answers.

5.2.1 Gender bias and CLASS scores

We coded the same 237 videotapes using the CLASS instrument for fourth to sixth grades (Pianta et al. 2008b). The details of the CLASS coding scheme are described in the Appendix. CLASS measures the quality of teacher–student interactions in three main domains: emotional support, classroom organization, and instructional support. Coders scored these interactions on a scale of 1–7 (Table 7). The average CLASS score for the teachers in our study was 3.95. Only one teacher scored above 5, and no teacher scored less than 2. Scores tended to be higher in class organization and lower in instructional support.

Table 7 CLASS scores of teachers in the sample

Score	Combined		Emotional support		Class organization		Instructional support	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
Low (less than 2)	0	0	0	0	0	0	35	18.5
Medium (2–5)	188	99.5	187	98.9	10	5.3	154	81.5
High (more than 5)	1	0.5	2	1.1	179	94.7	0	0
Average	3.95		3.60		5.93		2.31	

Source: Data from the Classroom Assessment Scoring System (CLASS)

Table 8 presents the results from estimating a regression of the gender bias indicators on the CLASS score (the first principal component of CLASS dimensions) and the disaggregated CLASS dimensions. All regressions include the following covariates:

- The ratio of boys to girls in the classroom
- Characteristics of the school (type of administration, income decile, and experience and tenure of the school principal)
- Characteristics of the classroom teacher (total experience, tenure in the school, and whether the teacher has or is pursuing a graduate degree).

In columns (3) and (4), we include school fixed effects, to control for any permanent differences in school idiosyncrasies in gender attitudes. We weight observations by class size (the results are very similar in unweighted specifications). In the specifications in which we do not control for school fixed effects for the two measures (TIA and SIA), the average CLASS score is negatively correlated with gender bias (worse teachers have higher gender bias).

The results are similar for the domains of the CLASS score:

- Better class organization is associated with less gender bias.
- The correlation with the emotional support score is negative for both TIA and SIA.
- Correlations for instructional support are not statistically significant. However, when school fixed effects are added, the association between the quality of teacher–student

Table 8 Correlation between CLASS (pca) and gender bias, as measured by teacher-initiated and student-initiated actions

Variable	(1)	(2)	(3)	(4)
	TIA	SIA	TIA	SIA
Class scores	−0.080** (0.036)	−0.097** (0.049)	−0.039 (0.119)	0.112* (0.064)
<i>R</i> squared	0.206	0.107	0.837	0.906
Class organization	−0.405** (0.177)	−0.468** (0.236)	0.233 (0.518)	0.582* (0.298)
<i>R</i> squared	0.121	0.102	0.832	0.904
Emotional support	−0.404*** (0.155)	−0.402** (0.193)	−0.120 (0.368)	0.322 (0.337)
<i>R</i> squared	0.128	0.097	0.831	0.900
Instructional support	−0.282 (0.208)	−0.375 (0.232)	0.156 (0.658)	0.668 (0.424)
<i>R</i> squared	0.100	0.075	0.831	0.904
School fixed effects	No	No	Yes	Yes
Observations	186	186	186	186

Source: Primary data and from the Classroom Assessment Scoring System (CLASS)

Note: Robust standard errors are in parentheses. All specifications include the following covariates: ratio of boys to girls in the classroom, characteristics of students (family income, school performance, household structure, and parents' education), characteristics of the school (type of administration, income decile, and experience and tenure of school principal), and characteristics of the classroom teacher (total experience, tenure at the school, and whether she or he has or is pursuing a graduate degree). Models in columns (3) and (4) include school fixed effects

pca principal component analysis, TIA teacher-initiated action, SIA student-initiated action

Significance level: * = 10%, ** = 5%, *** = 1%

interactions and the TIA indicator for gender bias disappears. For the SIA indicators in classes in which the quality of teacher–student interactions is better according to the CLASS measurements, there is more active participation of boys relative to girls. This finding may suggest that better interaction with students empowers boys at the expense of girls in the classroom.

- Overall, the results without fixed effects indicate an inverse correlation between measures of the quality of teacher–student interactions and various measures of gender bias, especially emotional support and class organization. The relationship between the quality of interactions and gender biases in teachers' attention are no longer significant in the fixed effects model. The results of the fixed effects model are driven by the 103 schools in the sample (47%) that have more than one fourth grade classroom. The results in this model are more consistent with the findings of the study of Ecuador by Carneiro et al. (2017). Although the comparison across schools suggests that higher CLASS scores are associated with less gender bias, the within-school variation seems to suggest that after controlling for permanent school characteristics, better CLASS scores are associated with significantly higher participation rates of boys relative to girls. Whether the measures of the quality of interactions are capturing all relevant dimensions of students' well-being within the classroom remains unclear.

5.2.2 SIMCE scores and gender bias

Paying more attention to some students could be associated with better learning outcomes. Our results cannot be interpreted as causal effects, however. Students were not randomly allocated to teachers, so the data do not allow other possible effects of unobserved variables to be isolated. For example, the best students (both girls and boys) may have more engaged parents, who get involved with the school to make sure their children get the best teachers, who may also be more capable of providing equal support to all students. The analysis below analyzes whether teacher gender bias is correlated with test scores.

To do so, we use the SIMCE, which fourth graders take every year. Table 9 presents the results of simple ordinary least squares regressions of SIMCE scores by discipline (reading, math, and science) for girls and boys. As explanatory variables, we use the two measures of gender bias (TIA and SIA). In all specifications, we include other covariates (to control for student, school, and teacher characteristics) as well as the CLASS score (to control for the quality of teacher–student interactions). All regressions control for the ratio of boys to girls in the classroom and school fixed effects. Along with the coefficients, we include the normalized coefficient that translates effects into percentages of a standard deviation of the test score for the corresponding gender group.

In the case of TIA, the results show significant negative correlations between gender biases and the test scores of girls in math and science: Girls whose teachers demonstrate greater gender bias have lower tests scores in these subjects. The effects are also big, particularly for math. For every 1 standard deviation increase in TIA, girls' scores on the SIMCE decrease by 34.6% of a standard deviation in math and 5% of a standard deviation in science, after controlling for student, school, and teacher characteristics.

Table 9 Correlation between gender bias and SIMCE scores

Variable	Reading		Math		Science	
	(1)	(2)	(3)	(4)	(5)	(6)
	Boys	Girls	Boys	Girls	Boys	Girls
Teacher-initiated action						
Coefficient	-7.286	-5.493	-23.255	-16.379*	-0.149	-0.234
Standard error	(5.037)	(5.505)	(16.619)	(9.812)	(4.458)	(4.526)
Normalized coeff.	-0.133	-0.105	-0.509	-0.346	-0.003	-0.005
R squared	0.250	0.287	0.315	0.394	0.284	0.301
Student-initiated action						
Coefficient	-5.859	-10.363**	7.006**	-9.538**	4.051	-4.989
Standard error	(4.774)	(5.216)	(3.459)	(4.331)	(4.649)	(4.537)
Normalized coeff.	-0.081	-0.150	0.200	-0.181	0.061	-0.080
R squared	0.250	0.289	0.315	0.394	0.285	0.302
Observations	1299	1198	577	560	1288	1188

Source: Primary data and data from the Sistema de Medición de la Calidad de la Educación (SIMCE)

Note: Normalized coeff. reports the association between gender bias and SIMCE scores in units of a standard deviation of the test score of the subject for the corresponding gender. Robust standard errors are in parentheses. All specifications include covariates: the ratio of boys to girls in the classroom, characteristics of students (family income, school performance, household structure, and parents' education), characteristics of the school (type of administration, income decile, and experience and tenure the school of principal), and characteristics of the classroom teacher (total experience, tenure in the school, and whether he or she has or is pursuing a graduate degree) and score on the Classroom Assessment Scoring System (CLASS). All specifications include school fixed effects

Significance level: * = 10%, ** = 5%, *** = 1%

SIA (call out) is significantly and negatively correlated with girls' test scores in reading and math. Coefficients suggest a larger impact on reading scores than TIA. A 1 standard deviation increase in SIA is associated with a decrease of 18% of a test score standard deviation in math and of 15% of a test score standard deviation in science. Girls perform significantly better in math relative to boys in classroom with higher gender equality in SIA.

Table 10 shows the association between gender bias and SIMCE scores by the socioeconomic status of the school. The negative correlation between gender bias and girls' learning is stronger in low socioeconomic status environments, and the positive correlation of gender bias on learning for boys is magnified in high socioeconomic status schools.⁹

In sum, the results suggest a correlation between gender bias (both from teachers and in terms of student behavior) and test scores. The magnitudes are large: the variation in SIMCE scores that is associated to the variation in gender bias explains 5–35% of the total variation in SIMCE scores. Our data do not permit claims to be made about causality, but the results are consistent with arguments well documented in the literature that point to the importance of teacher–student interaction for learning.

6 Conclusions

This study finds differences in the amount and type of attention teachers devote to girls and boys in the classroom in Chile. The fact that clear patterns already appear in fourth grade students is important, given that gender gaps in test scores tend to increase as students progress through school grades (Bharadwaj et al. 2015). As established in the education literature, receiving more teacher attention (positive or negative) may affect

Table 10 Correlation between gender bias and SIMCE scores, by socioeconomic status of the school

Variable	Reading		Math		Science	
	(1)	(2)	(3)	(4)	(5)	(6)
	Boys	Girls	Boys	Girls	Boys	Girls
Low socioeconomic status						
Teacher-initiated action						
Coefficient	-7.804	-2.007	-7.006	-32.140**	-4.079	2.088
Standard error	(5.263)	(5.795)	(9.875)	(12.709)	(4.520)	(4.766)
Normalized coeff.	-0.153	-0.040	-0.155	-0.678	-0.087	-0.046
R squared	0.262	0.297	0.302	0.409	0.284	0.308
Student-initiated action						
Coefficient	-7.322	-6.735	4.521	-8.381**	-1.118	-2.140
Standard error	(5.630)	(5.981)	(3.034)	(3.611)	(5.203)	(5.316)
Normalized coeff.	-0.109	-0.103	0.137	-0.166	-0.018	-0.036
R squared	0.262	0.298	0.302	0.409	0.284	0.308
Observations	1035	972	497	476	1028	963
High socioeconomic status						
Teacher-initiated action						
Coefficient	5.981	-8.382			14.671***	-6.990
Standard error	(5.759)	(5.297)			(5.316)	(4.810)
Normalized coeff.	0.069	-0.131			0.185	-0.119
R squared	0.200	0.179			0.277	0.234
Student-initiated action						
Coefficient	2.915	-12.558**			16.293**	-7.508
Standard error	(6.976)	(6.253)			(6.515)	(6.071)
Normalized coeff.	0.027	0.143			0.163	-0.094
R squared	0.197	0.184			0.272	0.233
Observations	264	226			260	225

Source: Primary data and data from the Sistema de Medición de la Calidad de la Educación (SIMCE)

Note: Normalized coeff. reports the association between gender bias and SIMCE scores in units of a standard deviation of the test score of the subject for the corresponding gender. Robust standard errors are in parentheses. All specifications include covariates: the ratio of boys to girls in the classroom, characteristics of students (family income, school performance, household structure, and parents' education), characteristics of the school (type of administration, income decile, and experience and tenure the school of principal), and characteristics of the classroom teacher (total experience, tenure in the school, and whether he or she has or is pursuing a graduate degree) and the Classroom Assessment Scoring System (CLASS) score. All specifications include school fixed effects. The association between student- and teacher-initiated actions and math scores cannot be estimated in the model with the full set of controls and fixed effects because of insufficient observations

Significance level: * = 10%, ** = 5%, *** = 1%

motivation, aspirations, and performance, as well as long-term outcomes, such as decisions about college, employment possibilities, and earnings.

Biases are often unconscious. They are based on myths and beliefs that are not necessarily grounded in evidence or even direct experience. For example, there is a general perception that girls talk more in class than boys. In one of their studies, Sadker and Sadker (1985) show a film of a classroom discussion and ask teachers and administrators which gender talked more. Although the quantitative data showed that boys talked three times as much, the majority of teachers claimed that girls talked more than boys.¹⁰

The fact that attention biases by gender are particularly related to math scores raises the question of why girls do better in language. An important factor behind these results could be that the process of learning is different in math than in other subjects. Reading can be improved by continued reading (something students can do on their own), whereas mathematical thinking requires “engaging students in posing and solving problems” (Fite 2002). The fact that certain subjects are more teacher-dependent than others could explain part of these differences.

Teachers’ use of instructional time, the quality of instructional support, the use of materials (including information and communications technology), the quality of classroom management, the ability to keep students engaged, and the emotional support provided to students seem to be important factors in learning. Making sure that all students benefit from good teaching practices is as important as how frequently good practices are used.

In terms of stimulating broad participation of all students in the classroom, certain pedagogical techniques—such as calling on the whole classroom, asking for quick answers or rushing students, and accepting call outs from students even if they have not respected their turn by raising their hands—seem to be especially problematic for girls. Raising teachers’ awareness of the gender biases implicit in these practices should be combined with a review of the extent of gender bias in textbooks.

Whether they favor boys or girls, gender gaps hinder the possibility of developing all students’ full potential. Educational policy should take into account impacts on and unintended consequences for both groups. Further research on why teachers pay more attention to boys than girls would help policy makers craft specific interventions to generate school environments more prone to learning for all students, regardless of their background.

Endnotes

¹The OECD launched the PISA in 1997, with the objective of developing regular, reliable, and policy-relevant indicators on student achievement. PISA 2015 is the sixth survey. It assesses the competencies of 15-year-olds in reading, math, and science (with a focus on science) in 72 economies. For more information, see <http://www.oecd.org/pisa/>.

²The teachers’ rationale was that the boys tended to be more demanding and that their tones and attitudes obliged teachers to respond.

³The Education Quality Agency in Chile, responsible for administration of the test, has been adding new disciplines and grades (2nd, 6th, 10th, and 12th grades) for representative samples of students and with lower frequency (every 2, 3, or 4 years).

⁴Some videos included classes for more than one subject. In those cases, each subject was coded separately.

⁵For example, Araujo et al. (2016) briefly review this literature for the USA and present a study for children in kindergarten in Ecuador.

⁶The video viewer aggregated the codes for time spent by the teacher with girls versus boys.

⁷The intercoder reliability rate assumes that two codings do not coincide if the difference in their scores differs by more than 1 standard deviation (by category). When this criterion is modified to 0.5 standard deviations, the intercoder reliability rate remains above 80%.

⁸Brown et al. (2010) report an intercoder reliability rate of 83%; Araujo et al. (2016) report a rate of 93%.

⁹The association between SIA and TIA and math scores cannot be estimated in the model with the full set of controls and fixed effects because of insufficient observations.

¹⁰LaFrance and Mayo (1978) find that males talk more in everyday life as well.

Appendix

Classroom Assessment Scoring System (CLASS)

CLASS is a classroom observational instrument developed by researchers at the Curry School Center for Advanced Study of Teaching and Learning at the University of Virginia. It was originally developed as a tool for research on early childhood development. It now includes versions for lower-elementary, upper-elementary, and secondary school.

CLASS focuses on interactions between teachers and students as the primary mechanism for student learning (Kane and Staiger 2012). It measures teachers' behaviors in three domains: emotional support, classroom organization, and instructional support. Each of the three domains includes several dimensions, listed and described in Table 11.

Trained coders look for specific behaviors of teachers in each dimension. Scores are assigned on a 7-point scale according to detailed descriptions provided by the CLASS protocol of what is expected at "low" (scores 1 and 2), "medium" (scores 3, 4, and 5), and "high" (scores 6 and 7) levels.

Araujo et al. (2016) (citing Pianta et al 2008a) present a table with examples of CLASS scores for behavior management dimension (in the emotional support domain). Coders have to analyze four behaviors in this dimension. One of them is setting clear behavior expectations. A low score is assigned when rules and expectations are absent, unclear, or inconsistently enforced. A middle score is assigned when rules and expectations may be stated clearly but are inconsistently enforced. A high score is assigned when rules and expectations for behavior are clear and consistently enforced.

Another example within the behavior management dimension is teacher's proactivity. A low score in proactivity is assigned when the teacher is reactive and monitoring is absent or ineffective. A middle score is assigned when the teacher uses a mix of proactive and reactive responses, sometimes monitors but other times misses early indicators of problems. A high score is assigned when the teacher is consistently proactive and monitors effectively to prevent problems from developing.

This paper uses videos and CLASS scores coded for another project (Bassi et al. 2016). All fourth grade teachers in the sample were filmed for four school hours (45 min each). According to the CLASS protocol, the first hour of the videotape is discarded, because teachers and students are likely to be aware of the camera and may modify their behaviors. The rest of the videos are cut into 15-min segments; only one segment per hour is coded.

After 4 days of training, 11 coders were certified according to CLASS protocols in the upper-elementary version of the instrument and participated in the coding of the

Table 11 CLASS dimensions by domain

Domain	Domain/ dimension	Description
Emotional support	Positive climate	Reflects overall emotional tone of classroom and connection between teachers and students
	Negative climate	Reflects overall level of expressed negativity in classroom between teachers and students (e.g., anger, aggression, irritability)
	Teacher sensitivity	Encompasses teachers' responsivity to students' needs and awareness of students' level of academic and emotional functioning
Classroom management	Behavior management	Encompasses teachers' ability to use effective methods to prevent and redirect misbehavior by presenting clear behavioral expectations and minimizing time spent on behavioral issues
	Productivity	Considers how well teachers manage instructional time and routines so that students have maximum number of opportunity to learn
	Instructional learning formats	Reflects degree to which teachers maximize students' engagement and ability to learn by providing interesting activities, instruction, centers, and materials
Instructional support	Concept development	Reflects degree to which instructional discussions and activities promote students' higher-order thinking skills versus focus on rote and fact-based learning
	Quality of feedback	Considers teachers' provision of feedback focused on expanding learning and understanding (formative evaluation), not correctness or the end product (summative evaluation)
	Language modeling	Reflects the quality and amount of teachers' use of language-stimulation and language-facilitation techniques during individual, small-group, and large-group interactions with children.

Source: Pianta et al. 2008a, 2008b

videos, which took place between November 30, 2012, and January 3, 2013. Double coding was applied to 52% of the segments, decreasing gradually from the first week of coding to the last (starting at 100% double coding in the first week), as good levels of reliability were confirmed. The intercoder reliability rate was 84.2%, which is comparable to other studies using CLASS.

Acknowledgements

We are grateful for the assistance and guidance of many people. Gloria Burgoa, Olivia Feldman, Daniel Alonso, and Laura Becerra provided outstanding assistance for the literature review and analysis of the data. At the University of Virginia, Barbara Selby helped us find helpful sources, and Nancy Kechner, an information technology expert, helped us solve challenging problems with the coding of the videos. We are also grateful to David Sadker, Andy Morrison, and Norbert Shady for their insights; to the editor and anonymous peer reviewers; and to participants at the Asia-Pacific Economic Cooperation (APEC) Annual Meeting and a seminar on gender and education organized by the Chilean Ministry of Education and Ministry of Finance, both held in Santiago, Chile. All remaining errors are our own. Responsible editor: Pierre Cahuc

Funding

Source of funding: Inter-American Development Bank. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

Competing interests

The IZA Journal of Labor Economics is committed to the IZA Guiding Principles of Research Integrity. The authors declare that they have observed these principles.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹World Bank Group, Av. Kenneth Kaunda 1224, Maputo, Mozambique. ²Inter-American Development Bank, 1300 New York Avenue, NW, Washington, DC, USA. ³University of Virginia, 102 Randall Hall, Charlottesville, VA, USA. ⁴University of Michigan, 611 Tappan Avenue, Ann Arbor, MI, USA.

Received: 28 February 2018 Accepted: 24 October 2018

Published online: 13 November 2018

References

- AAUW (American Association of University Women) (1992) How schools shortchange girls. AAUW, Annapolis Junction
- Araujo MC, Carneiro P, Cruz-Aguayo Y, Schady N (2016) Teacher quality and learning outcomes in kindergarten. In: IDB Working Paper Series No. IDB-WP-665. Inter-American Development Bank, Washington, DC
- Autor D, Figlio D, Karbownik K, Roth J, Wasserman M (2016) Family disadvantage and the gender gap in behavioral and educational outcomes. In: NBER Working Paper 22267. National Bureau of Economic Research, Cambridge
- Banjong DN (2014) Same performance but different perception: female stereotypes in mathematics emerge in fifth grade. *Int Online J Educ Sci* 6(2):258–268
- Bassi M, Busso M, Muñoz JS (2015) Enrollment, graduation, and dropout rates in Latin America: is the glass half empty or half full? *Econ J Lat Am Caribbean Econ Assoc* 16(1):113–156
- Bassi M, Meghir C, Reynoso A (2016) Education quality and teaching practices. In: NBER working paper 22719. National Bureau of Economic Research, Cambridge
- Bertrand M, Pan J (2013) The trouble with boys: social influences and the gender gap in disruptive behavior. *Am Econ J Appl Econ* 5(1):32–64
- Bharadwaj P, De Giorgi G, Hansen D, Neilson C (2015) The gender gap in mathematics: evidence from a middle-income country. In: Staff report 721. Federal Reserve Bank of New York, New York
- Bos MS, Ganimian A, Vegas E (2014) América Latina en PISA 2012: Cómo se desempeñan los varones y las mujeres? Inter-American Development Bank, Washington, DC https://publications.iadb.org/handle/11319/6355?scope=123456789/4&thumbnail=true&hpp=5&page=0&group_by=none&etal=0
- Brown JL, Jones SM, LaRusso MD, Aber JL (2010) Improving Classroom Quality: Teacher Influences and Experimental Impacts of the 4Rs Program. *J Educ Psychol*. 102(1):153–167. <http://www.morningsidecenter.org/sites/default/files/documents-pdfs/JournalEdPsych2010.pdf>.
- Bruns B, De Gregorio S, Taut S (2016) Measures of effective teaching in developing countries. In: Research on Improving Systems of Education (RISE) Working Paper 16(009)
- Bruns B, Luque J (2015) Great teachers: how to raise student learning in Latin America and the Caribbean. In: Working paper 89514. World Bank, Washington, DC <http://documents.worldbank.org/curated/en/2014/01/19798994/great-teachers-raise-student-learning-latin-america-caribbean>
- Carneiro PM, Cruz-Aguayo Y, Schady NR (2017) Where the girls are not: households, teachers, and the gender gap in early math achievement. In: IDB working paper IDB-WP-807. Inter-American Development Bank, Washington, DC. <https://doi.org/10.18235/0000700>
- Chetty R, Friedman JN, Rockoff JE (2011) The long-term impacts of teachers: teacher value-added and student outcomes in adulthood. In: NBER working paper no. 17699. National Bureau of Economic Research, Cambridge <http://www.nber.org/papers/w17699>
- Chetty R, Hendren N, Lin F, Majerovitz J, Scuderi B (2016) Childhood environment and gender gaps in adulthood. *Am Econ Rev* 106(5):282–288
- Consuegra E, Engels N, Willegems V (2016) Using video-stimulated recall to investigate teacher awareness of explicit and implicit gendered thoughts on classroom interactions. *Teach Teach* 22(6):683–699
- Davis KL (2000) A case study in gender equity for basic instruction programs. *Phys Educ* 57(1):47
- Erden F, Wolfgang C (2004) An exploration of the differences in teachers' beliefs related to discipline when dealing with male and female students. *Early Child Dev Care* 174(1):3–11
- Fennema E, Peterson PL, Carpenter TP, Lubinski CA (1990) Teachers' attributions and beliefs about girls, boys and mathematics. *Educ Stud Math* 21:55–69
- Fite G (2002) Reading and math: what is the connection? A short review of the literature. *Kans Sci Teach* 14:7–11
- Guiso L, Monte F, Sapienza P, Zingales L (2008) Culture, gender, and math. *Science* 320(May 30):1164–1165 (plus supporting online data)
- Heckman JJ (2011) The Economics of Inequality: The Value of Early Childhood Education. *Am Educ* 35(1):31–47. <https://eric.ed.gov/?id=EJ920516>
- Kane TJ, Staiger DO (2012) Gathering feedback for teaching: combining high-quality observations with student surveys and Achievement Gains. Research Paper. MET Project. Bill & Melinda Gates Foundation. <https://eric.ed.gov/?id=ED540960>
- Kersey AJ, Braham EJ, Csumitta KD, Libertus ME, Cantlon JF (2018) No intrinsic gender differences in children's earliest numerical abilities. *Sci Learn* 3:12
- LaFrance M (1991) School for scandal: different educational experiences for females and males. *Gen Educ* 3(1):1–13
- LaFrance M, Mayo C (1978) Moving bodies: nonverbal communication in social relationships. Brooks/Cole, Monterey
- Lavy V, Sand E (2015) On the origins of gender human capital gaps: short and long term consequences of teachers' stereotypical biases. In: NBER Working Paper 20909. National Bureau of Economic Research, Cambridge
- Li Q (1999) Teachers' beliefs and gender differences in mathematics: a review. *Educ Res* 41(1):63–76
- Measor L, Sykes P (1992) Gender and schools. Cassell, New York
- Mizala A (2014) Brecha de género, desempeño en matemáticas y elección de carreras. In: Proyecto FONDECYT. Fondo Nacional de Desarrollo Científico y Tecnológico, Santiago
- Murnane RJ, Willett JB, Levy F (1995) The growing importance of cognitive skills in wage determination. *Rev Econ Stat* 77: 251–266
- Ñopo H (2012) New century, old disparities: gender and ethnic earnings gaps in Latin America and the Caribbean. World Bank, Washington DC
- OECD (Organisation for Economic Co-operation and Development) (2014) Education at a glance 2014: OECD indicators. OECD Publishing, Paris
- OECD (Organisation for Economic Co-operation and Development) (2015a) The ABC of gender equality in education: aptitude, behaviour, confidence. OECD Publishing, Paris

- OECD (Organisation for Economic Co-operation and Development) (2015b) PISA 2015 Data File. Interactive Selection, Variable STO4Q01, Gender. OECD, Paris <http://www.oecd.org/pisa/>
- Olszewski-Kubilius P, Turner D (2002) Gender differences among elementary school-aged gifted students in achievement, perceptions of ability and subject preference. *J Educ Gift* 25(3):233–268
- Paglin M, Rufolo AM (1990) Heterogeneous human capital, occupational choice, and male-female earnings differences. *J Labor Econ* 8:123–144
- Pianta RC, LaParo KM, Hamre BK (2008a) Classroom assessment scoring system manual: pre-K. Brookes Publishing, Baltimore
- Pianta RC, Mashburn AJ, Downer JT, Hamre BK, Justice L (2008b) Effects of web mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. *Early Child Res Q* 23(4):431–451
- Rivkin SG, Hanushek EA, Kain JF (2005) Teachers, schools, and academic achievement. *Econometrica* 73(2):417–458
- Robinson JP, Lubienski ST (2011) The development of gender achievement gaps in mathematics and reading during elementary and middle school: examining direct cognitive assessments and teacher ratings. *Am Educ Res J* 48(2):268–302
- Robinson JP, Lubienski ST, Copur Y (2011) The effects of teachers' gender-stereotypical expectations on the development of the math gender gap. Society for Research on Educational Effectiveness, Evanston
- Sadker M, Sadker D (1985) Sexism in the schoolroom of the Eighties. Reprinted in *The World*. Reprinted in Annual Editions: Education 1986-87, 87-88, 88-89, (Dushkin Press). *Psychol Today*. <https://www.sadker.org/CV3.html>
- Sadker M, Sadker D (1994) *Failing at fairness: how America's schools cheat girls*. Scribner's, New York
- Sadker M, Sadker D, Stulberg LM (1993) Fair and square: creating a nonsexist classroom. *Instructor* 102(7):44–46 67–68
- Shumow L, Schmidt JA (2013) Academic grades and motivation in high school science classrooms among male and female students: associations with teachers' characteristics, beliefs and practices. *J Educ Res* 7(1):53–71
- Stallings J, Knight S (2003) Using the Stallings Observation System to investigate: time on task in four countries. In: Unpublished paper for the International Time on Task Project. World Bank, Washington, DC
- Straus B (2015) Boys and literacy. In: Research apprenticeship report. University of Virginia, Charlottesville
- Terrier C (2016) Boys lag behind: how teachers' gender biases affect student achievement. In: IZA Discussion 10343. Institute of Labor Economics, Bonn
- Tiedemann J (2002) Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educ Stud Math* 50:49–62
- Van Duzer E (2006) Gender bias in predictions of boys' academic failure. *Action Teach Educ* 28(1):94–98
- WEF (World Economic Forum) (2015) The global gender gap report 2015. World Economic Forum, Geneva

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
