School Discipline across Countries: Theory, Measurement and Effect

Gruber, Noam

1 September 2020

School Discipline across Countries: Theory, Measurement and Effect *

Noam Gruber †

September 1, 2020

Abstract

Using PISA truancy and tardiness data to generate estimates of school discipline comparable across countries, this paper finds a strong relation between both individual and school-level discipline and student performance. Furthermore, the data shows that the effect of discipline grows with class size, so that students in large classes can benefit the most from an atmosphere of discipline. This finding explains how Asian education systems in Japan, Korea, Taiwan, Hong-Kong and Singapore are top performers in international student achievement tests while having exceptionally large classes. It also implies that some Western countries, enjoying high levels of discipline but opting for small classes, are inefficient in the use of their educational resources, leading to sub-optimal results by their students.

Keywords: Education, PISA, International Tests, Discipline, Tardiness, Punctuality, Truancy, Class Size

JEL Classification Codes: I21, I28

*I am grateful to Dan Ben-David, Ayal Kimhi and Ity Shurtz, as well as participants of the Israeli Economic Association's convention, for their helpful comments. Opinions expressed here are solely my own and do not represent the views on the National Economic Council of Israel. All errors are my own.

†National Economic Council of Israel. Email: grubern@pmo.gov.il.

1 Introduction

It is well established that discipline is a major factor in the efficacy of the scholastic process. However, the role of discipline in explaining the differences in scholastic achievement levels between countries has not thus far been explored, mostly due to the substantial cultural bias present in estimating the level of discipline across countries. This paper uses truancy and tardiness data, which are less vulnerable to cultural bias, from the PISA (the OECD's Programme for International Student Assessment) 2012, 2015 and 2018 rounds, to construct a measure of school discipline that is valid across countries. This measure is well-correlated with subjective within-country measurements of discipline, and is shown to strongly correlate to student achievements, both within and between countries.

Attentiveness and discipline have a number of dimensions, including students' ability to listen and concentrate for lengthy periods; their ability to keep quiet, thereby facilitating concentration and a positive learning atmosphere for their classmates; and their self-discipline, which translates into diligence and studiousness. As shall be demonstrated further on, discipline rises in importance as classes grow in size and the number of potential disturbances increases. These empirical results match the theoretical predictions of Lazear (2001).

Numerous studies debate the effect and importance of class size in student achievement, with ambiguous results. However, PISA results show the top 5 performing developed countries – Singapore, Hong-Kong, Taiwan, Korea and Japan – to have exceptionally large classes.[1] As will be shown, these countries are all able to maintain a strong disciplinary climate in their schools. This paper will argue that it is this discipline that enables students in large classes to study and test well, providing school systems in these countries with substantial advantages over their counterparts. Likewise, it is argued that some Western school systems, enjoying high levels of discipline but opting for small classes, are inefficient in the use of their resources, leading to sub-optimal results by their students.

[1] It is worth noting that while China, a top performing country, is not included in this analysis due to its developing country status, it too is characterized by exceptionally large classes.

Section 2 overviews the relevant literature. Section 3 presents a theoretical model. Section 4 describes the PISA data used in this paper. Section 5 analyzes the disciplinary atmosphere as subjectively reported by students. Section 6 introduces an 'objective' measure of discipline, based on truancy and tardiness data. Section 7 contains econometric analysis of the effect of discipline on students' achievements. Section 8 discusses the issue of identification, as well as first-difference estimation results. Section 9 concludes.

2 Literature Review

Many papers have used data from international surveys such as PISA in attempts to identify key elements of the 'education production function' and explain the variance in outcomes across students, schools, education systems and countries, using student, family and school attributes (for a comprehensive review see Hanushek and Woessmann (2011)). This strand of literature has generally not used indicators of discipline levels, likely due to their culturally-biased nature.

While school discipline and its effect on student outcomes has long been explored (for a review of the topic, see Arum and Velez (2012)), it has mostly been studied separately from other factors influencing the educational process. In essence, measurement issues have thus far hampered a more thorough and systematic empirical analysis of the effect of school discipline across classes, schools and countries, with respect to pertinent educational factors such as class size and teaching quality.

The 'class size' literature attempts to explain the apparent paradox of positive correlation between class size and student performance. This vast empirical literature seeks to deal with selection issues and properly identify the effect of class size on the quality of education. Some meta-analysis papers, such as Krueger (2003) find that reducing class size has a positive impact on student performance. Others, such as Hoxby (2000) and Hanushek (2003), find little to no effect. Of two exceptional papers in this literature, Angrist and Lavy (1999), which uses the maximum class size regulation in Israel for identification, and Krueger (1999), which uses random assignment, both find that reducing class size does have a small positive effect on scholastic achievements.

Lazear (2001) attempts to explain the positive correlation between class size and student performance using a theoretical model, arguing that for higher levels of discipline, larger classes may indeed be optimal and lead to better results. Lazear (2001) is seminal in presenting a model which ties student outcomes to class discipline, class size and teaching quality together in one theoretical framework, and exploring the potentially complex co-relations between these factors of education production.

It could be argued that the vast majority of research on the effect of class size is done in a partial equilibrium framework. This literature focuses on overcoming the selection bias causing better students to be assigned to larger classes, and often assumes that changing class sizes has no impact on the overall quality of teaching.[2] Lazear (2001) breaks this mold by modeling the trade-offs between class sizes and teaching quality, and how they depend on class discipline. This paper's contributes to existing economic literature by extending the model in Lazear (2001), developing a measure of discipline that is culturally unbiased, and using it to empirically estimate and establish the links between school discipline, class size and student performance, as well a their role in explaining the cross-country variation in student performance, as measured in the PISA test.

3 Theoretical Analysis

Equation 1, the class-level education production equation, is based on the model presented in Lazear (2001):

$$e = nqp^n \tag{1}$$

[2]There are some exceptions, such as Jepsen and Rivkin (2009).

where e is the class-level educational product, n is the number of students in the class and q is the study quality without interruptions. p is the discipline level in the class, representing the probability that the representative students will study without interrupting within a given time frame. Therefore $0 \leq p \leq 1$ and the probability on uninterrupted study in the class is p^n.

Given Equation 1, E, the country-level educational product, is as follows:

$$E = Nqp^n \tag{2}$$

Where N is the total number of students in the country.

Equation 3 formulates the budget constraint faced by the national school system:

$$C = w\frac{N}{n} \tag{3}$$

Where C is the given education budget, w is the expenditure per class, consisting of the teacher's wage and other factor inputs, and $\frac{N}{n}$ is the total number of classes.

Equation 4 assumes that an increase in w, the expenditure per class, will increase q, the quality of study, *ceteris paribus*.

$$q = w \tag{4}$$

This is a simplification, abstracting away from many factors affecting the quality of study, such as parents education, the effect of the number of class groups on the demand for teachers,[3] and the effect of school discipline on the teacher's profile (the teaching profession may be less attractive if students are unruly).[4]

Maximizing the educational product (Equation 2) under the budget constraint (Equation

[3]If less teachers are required, then the school system can be more selective and hire the more capable ones, as argued in Jepsen and Rivkin (2009).

[4]**?** shows that teachers' decisions on leaving a school or staying in it are strongly influenced by student characteristics in the school.

3) yields n^*, the optimal number of students per class:

$$n^* = -(\ln p)^{-1} \tag{5}$$

Recall that $0 \leq p \leq 1$, so that $-\infty \leq \ln p \leq 0$. Equation 5 implies that n^*, the optimal class size, is increasing in p, the level of discipline. Combining Equation 5 with Equation 3 yields w^*, the optimal expenditure per class:

$$w^* = -\frac{C}{N \ln p} \tag{6}$$

As Equation 6 demonstrates, for a given budget C, a higher level of discipline (p) enables larger classes, which in turn increases expenditure per class (w^*), meaning better teachers and equipment. The effect of discipline can be broken down to two parts: the direct effect is to improve educational outcomes for a given level of inputs, while the indirect effect is to enable larger classes.

Equation 5 yields the following:

$$p^{n^*} = p^{-\frac{1}{\ln p}} = e^{-1} \tag{7}$$

As Equation 7 implies, in this model an increase in discipline (p) does not increase p^{n^*}, but translates completely to an increase in class size and resources per class. Combining Equation 2 and Equation 7 results in Equation 8:

$$E^* = N w^* p^{n^*} = -\frac{C}{e \ln p} \tag{8}$$

The model clearly indicates that a higher level of discipline enables better educational results on a given budget, or alternatively similar results at a lower expense. Since a higher level of discipline can be expected to attract a stronger profile of teachers, and since fewer classes (and fewer teachers needed) would enable schools to be more selective with regard to personnel choice, this model is likely to only partially represent the importance of discipline in the educational process.

This model also demonstrates that education systems that enjoy high levels of school discipline, but choose to have small classes, are sub-optimal in terms of the educational product they produce.

4 PISA Data: 2012, 2015 and 2018 Rounds

The PISA test is administered to 15 year olds[5] every three years.[6] This work uses data from the 2012, 2015 and 2018 rounds. Of the three main subjects covered by PISA - mathematics, literacy and science - math was chosen as the focus. Beyond its intrinsic importance, math is an international language and is therefore uniquely suited for cross-country comparison. The PISA math questions are not meant to test specific knowledge - they do not encompass theoretical material but rather require quantitative thinking.[7] It should be noted that there is a high correlation between countries' math achievements and their attainments in literacy and science, meaning that an education system's success in teaching quantitative thinking skills is a good indicator of success in the teaching of other skills.

The PISA exams were calibrated in 2000 to an OECD country average of 500 and standard deviation of 100. All subsequent rounds were calibrated with respect to the 2000 round using IRT (Item Response Theory) methodology. In addition to the test questions, PISA respondents, as well as school administrators, also fill out survey questionnaires, which provides many details on the students' family background, school attributes, attitude towards math etc. The present study uses data for the 34 OECD member countries,[8] as well as Singapore, Hong-Kong[9] and Taiwan. For these 37 countries there were 311,678 test-takers

[5] The cohort of 15 years and 3 months old to 16 years and 2 months old.

[6] See the OECD website: http://www.oecd.org/pisa/aboutpisa.

[7] For sample questions from the PISA exam, see: http://www.oecd.org/pisa/test.

[8] Latvia and Lithuania, which joined the OECD in 2016 and 2018, are not included.

[9] Hong-Kong is not an independent state but rather a special administrative region of China. Because its education system is separate from the mainland Chinese system, this study treats Hong-Kong as a separate

in 2012, 262,933 test-takers in 2015, and 294,527 test-takers in 2018.

In addition to PISA data, Section 8 also makes use of PIAAC (the OECD's Programme for the International Assessment of Adult Competencies) survey data.

4.1 PISA Performers: top and bottom Countries

Figure 1 presents the math scores by country, averaged across the PISA test 2012, 2015 and 2018 rounds. The 37 developed countries that appear in the figure are the 34 OECD countries (as of 2012) with the addition of Singapore, Hong-Kong and Taiwan.

It is interesting to note that there are striking similarities between some of the countries leading the score board and some of the countries at the bottom. For example, the country pairs of the United States and Canada, Sweden and Finland, and Italy and Switzerland are neighboring economically developed countries, sharing many cultural traits. Poland, Hungary, Slovakia and Estonia are all Eastern European countries which transitioned away from communism during the final decade of the 20th century. Both Israel and Singapore are small, isolated, multi-ethnic countries with high defense expenses. The aim of this paper is to identify the role of school system disciplinary norms in the scholastic success of some countries, and the failure of others.

5 Subjective Discipline: Disciplinary Atmosphere as Reported by Students

In the PISA questionnaire three statements were posed to the students that relate directly to classroom discipline levels: "Students don't listen to what the teacher says," "There is noise and disorder," and "The teacher has to wait a long time for students to quiet down." For each of these statements, the students had to mark one of the following responses: "Every lesson," "Most lessons," "Some lessons," and "Never or hardly ever."

country.

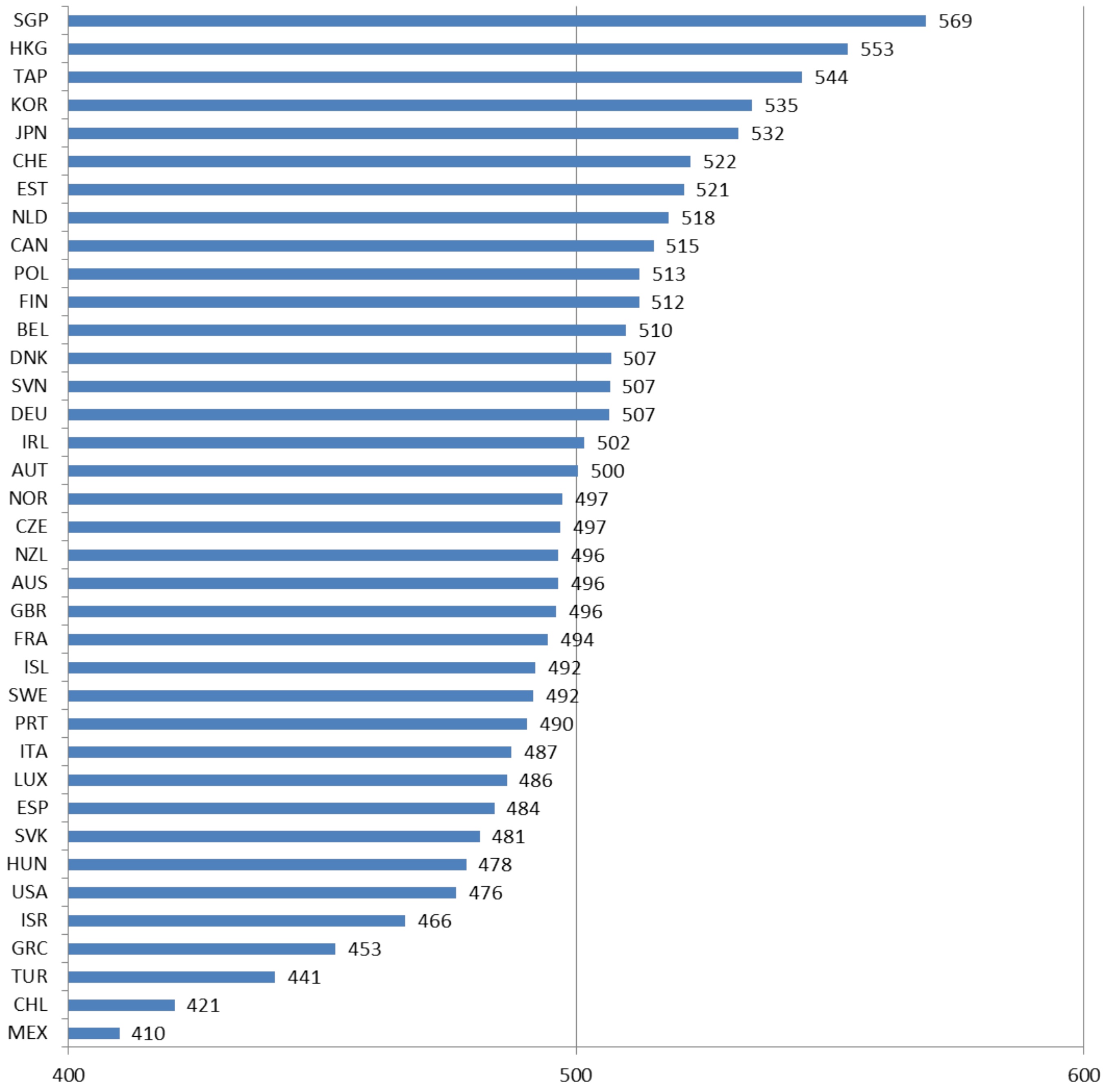

Figure 1: PISA math score by country, averages of the 2012, 2015 and 2018 test rounds

Figure 2 demonstrates the correlation between the perceived disciplinary climate reflected by the response to the statement "Students don't listen to what the teacher say" and the students' average score, compared to the respective national average. This correlation is clear, with students who perceive a good disciplinary climate achieving higher average scores.

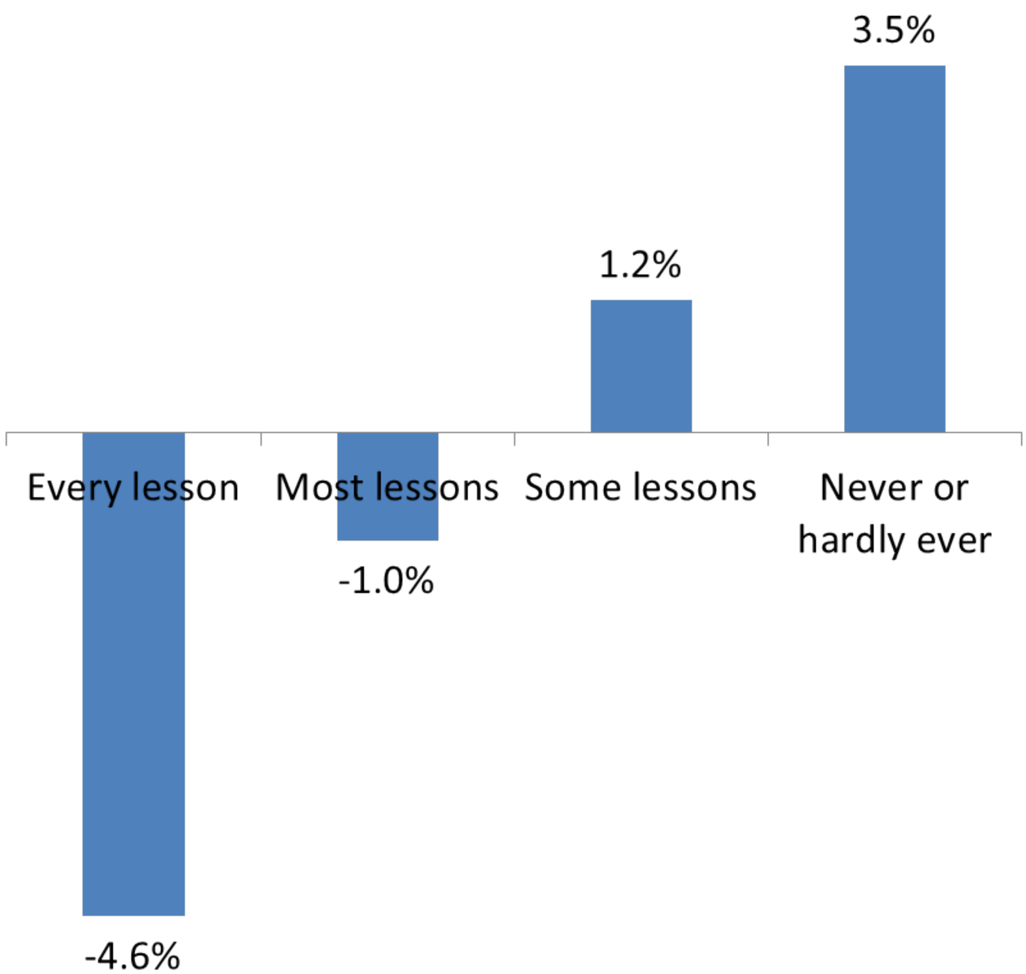

Figure 2: Percent difference from country average score, by response to statement "Students don't listen to what the teacher says"

The results for the statements "There is noise and disorder" and "The teacher has to wait a long time for students to quiet down" are very similar. For these statements as well the average score rises along with the level of classroom quiet and order attested to by students, as can be seen in Figures 14–15 in Appendix A.

On the assumption that cultural bias is relatively small within a given education system (in contrast to bias between different education systems, which can be considerable), it is not surprising that these figures suggest a major advantage for students in classrooms where discipline levels are high.

Using the factor analysis method, the responses to the three statements above ("Students don't listen to what the teacher says," "There is noise and disorder," and "The teacher has to wait a long time for students to quiet down") are reduced to a single 'subjective discipline' index. Country ranking and index scores are presented in Figure 16 in Appendix A. This discipline index is clearly subject to cultural bias, as it is influenced by students' perception of discipline in each locale. Figure 3 displays the 'subjective discipline' index versus PISA math score (both averaged over the 2012, 2015 and 2018 rounds), by country.

While the leading 5 Asian countries (Singapore, Hong-Kong, Taiwan, Korea and Japan) enjoy high scores in conjunction with high subjective discipline, and some of the bottom scoring countries (Chile, Greece, Hungary, Slovakia, Spain and Luxemburg) suffer from both low scores and low subjective discipline, the correlation between the two is far from evident. Norwegian students report a high level of discipline, yet their scores are comparable to those of their French counterparts, who report the lowest levels of discipline in the developed world. Of the bottom scoring countries, Mexican, Turkish and American students all report a strong disciplinary climate, while among the leading countries Finish students report a weak one.

Table 1 uses simple OLS regressions to show the correlation between PISA math scores (on the left-hand side) and subjective discipline (on the right-hand side) at the country, school and student level. At the country level (column 1), there is some correlation: the subjective discipline coefficient is statistically significant at the 5% level (the t-value is 2.38), and the two variables are 0.37 correlated. Considering the small sample size (37 countries), and eyeballing Figure 3, this is far from conclusive.

At the school level (column 2) the subjective discipline coefficient is statistically signif-

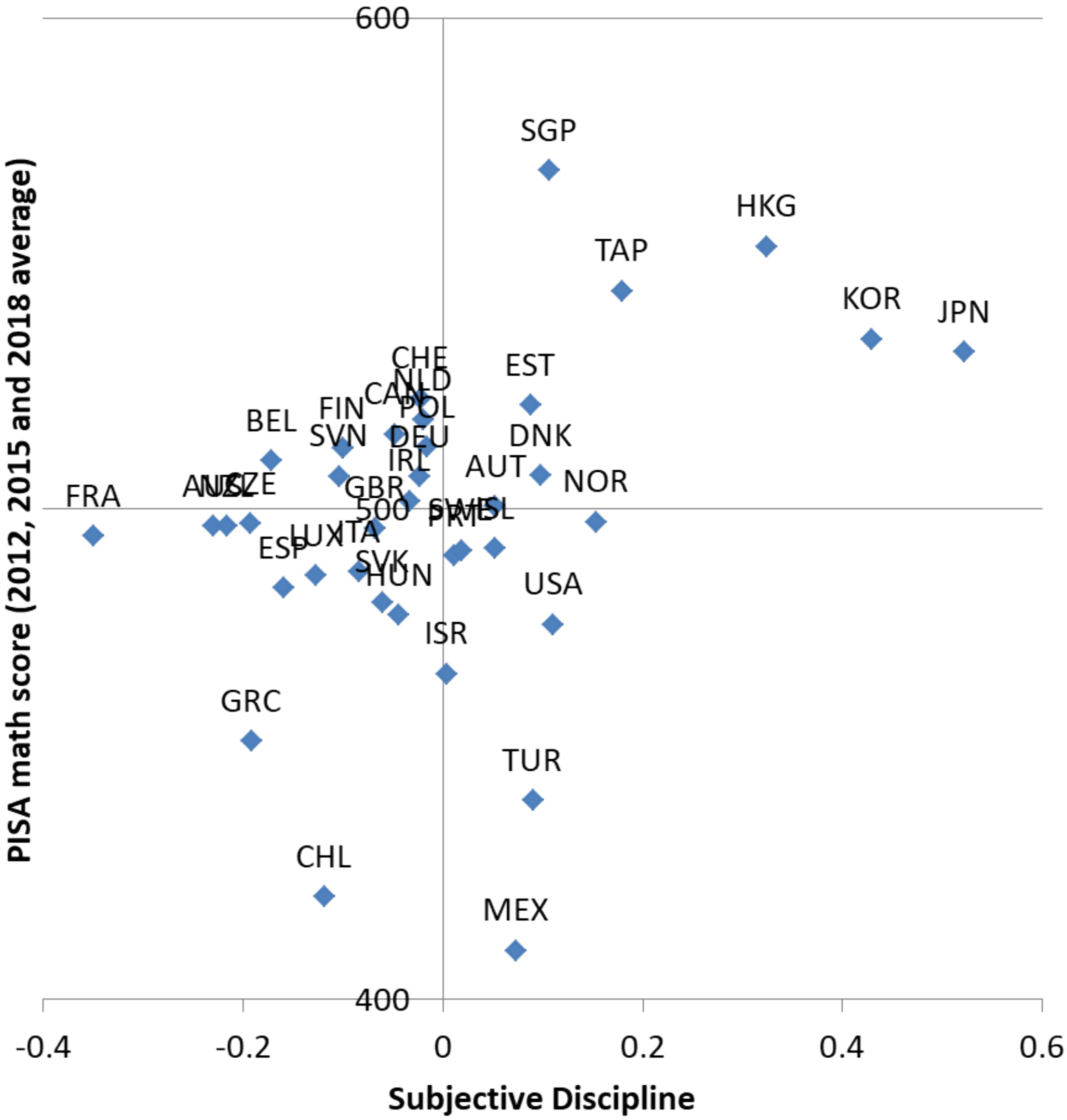

Figure 3: PISA score versus subjective discipline (2012, 2015 and 2018 test rounds average)

	Countries	Schools	Schools with country dummies	Students	Students with country dummies
	(1)	(2)	(3)	(4)	(5)
Subjective Discipline	68.465	40.795	40.904	13.548	12.750
	(28.789)**	(6.009)***	(.768)***	(1.267)***	(.101)***
Const.	497.489	489.069	500.521	499.085	497.418
	(4.970)***	(6.422)***	(1.028)***	(5.440)***	(.473)***
Obs.	37	32456	32456	707564	707564
R^2	.139	.083	.308	.023	.13
F statistic	5.656	46.084	402.576	114.309	3200.916

Table 1: PISA score versus subjective discipline (robust standard errors)

icant at the 1% level (the t-value is 6.69), and the two variables are 0.29 correlated. With 32,456 schools,[10] this result is much more meaningful. However, when country dummies are added (column 3), the statistical significance of subjective discipline to what is now the within-country distribution of scores becomes much stronger, with a t-value of 53.26. Clearly, subjective discipline is much more able to explain scholastic achievements when cultural bias is accounted for, at least partly, by the country dummies. This result highlights the difficulty of comparing discipline levels across countries using student reports. At the student level (column 4), correlation between subjective discipline and score is 0.15, and the coefficient of subjective discipline is statistically significant at the 1% level, with a high t-value of 10.69. However, adding country dummies (column 5) once again makes subjective discipline a much better predictor of student (within country) achievement level, with a t-value of 126.71.

6 Objective Discipline: Disciplinary Atmosphere as Reflected by Truancy and Tardiness

Truancy and Tardiness have a direct negative impact on the students themselves, as they lose study time; they also affect the entire class by creating gaps in knowledge that have to be addressed. But beyond this, data on punctuality and truancy also testify to classroom and school-wide discipline levels.[11] The PISA questionnaire questions students about the number of times they arrived late to school, the number of their full-day absences, and the number of times they skipped class (without skipping the entire school day), during the two weeks that preceded the test.

As with the subjective statements regarding discipline, the data on tardiness correlate with student exam scores. This can be seen in Figure 4. There is a simplistic explanation

[10]Over the three test rounds. It is possible that some schools were sampled more than once, but this cannot be ascertained through the data.

[11]The effects of truancy on scholastic achievement have been explored in papers such as Bosworth (1994) and Goodman (2014).

for the correlation between number of lateness instances and test score. It is likely that, on some of the occasions, it was to their math lesson that the students were late, meaning that they missed class time and material and accordingly lagged behind the rest of the class. Beyond this explanation, however, it is also likely that the students who were late are also less disciplined and attach less importance to their studies. Additionally, it is reasonable to assume that the classes in which a high percentage of the students who came late are classes whose discipline level is low. As will be seen below, the data match well with these statements.

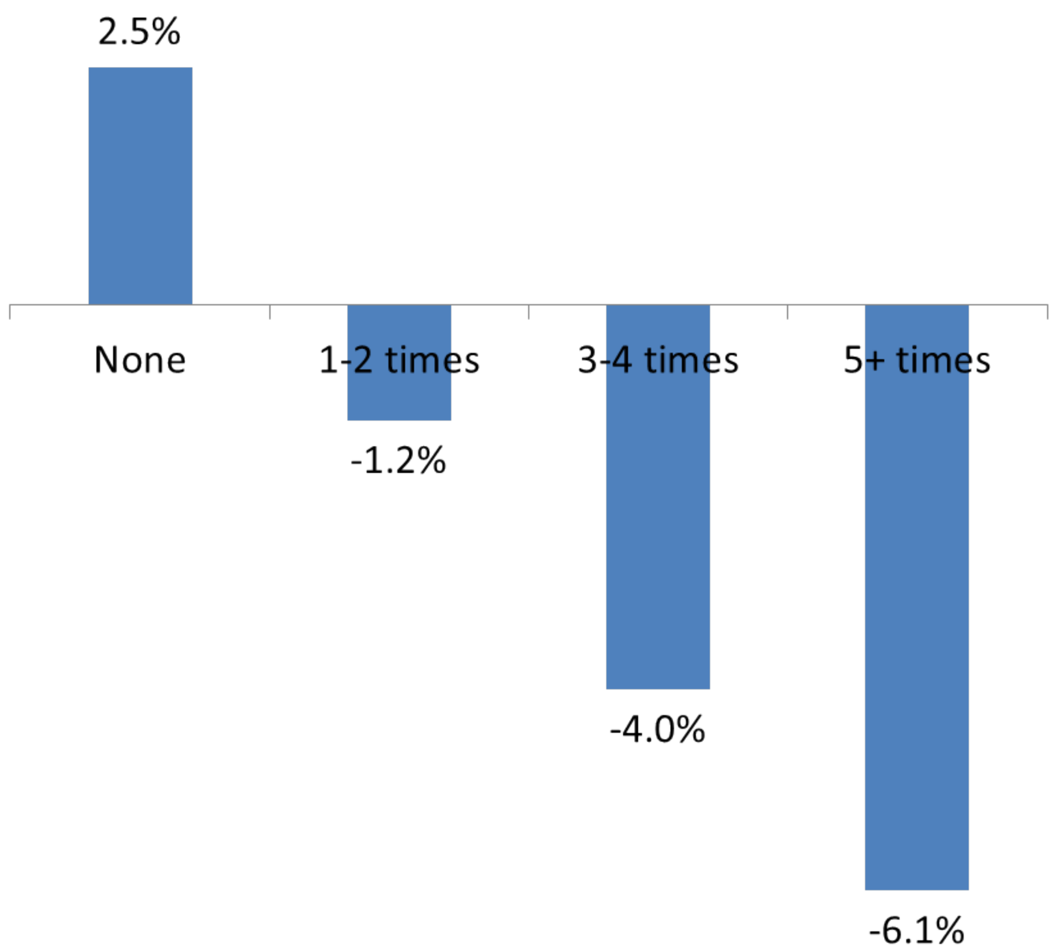

Figure 4: Score percent difference from country average, by number of late arrivals to class

Classroom breakdown by the level of tardiness enables a look at the peer effect - the correlation between classmates' tardiness and student scores. As Figure 5 shows, student scores correlate with their own behavior (the number of times they were late during the preceding two weeks), but no less so than they do with their classmates' behavior. The average score of a 'punctual' student, who did not arrive late during the two weeks that preceded the exam and who is enrolled in a 'tardy' class, where 75% of the students were late at least once during that period, is lower than the average score of a student who came late 3-4 times but who is in a 'punctual' class where the percentage of latecomers does not exceed 25%.

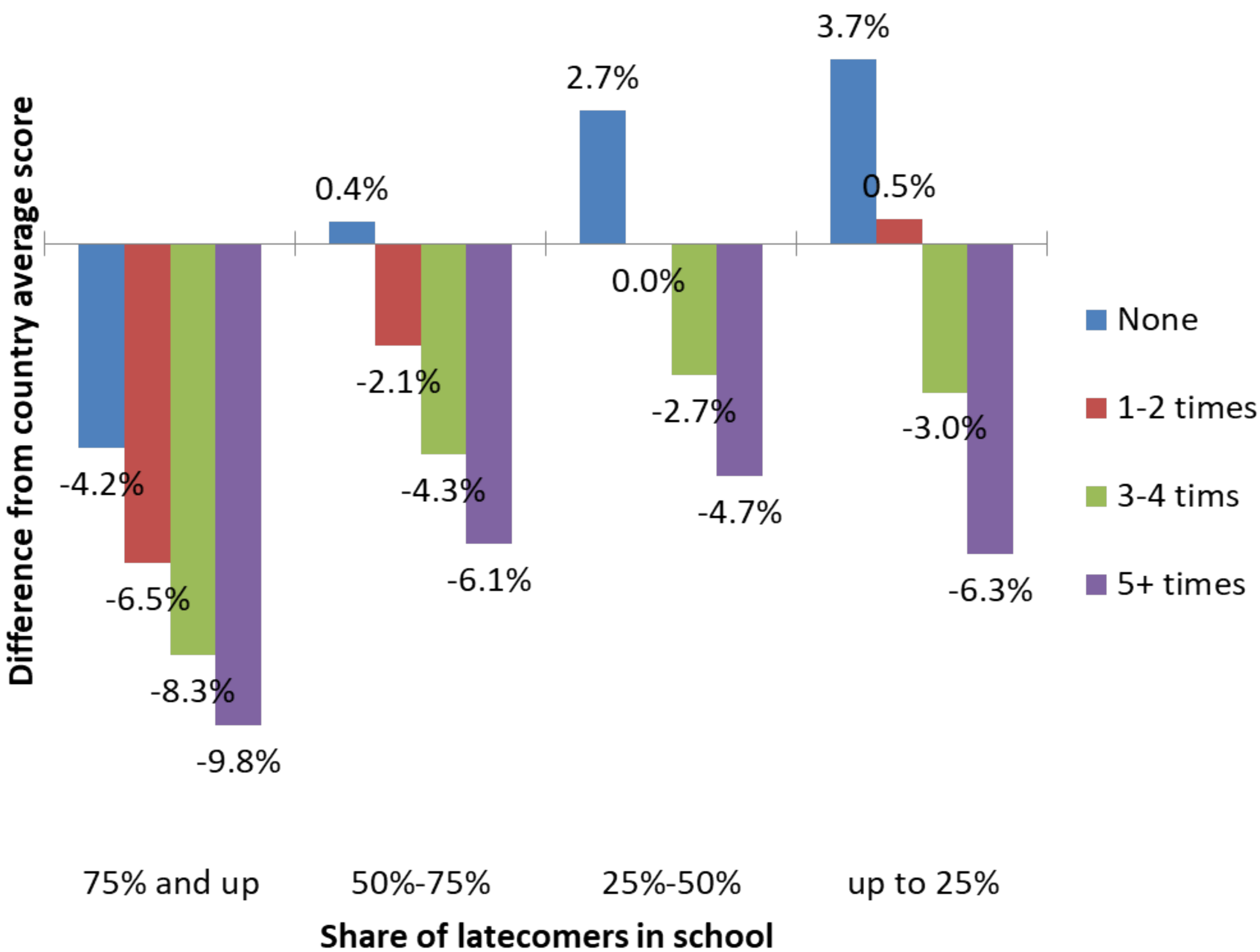

Figure 5: Score percent difference from country average, by school and individual level late arrivals

The remaining data on truancy (full-day absences) and skipping class paint a similar picture. In both cases, these forms of truancy are correlated to lower scores. The data are presented in Figures 17-18 in Appendix A.

Using the factor analysis method, the tardiness and truancy data (late arrivals, full-day absences and skipping classes) are reduced to a single 'objective discipline index.' Country ranking and index scores are presented in Figure 19 in Appendix A.

Figure 6 displays the 'objective discipline' index versus PISA math score (both averaged between the 2012, 2015 and 2018 rounds), by country. A partial correlation can be readily detected between objective discipline level and mean country score. Of the top performing countries, the 5 Asian nations, Switzerland and the Netherlands all exhibit very high discipline levels; by contrast, discipline levels in Finland, Estonia and Canada are low to average. Some of the low-scoring countries, such as Turkey, Italy, Greece and Israel, and to a lesser degree Spain, Portugal, Slovakia, Mexico and the US, are characterized by low discipline levels. However, of the low-scoring countries, data from Chile and Sweden indicate a respectable level of discipline, and Luxemburg and Hungary's level of discipline is in fact quite high.

The 'objective discipline' index reinforces many cultural stereotypes: the East Asian and Northern European countries are at the top of the ladder, while the Mediterranean countries (and Portugal) occupy the bottom rungs.

Similarly to Table 1 with regard to the 'subjective discipline' index, Table 2 uses simple OLS regressions to show the correlation between PISA math scores (on the left-hand side) and objective discipline (on the right-hand side) at the country, school and student level. At the country level (column 1), there is some correlation: the objective discipline coefficient is statistically significant at the 1% level, with a t-value of 4.19, compared to 2.38 for subjective discipline. The two variables (scores and objective discipline) are 0.58 correlated, compared to 0.37 for subjective discipline. R^2 is 0.33, more than double that of the respective subjective discipline regression (0.14, column 1 in Table 1).

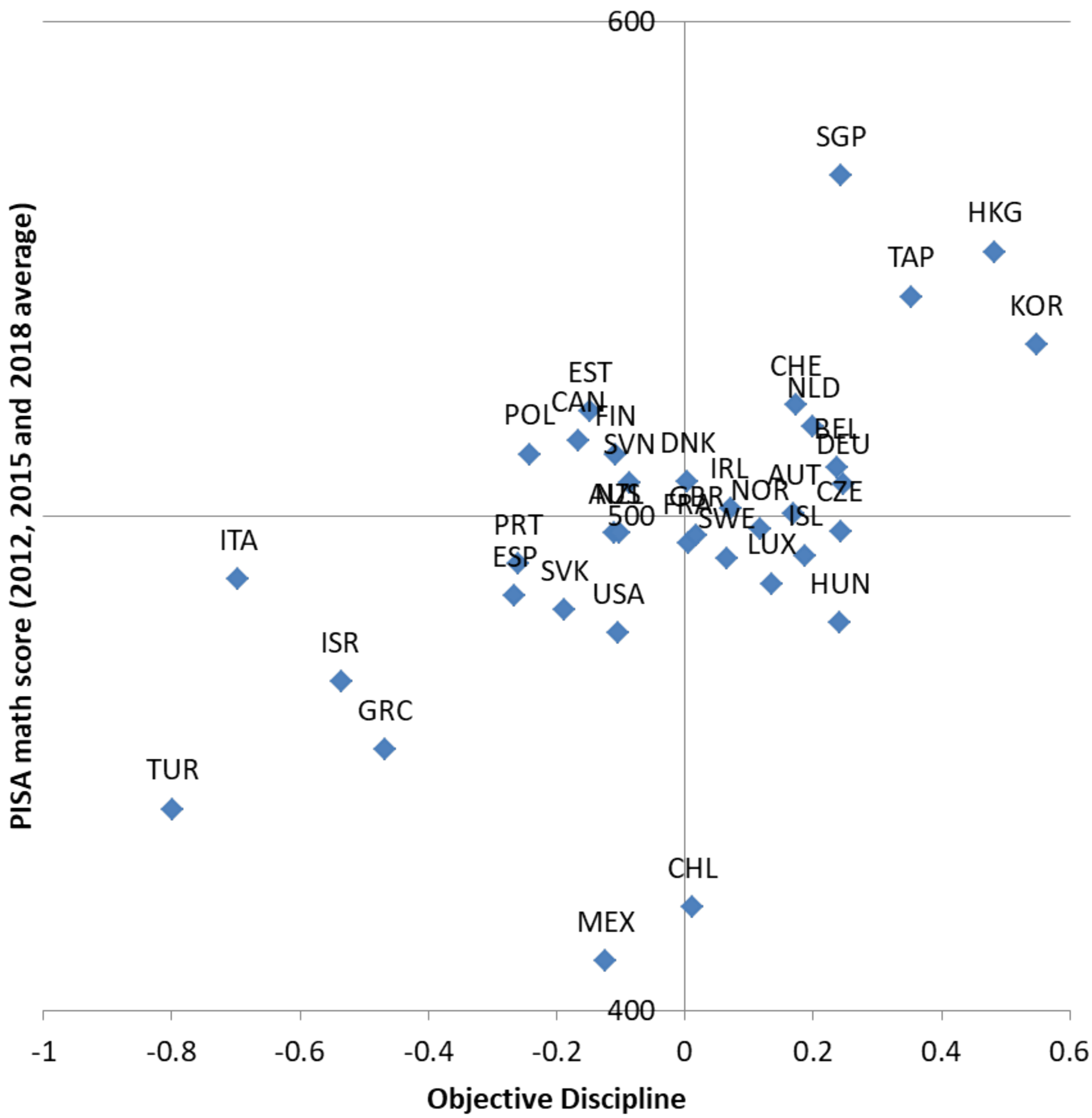

Figure 6: PISA score versus objective discipline

	Countries	Schools	Schools with country dummies	Students	Students with country dummies
	(1)	(2)	(3)	(4)	(5)
Objective Discipline	59.728	48.570	47.025	20.926	18.130
	(14.264)***	(4.451)***	(.836)***	(1.612)***	(.095)***
Const.	497.428	491.922	496.793	498.548	494.850
	(4.372)***	(6.002)***	(1.008)***	(5.523)***	(.445)***
Obs.	37	32386	32386	801073	801073
R^2	.334	.145	.328	.056	.15
F statistic	17.534	119.061	419.572	168.471	4186.618

Table 2: PISA score versus objective discipline (robust standard errors)

At the school level (column 2) the objective discipline coefficient is statistically significant at the 1% level, with a t-value of 10.91, which is only slightly higher than the 6.79 t-value in the subjective discipline case (column 2 in Table 1). However, objective discipline is better correlated with PISA score – correlation of 0.38 versus 0.29 – and the 0.145 R^2 value is, again, much larger than that of the respective subjective discipline regression (0.083, column 2 in Table 1). Similarly to the subjective discipline case, the statistical significance of objective discipline increases when country dummies are added (column 3), with a t-value of 56.25 (compared with 53.26 for subjective discipline). The gap in fit, as measured by the R^2 statistic (0.36 versus 0.34), narrows with the use of country dummies, which eliminate country-level cultural bias in the reporting of the disciplinary climate by students.

At the student level (column 4), objective discipline again exhibits better correlation with the PISA score (0.24 versus 0.15), and while at 12.98 the t-value of the coefficient of objective discipline is statistically significant at roughly the same level as that of subjective discipline (a t-value of 10.69), the R^2 statistic (0.056) is again much larger than that of the respective subjective discipline regression (0.023). As in the school-level regressions, adding country dummies (column 5) once again greatly narrows the gaps in fit.

7 The Effects of Discipline: Data Analysis

7.1 The Relation Between Objective and Subjective Measures of Discipline

As Figure 7 demonstrates, the picture that emerges from truancy and tardiness data (reflected by the 'objective discipline' index) can be quite different from the one painted by students' subjective impressions regarding the discipline levels in their classes. While some countries are located close to the diagonal of the graph, meaning that their subjective and objective indexes are similar, or in other words, that the discipline levels reported by their students are realistic, for some countries this is clearly not the case. Turkish, Italian and Israeli students report a disciplinary climate much better than what the more objective truancy and tardiness data indicate, while for France, Belgium and the Czech Republic the opposite is true. Overall, the two indexes are 0.45 correlated at the country level.

As discussed before, the major advantage of using data on truancy and tardiness to evaluate discipline levels is the fact that they are less susceptible to cultural bias compared to direct questions on classroom disciplinary atmosphere. What is left is to establish that there is indeed a strong relation between truancy and tardiness and classroom discipline level.

In order to examine this relation, Table 3 presents regressions of the 'subjective discipline' index, on the left hand side, versus the 'objective discipline' index, on the right hand side, at the country, school and student levels. As can be seen in column 1, at the country level the objective discipline index coefficient is significant at the 1% level (a t-value of 2.95). At the school level (column 2) the objective discipline index coefficient is even more significant, with a t-value of 6.86. When country dummies are added (column 3) the t-value becomes much higher – 34.4. Clearly, with cultural bias accounted for by country dummies, truancy and tardiness data provide strong indications for disciplinary climates in schools. Columns 4 presents the results at the student level, with country dummies added in column 5. Once

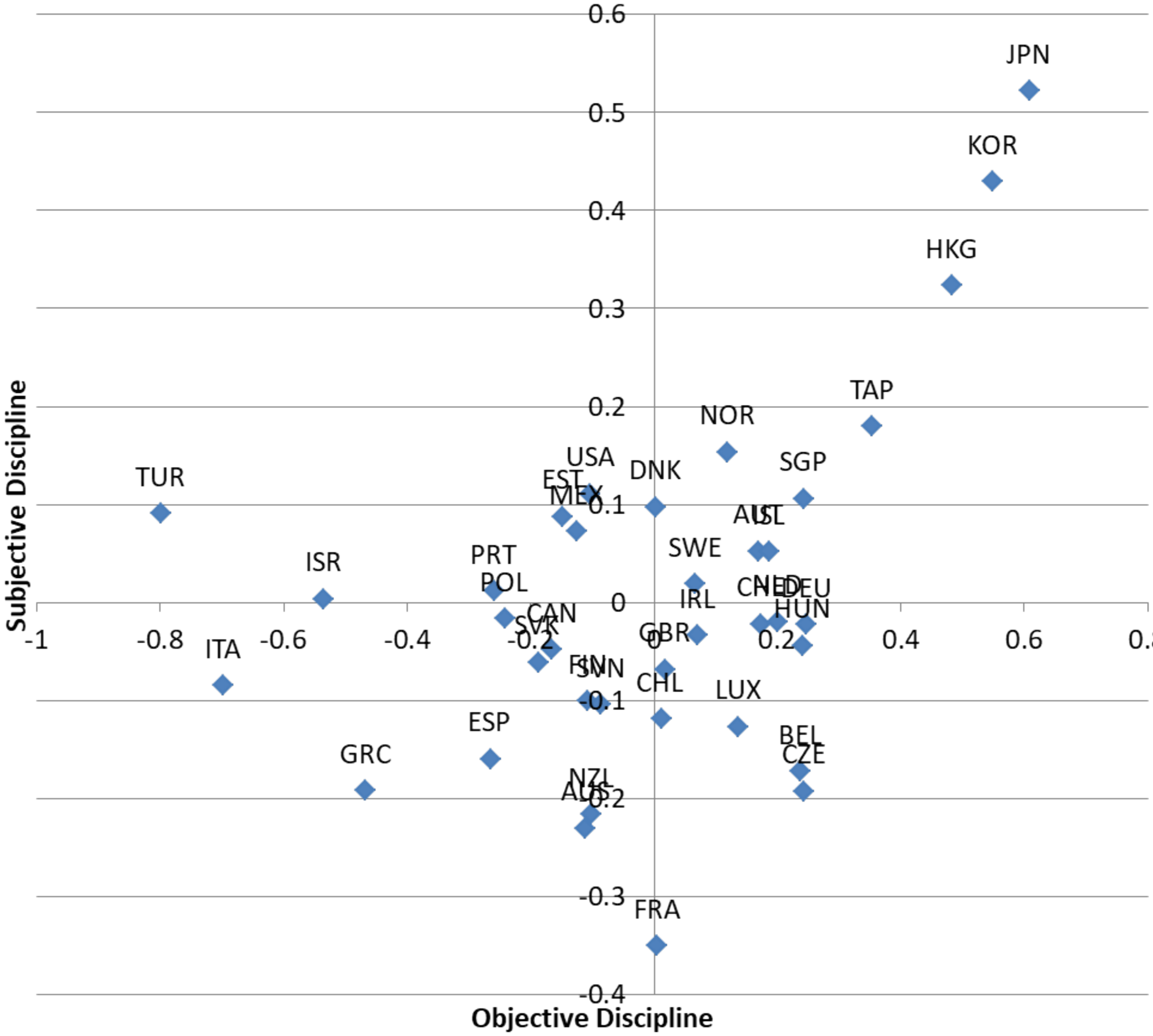

Figure 7: Subjective versus objective discipline

	Countries	Schools	Schools with country dummies	Students	Students with country dummies
	(1)	(2)	(3)	(4)	(5)
Objective Discipline	.251	.247	.241	.138	.125
	(.085)***	(.036)***	(.007)***	(.013)***	(.001)***
Const.	-.002	-.021	-.220	-.012	-.239
	(.026)	(.025)	(.010)***	(.026)	(.006)***
Obs.	37	32123	32123	660524	660524
R^2	.199	.075	.172	.02	.043
F statistic	8.675	47.676	263.913	104.276	1071.067

Table 3: Subjective versus objective discipline (robust standard errors)

again, the ‘objective discipline’ index coefficient is highly significant (t-value of 10.2), and becomes much higher (t-value of 96.2) when adding country dummies.

7.2 Discipline and Class Size

Figure 8 presents class sizes and PISA math scores across countries (2012, 2015 and 2018 averages). Clearly, class size by itself is not a good indicator of PISA scores. PISA scores in Iceland and France are virtually identical, despite the former having an average class size of 18 and the latter a class size of 29. Chile, Israel, France, Korea and Hong-Kong all have very similar class sizes, yet their PISA scores differ wildly. As discussed in Section 2, most of the current literature on the effect of class size on scholastic achievements focuses on the micro-level – what would happen to students if class size is changed *Ceteris Paribus*. This is very much a partial equilibrium analysis, as does not take into account the system-level effect of changes in average class size on budget needs and on the quality of teachers available. Also absent from the ‘class size’ literature is the discussion on the role of discipline over the effect of class size on students’ achievements.

Figure 6 and Tables 1 and 2 suggest that although a correlation between discipline levels and achievements exists, it is far from perfect, and that there are additional influencing factors. The theoretical analysis in Section 3 points to an interaction - a reciprocal relationship - between discipline level and class size, meaning that the bigger the class, the greater the

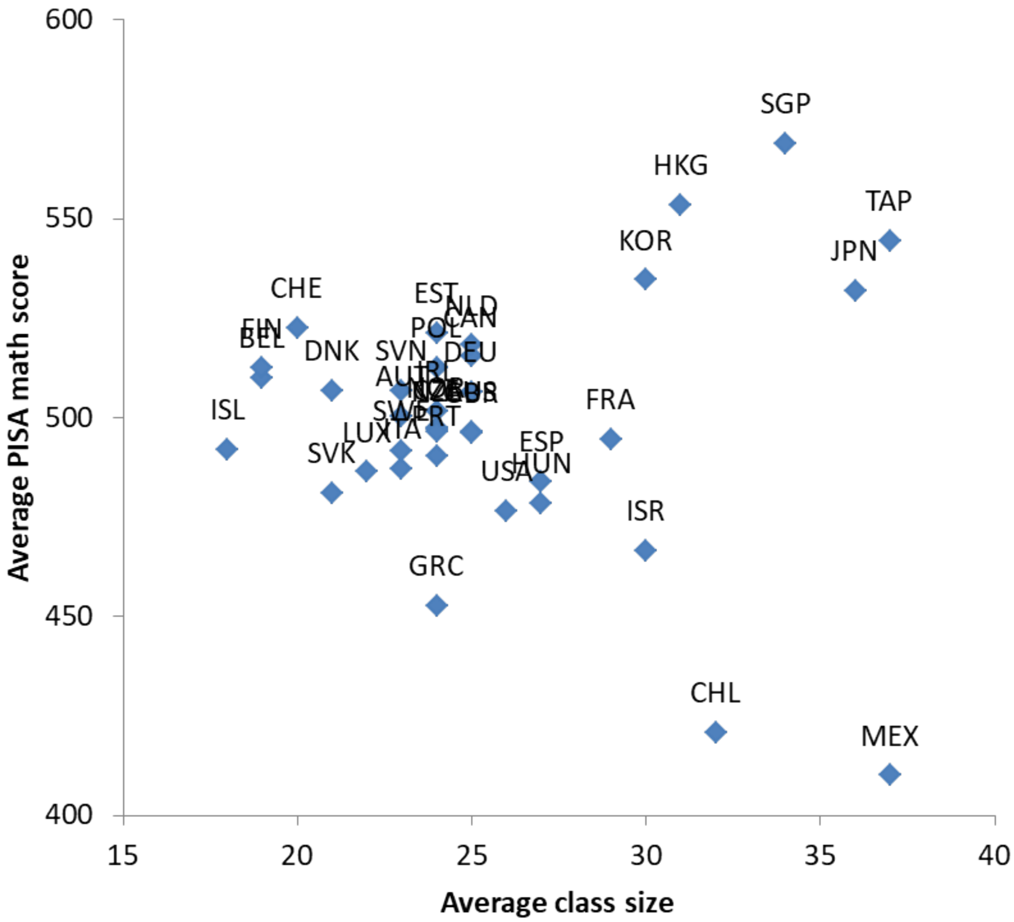

Figure 8: PISA math scores versus class size, by country

degree to which low discipline levels will compromise achievements. The present study is first in establishing empirical support for this theoretical hypothesis, as shall be demonstrated later on.

Figure 9 presents the objective discipline level versus class size, reported by the schools. As can be seen, the 5 leading countries, located in Asia, display high discipline levels and large classes, while most Western countries are characterized by smaller classes and differing levels of discipline. Of the latter, countries such as Germany, Belgium, the Czech Republic, Iceland, the Netherlands and Switzerland enjoy high levels of student discipline, meaning that it could be more optimal for them to have larger classes, allowing them to improve teaching quality by sending away the least-capable teachers as well as using the saved resources to improve teaching quality. Of the low-performing countries, Chile and Mexico have large classes, but only a medium level of student discipline, and Turkey, Israel, Greece and Spain have medium sized classes, but low discipline levels. It can therefore be argued that in order to improve their achievements, the school systems in these countries need to either lower class sizes, at a considerable cost (financially, and/or in terms of teachers' quality), or increase discipline.

It is worthy of note that class size is correlated with student discipline not only between countries, but also within countries. Figure 10 presents average objective discipline by class sizes, among all PISA participants.

To clarify the point that this correlation is just as apparent within countries as between them, Table 4 shows that this relation remains strong when using country fixed-effects for within-country estimation. The rational at the root of the theoretical model presented at Section 3 is that higher discipline enables efficient teaching in bigger classes. Here as well data matches with theory. The more disciplined the students, the less individual attention they require and the more feasible it is to teach them in large groups.

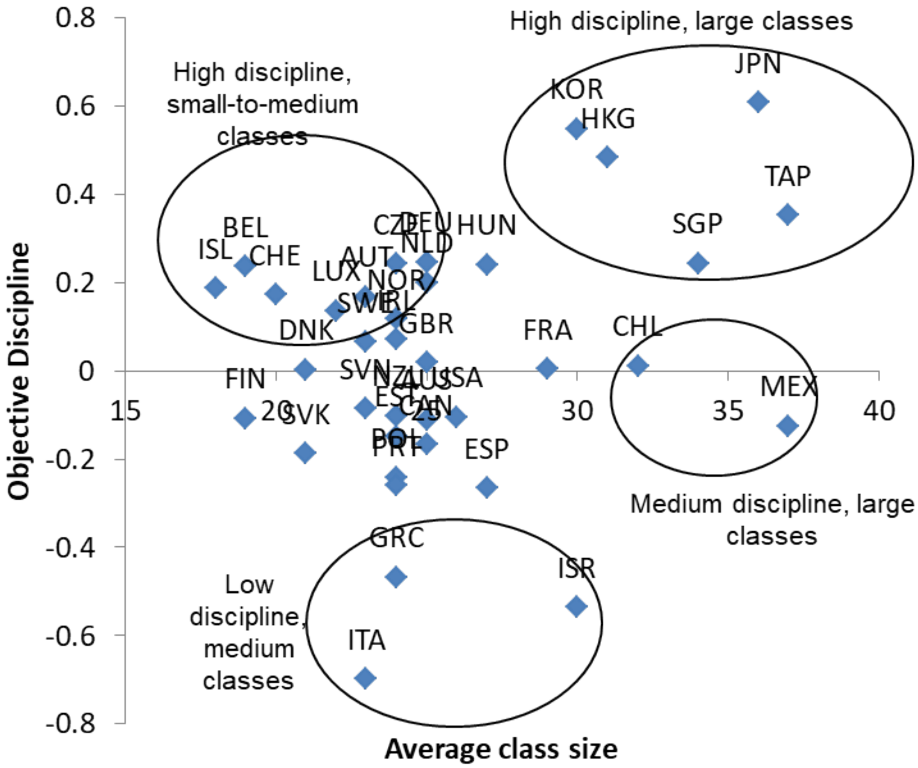

Figure 9: Objective discipline versus class size

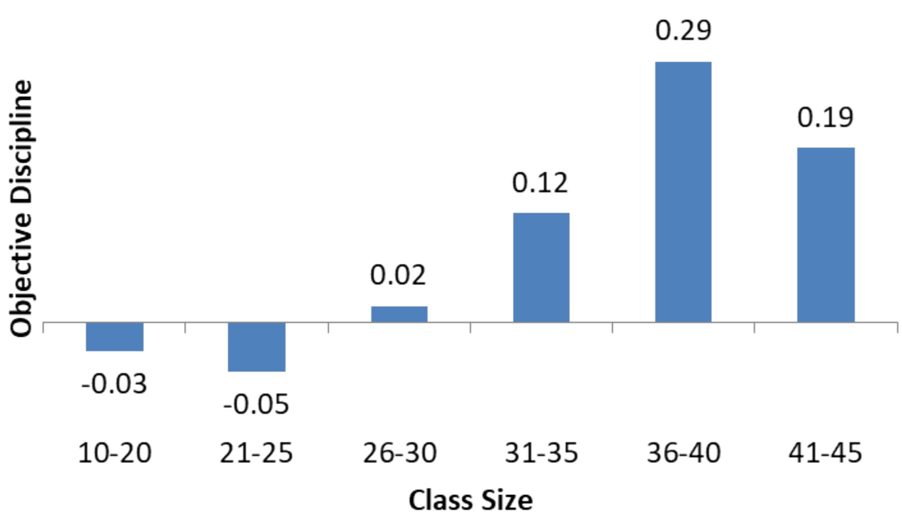

Figure 10: Objective discipline by class size

Class Size	
Objective Discipline	.155
	(.008)***
Const.	27.410
	(.008)***
Obs.	731880
R^2	.453
F statistic	360.464

Table 4: Class size versus objective discipline, OLS with country dummies, robust standard errors

7.3 A Multivariate Analysis of Student Achievements

As noted before, students' scholastic achievements are determined by many factors simultaneously. To weigh the effects of these factors, three regressions were performed, presented in Table 5.This is in effect a cross-sectional estimation of the 'education production function,' as discussed in Krueger (1999). A generalized form of the 'education production function' estimated is formulated in Equation 9 below:

$$y_{i,j,c,t} = \alpha + \beta_1 Student_i + \beta_2 Class_j + \beta_3 Country_c + \beta_4 Round_t + \epsilon_{i,j,c,t} \tag{9}$$

Where $y_{i,j,c,t}$ is the PISA math score of student i in class j at country c, tested at PISA round t. $Student_i$ is a vector of student-level data, i.e. individual discipline and parents' education. $Class_j$ is a vector of class-level data, i.e. class size, average school discipline,[12] the share of parents with high education, and the number of math classes per week. $Country_c$ and $Round_t$ are country and PISA round dummies, respectively, and $\epsilon_{i,j,c,t}$ is the stochastic error term. For an in-depth analysis of each of these factors, see Gruber (2017).

Of particular interest to the analysis is the coefficient of the interaction between class size and school discipline, as will be discussed shortly.

[12]In PISA sampled schools students in the PISA age bracket (15 and three months to 16 and two months) are randomly selected across classes. Therefore discipline, averaged across students from the same school, is a school-level measure.

	(1)	(2)	(3)	(4)
Objective Discipline			11.407 (.141)***	11.611 (.134)***
Subjective Discipline	6.444 (.152)***	6.582 (.143)***		
Objective School Discipline			10.880 (.896)***	13.668 (.994)***
Subjective School Discipline	5.833 (1.279)***	13.419 (1.272)***		
Obj. Sch. Dis. * Class Size			.732 (.029)***	.484 (.034)***
Sub. Sch. Dis. * Class Size	.971 (.044)***	.504 (.044)***		
Class Size	.286 (.017)***	.893 (.023)***	.614 (.016)***	.873 (.021)***
Country Dummies	No	Yes	No	Yes
Const.	369.872 (1.555)***	359.011 (1.601)***	363.726 (1.352)***	351.501 (1.424)***
Obs.	462806	462806	567319	567319
R^2	.208	.311	.235	.321
F statistic	3749.191	2766.732	5231.124	3457.571

Table 5: Regressions on student-level PISA math scores. Robust standard errors. Variables used but not displayed: maternal and paternal education, share of schoolmates whose parents have post-secondary education, number of weekly math classes and PISA round dummies.

Regressions (1) and (2) (columns 1 and 2 in Table 5) make use of the subjective discipline measure, both at the student and at the school level. The impact of discipline is assessed at the individual student level, and at the school level; the school discipline average is multiplied by dummy variables for class size. Regression (2) adds country dummies to the estimation. While, as discussed before, subjective discipline is prone to cultural bias and thus is a problematic measure to explain cross-country differences, its within-country correlation with student scores, as well as the correlation with objective discipline (shown in Table 3), make student-level and school-level subjective discipline statistically significant at a 1% level. The interaction term of school-level subjective discipline and class size is also positive and statistically significant at a 1% level. Adding country dummies in regression (2) greatly improves the fit, as it removes the country-level bias in subjective discipline.

Regressions (3) and (4) (columns 3 and 4 in Table 5) make use of the objective discipline measure, but are otherwise identical to regressions (1) and (2). The comparison between regressions (1) and (3), which do not include country dummies, once again shows the objective discipline to fit the data better, with higher T and R^2 values. The same is true for the comparison between regressions (2) and (4), which include country dummies, though the difference in fit is smaller, as can be expected, since the country fixed effects eliminate the cultural bias in the perception of discipline between countries.

The interaction term of school-level subjective discipline and class size is once again, positive and statistically significant at a 1% level. It is therefore clear that classroom discipline gains in importance as class size grows. This new empirical finding matches well with the theoretical model presented in Section 3.

Table 6 presents the results of a first-difference estimation using the 2012, 2015 and 2018 PISA rounds. Since there is no way to identify schools across PISA rounds, artificial school bins are defined, and the first-difference is performed between the equivalent school bins across rounds. These bins are defined using the country, school type (private independent / private government-dependent / public), ISCED level (lower or upper secondary school), test language and community type (village / small town / town / city / large city) variables. A total of 467 school bins with support in all three PISA rounds were created.[13]

The first difference estimation in Table 6 aim to look at similar school groups between rounds, and break down change in school performance to its underlying factors. Even among schools grouped by similar characteristics, changes in objective discipline are better correlated to changes in school performance than changes in subjective discipline, as evidenced by the higher t-values and R^2 in regression (2) compared to regression (1).

[13]While PISA country-level data is stratified into sub-country school clusters, this stratification differs between rounds, and thus cannot be used for creating school bins.

	Objective Discipline	Subjective Discipline
	(1)	(2)
D(Subjective Discipline)	13.621	
	(3.497)***	
D(Objective Discipline)		20.360
		(3.004)***
D(Class Size)	1.221	1.162
	(.265)***	(.270)***
D(Parents' Education)	111.222	116.048
	(11.509)***	(10.396)***
D(Number of Math Classes)	6.674	6.249
	(2.638)**	(2.681)**
Obs.	886	886
R^2	.284	.327
F statistic	45.168	45.197

Table 6: PISA score first difference regressions (robust standard errors)

8 Identification and Causation

As established, PISA data demonstrates strong correlation between measures of objective discipline, based on truancy and tardiness data, and test scores. Further more, this correlation grows stronger with class size, as predicted by the model presented in Section 3. The model predicts higher discipline to improve educational outcomes through two channels - the direct one, in which higher discipline leads to better class atmosphere and higher grades, and the indirect one, in which higher discipline leads to a better allocation in terms of class size and teaching quality. Both predictions match the data, where higher discipline is positively correlated with test score both directly and via an interaction with class size.

Whether these correlations stems from causation, i.e. that higher discipline leads to better test scores, can and should be challenged. Is there a third factor, a missing variable, effecting both discipline and scores and thus rendering the correlation between the latter two spurious? The main two confounding factors that come to mind are: a. student motivation, and b. teacher quality. It can be argued that one or both of these factors impact both student discipline and student scores.

When it comes to student motivation, it is first interesting to note that there is very little between-country correlation between motivation, as reported by students in the PISA questionnaire, and scores. For example, Israeli students lead the world in the level of importance they attribute to math studies, yet their scores are lower than the OECD average, while student in Japan, Taiwan, Hong-Kong and Korea attribute the math studies the least importance on average.[14] However, students' reports on motivation may be prone to cultural bias, much like their reports on classroom discipline. It is reasonable to expect a more motivated student to be more disciplined, and for the individual student it may indeed be the case that both self-discipline and performance are a manifestation of motivation. However, as the 'Objective School Discipline' variable in Table 5 demonstrates and Figure 5 graphically illustrates, individual student performance is heavily influenced by school-level discipline. Whether or not the discipline maintained by the individual student's peers is caused in turn by the peers' inner motivation seems beside the point. Finally, the 'motivation' explanation fails to account for the effect of discipline increasing with class size.

With regard to teacher quality, it can be argued that better teachers are able to better maintain a disciplined atmosphere in the classroom. Furthermore, the theoretical analysis in Section 3 suggests that higher discipline would enable school systems to have larger classes, thus reducing the number of teachers needed, vacating resources for teaching quality and being more selective vis-à-vis the teachers employed, as argued in Jepsen and Rivkin (2009). It could also be argued that higher student discipline would attract a higher caliber of teachers to specific schools or the teaching profession in general.[15] The causal relationship between country- and school-level discipline and teacher quality are therefore potentially complex and bi-directional.

As discussed in Hanushek and Rivkin (2006), teacher quality is hard to measure. The added-value approach, measuring teacher quality as a residual on the effect of other inputs,[16] is not applicable for attempting to distinguish between the possibly correlated effects of discipline and teacher quality. The PISA test does not include a measurement of teacher

[14]See Gruber (2017) for details.

[15]For example, /citetHKR2004 shows that student characteristics are important factors in the decisions of teachers whether to leave specific schools.

[16]See Hanushek et al. (2005) and Hanushek and Rivkin (2006) for details.

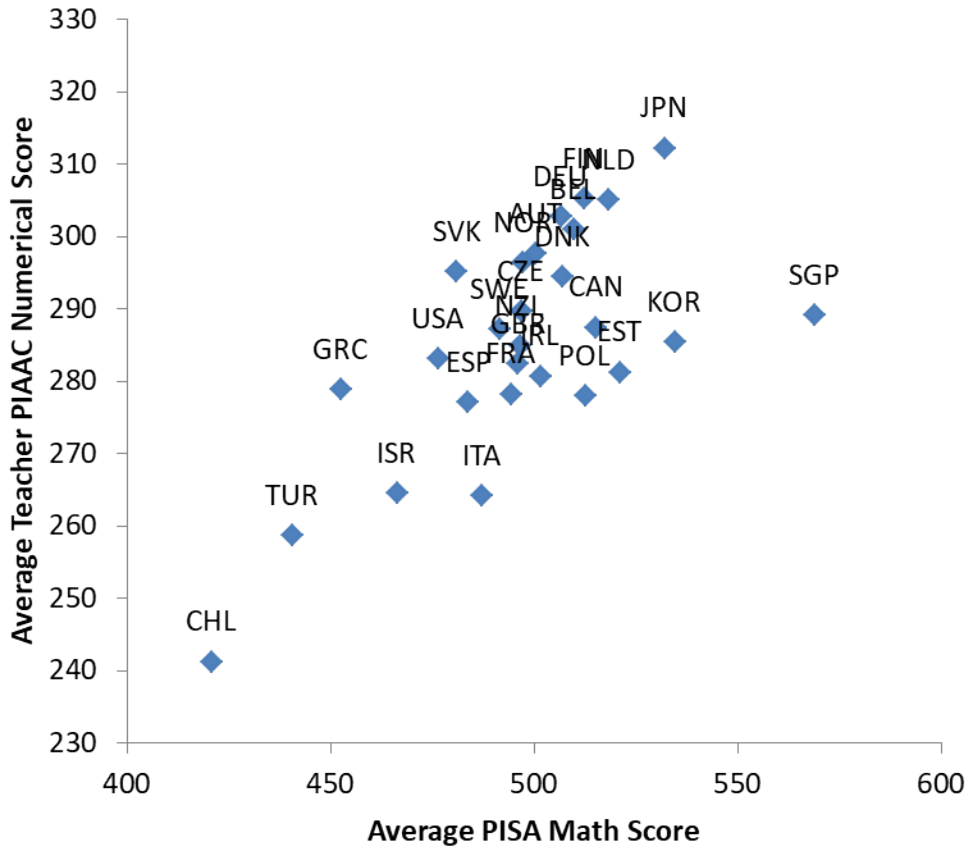

Figure 11: Teachers' skills versus student scores, correlation: 0.68

quality. This paper uses PIAAC (the OECD's Programme for the International Assessment of Adult Competencies) numerical skills test scores by teachers, at the country level.[17] While the literature is undecided on whether or not teachers' test scores are a good indication on their quality as teachers,[18] their numerical skill level is arguably a good indication of their ability to comprehend and teach mathematical thinking and of the alternative career paths open to them.

The intersection of PISA and PIAAC provides a set of 27 developed countries. While this is too small a sample to use for statistical analysis, eyeballing the data can provide some insights.

[17] For a recent paper comparing teachers' PIAAC scores to students' PISA scores, see Hanushek et al. (2019).

[18] See Hanushek and Rivkin (2006) for review.

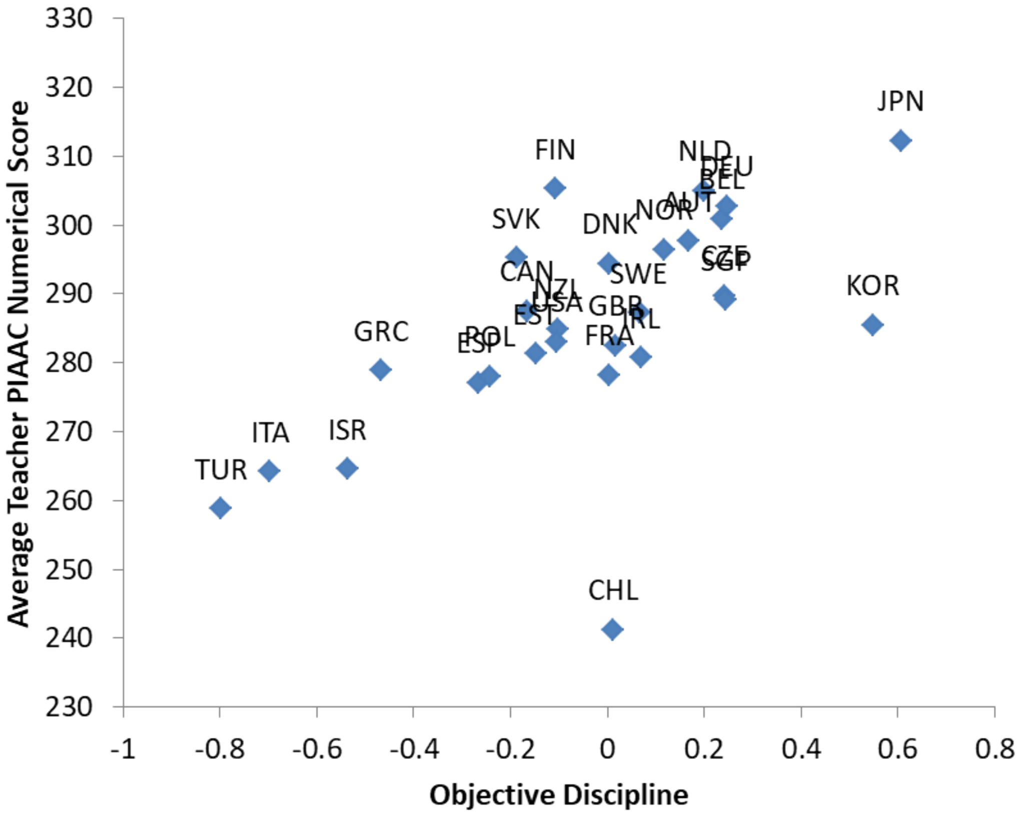

Figure 12: Teachers' skills versus objective discipline, correlation: 0.62

Figure 11 shows average teacher PIAAC numerical score versus average PISA math score by country. The 0.68 correlation is visually evident. However, some countries that are top performers in PISA, namely Singapore, Korea, Canada, Estonia and Poland, are quite mediocre when it comes to teachers' PIAAC scores. Here as well correlation is not necessarily causation – for example, it could be that a highly skilled society (as reflected by PISA scores) naturally leads to high skills in all professions, including teaching.

Figure 12 correlates teachers' PIAAC numerical scores with the 'objective discipline' index. Here as well the correlation is clear to the eye, but the direction of causality is not. It may be that better teachers create a more disciplines class atmosphere, it could be that societal norms of student discipline attract more able individuals to become teachers, and it could indeed be both.

It is informative to see that there is no visually apparent correlation between teacher's

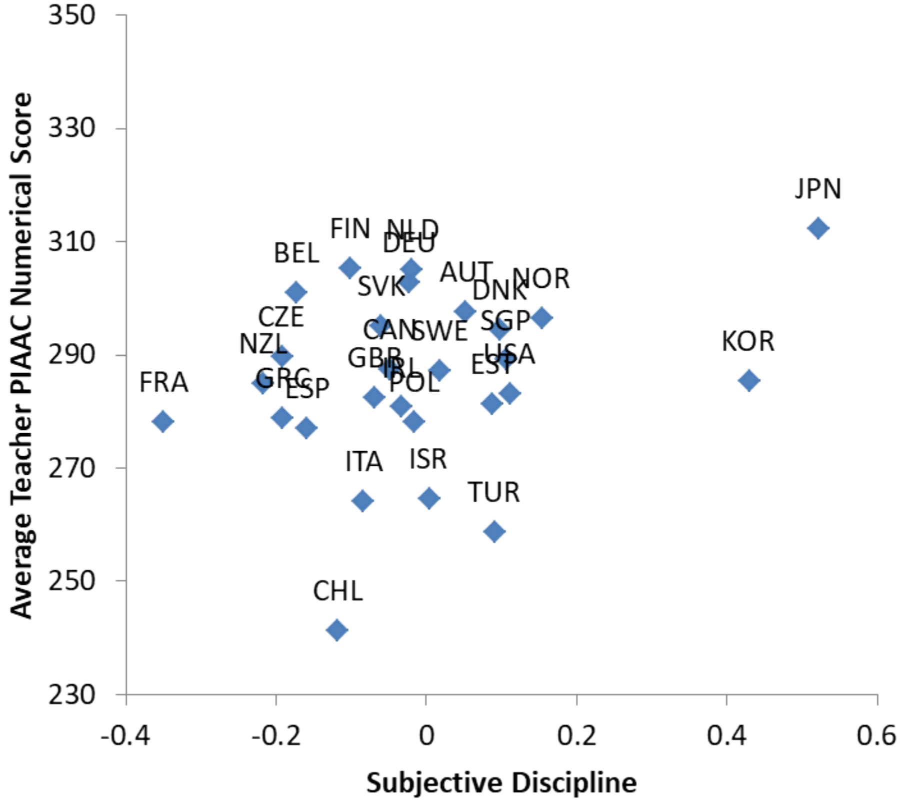

Figure 13: Teachers' skills versus subjective discipline, correlation: 0.27

skills and subjective discipline, as Figure 13 demonstrates.

Unfortunately, the structure of the PISA test, which does not allow for the identification of specific schools between rounds, and which does not provide quantitative data on disciplinary measures and enforcement, makes it difficult to disentangle the effects of discipline from those of teacher quality. However, as with the alternative explanation of student motivation, teacher quality on its own is unable to account for the effect of discipline increasing with class size, which is predicted by the theory (see Section 3) and found to match very well with the data (see Table 5). It can hardly be argued that the quality of teachers has a greater impact on individual students the larger the is number of students.

9 Conclusion

Beyond simple intuitiveness, the importance of discipline to the classroom learning experience has a solid basis in research, both theoretical and empirical. It should not be assumed that stricter discipline is always better; excessive strictness can arguably lead to over-conformity and lack of creativity. Every society may choose the level of discipline that suits it culturally; but it is important to note that while high disciplinary levels foster quality learning in large classes, lower levels entail small classes, more - possibly less qualified - teachers, and more individual attention to students in place of actual teaching. Moreover, attentiveness and discipline make the teacher's job easier – more rewarding and less taxing - thereby affecting the segment of the population that is attracted to teaching and that remains in the profession.

Research into the role of discipline in education production has been greatly limited due to the lack of an objective measure for it: the cultural bias involved in the reporting of disciplinary atmosphere has thus far limited the discussion on its effect across school systems and countries, and with respect to other educational factors such as class size and teacher quality, to the theoretical realm.

This paper starts out by establishing a theoretical framework, based on Lazear (2001), which predicts student discipline to have two positive effects on education production: a direct effect of reducing disturbances and thus improving outcomes - this direct effect is predicted to be increasing in class size; and an indirect effect, of making larger class sizes optimal, thus reducing the number of teachers and increasing their quality.

Using truancy and tardiness data from the PISA (the OECD's Programme for International Student Assessment) 2012, 2015 and 2018 rounds, this paper constructs an index of school discipline that is far more objective (less culturally biased) than students' reports on classroom atmosphere. This index of objective discipline, which is shown to be strongly correlated to reported discipline within countries (where cultural bias is expected to be much smaller), is well correlated with PISA math scores, and is argued to be a major factor in explaining variation in educational outcomes across students, schools and countries.

In accordance with theory, the effect of discipline is found to increase in class size. This finding is of great relevance to the 'class size' literature, showing that the effect of class size is tightly connected to class discipline, and explaining how the top 5 performing developed countries – Singapore, Hong-Kong, Taiwan, Korea and Japan – benefit from high discipline by having exceptionally large classes, which allows them to be more selective in the choice of teachers. Some Western countries, such as Germany, Belgium, the Czech Republic, Iceland, the Netherlands and Switzerland, which enjoy high levels of discipline, only partially benefit from it due to their choice of small classes.

While results presented in this paper clearly point to student discipline as a crucial factor in explaining gaps in scholastic performance between students, schools and countries, there is much that remains to be studied. Of particular interest is the interaction between discipline and teacher quality, which is difficult to measure and which theory predict to be correlated with discipline.

References

Angrist, Joshua and Victor Lavy, "Using Maimonides' Rule To Estimate The Effect Of Class Size On Scholastic Achievement," *The Quarterly Journal of Economics*, 1999, *114* (2), 533–575.

Arum, Richard and Melissa Velez, *Improving learning environments: School discipline and student achievement in comparative perspective*, Stanford University Press, 2012.

Bosworth, Derek, "Truancy and Pupil Performance," *Education Economics*, 1994, *2* (3), 243–264.

Goodman, Joshua, "Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time," *NBER Working Paper*, 2014, (20221).

Gruber, Noam, "Why Israel does poorly in the PISA exams - perceptions versus reality," *Shoresh Institution Research Paper*, 2017.

Hanushek, Eric A., "The Failure of Input-Based Schooling Policies," *The Economic Journal*, 2003, *113*, F64–F98.

_ **and Ludger Woessmann**, "The Economics of International Differences in Educational Achievement," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 3, Amsterdam: North-Holland, 2011, pp. 89–200.

_ **and Steven G. Rivkin**, "Teacher Quality," in Eric A. Hanushek and Finis Welch, eds., *Handbook of the Economics of Education*, Vol. 2, Amsterdam: North-Holland, 2006.

_ **, John F. Kain, Daniel M. OBrien, and Steven G. Rivkin**, "The Market for Teacher Quality," *NBER*, 2005, (11154).

_ **, Marc Piopiunik, and Simon Wiederhold**, "Do Smarter Teachers Make Smarter Students? International Evidence on Teacher Cognitive Skills and Student Performance," *Education Next*, 2019, *19* (2), 57–64.

Hoxby, Caroline M., "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *The Economic Journal*, 2000, *113*, F64–F98.

Jepsen, Christopher and Steven Rivkin, "Class size reduction and student achievement the potential tradeoff between teacher quality and class size," *Journal of human resources*, 2009, *44* (1), 223–250.

Krueger, Alan B., "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, 1999, *114* (2), 497–532.

_ , "Economic Considerations and Class Size," *The Economic Journal*, 2003, *113* (485), F34–F63.

Lazear, Edward P., "Educational Production," *The Quarterly Journal of Economics*, 2001, *116* (3), 777–803.

A Additional Graphs

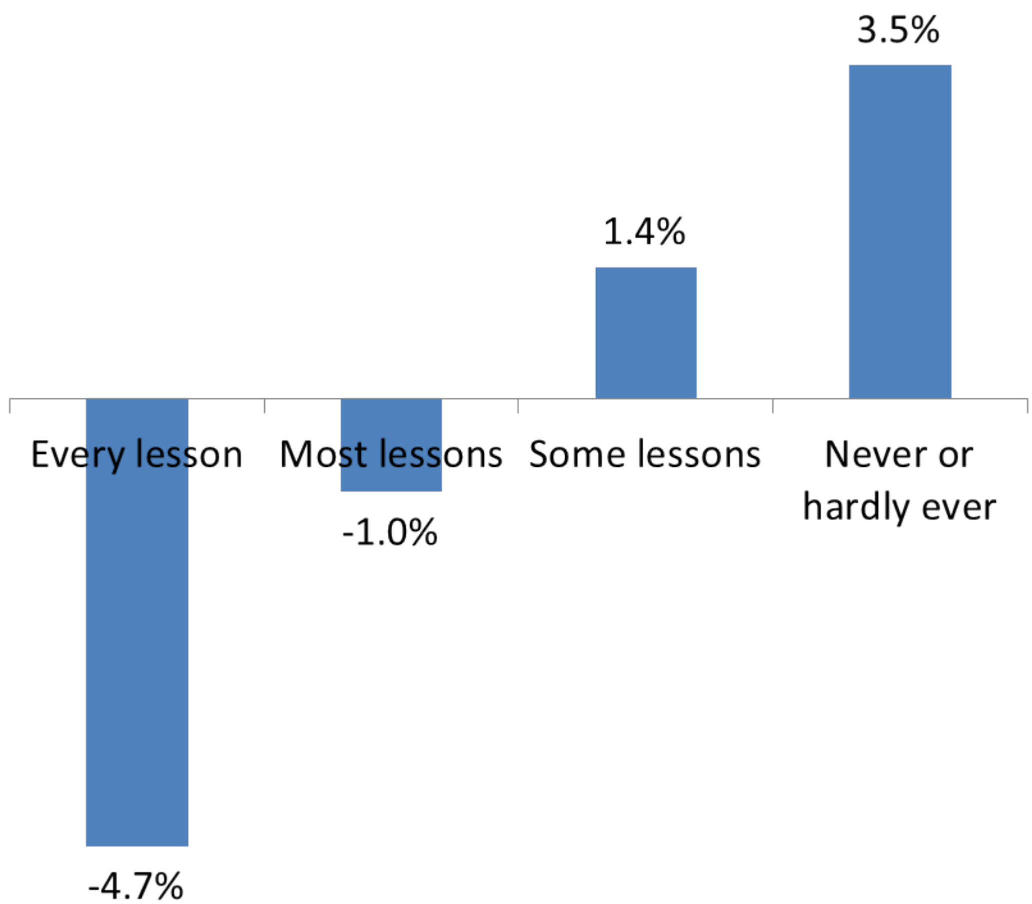

Figure 14: Percent difference versus country average, by response to statement "There is noise and disorder"

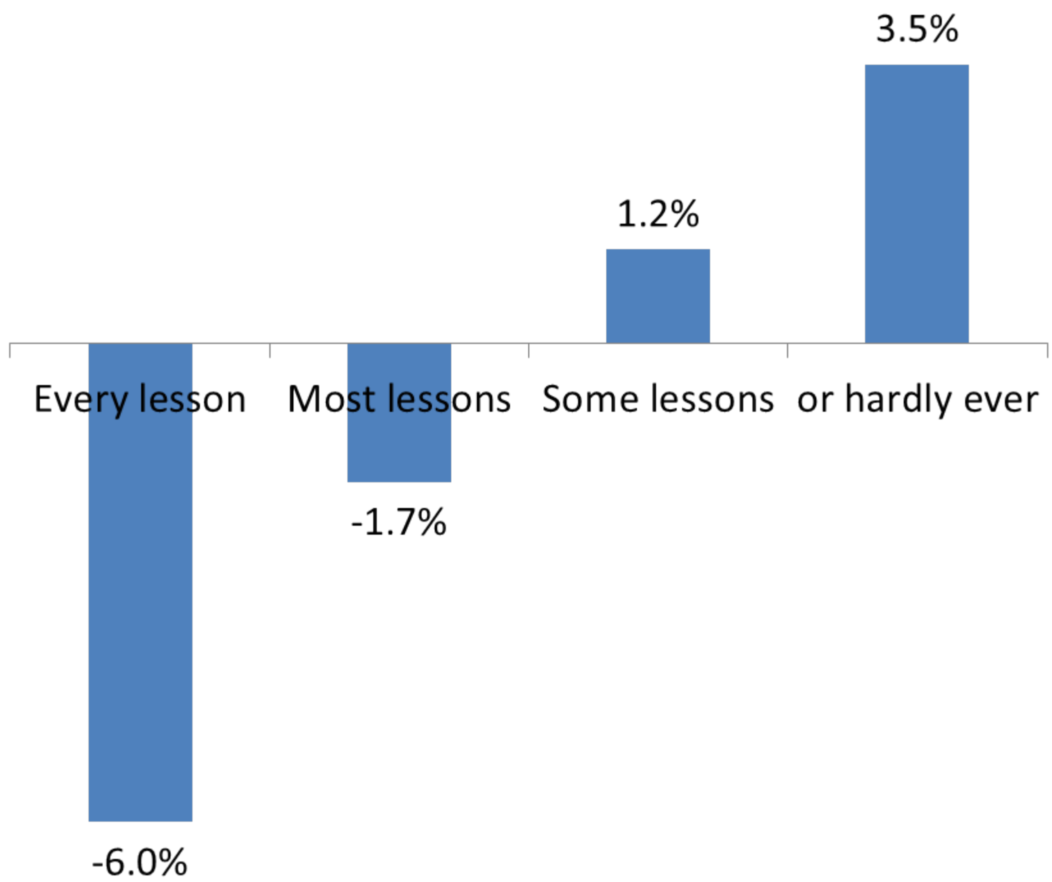

Figure 15: Percent difference versus country average, by response to statement “The teacher has to wait a long time for students to quiet down”

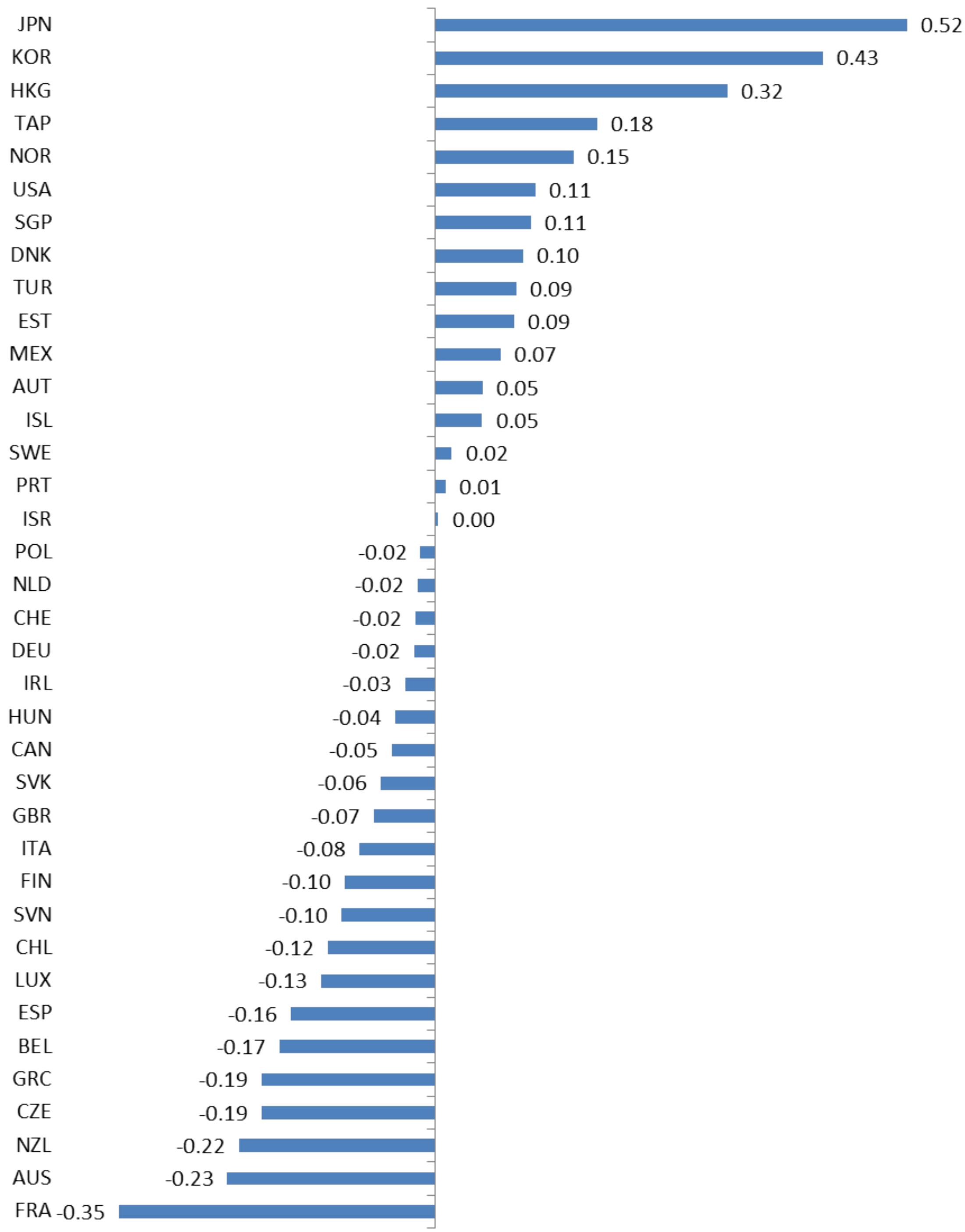

Figure 16: Subjective discipline index (2012, 2015 and 2018 round averages)

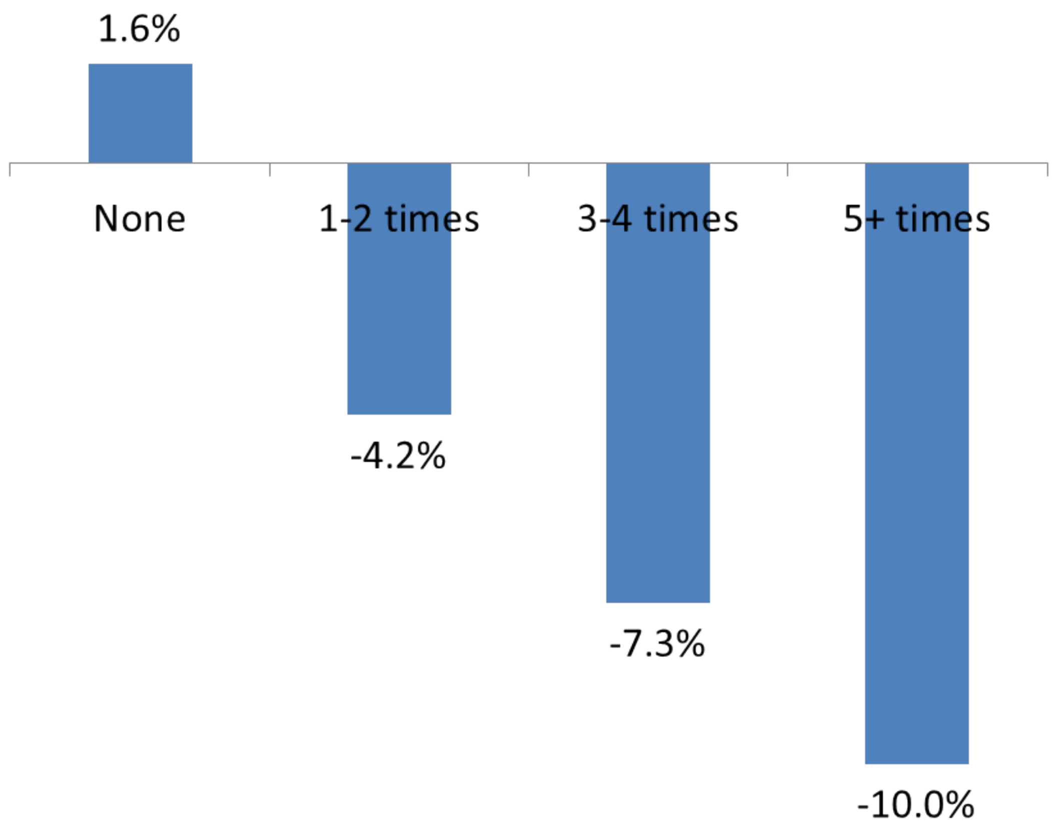

Figure 17: Percent difference versus country average, by whole days skipped

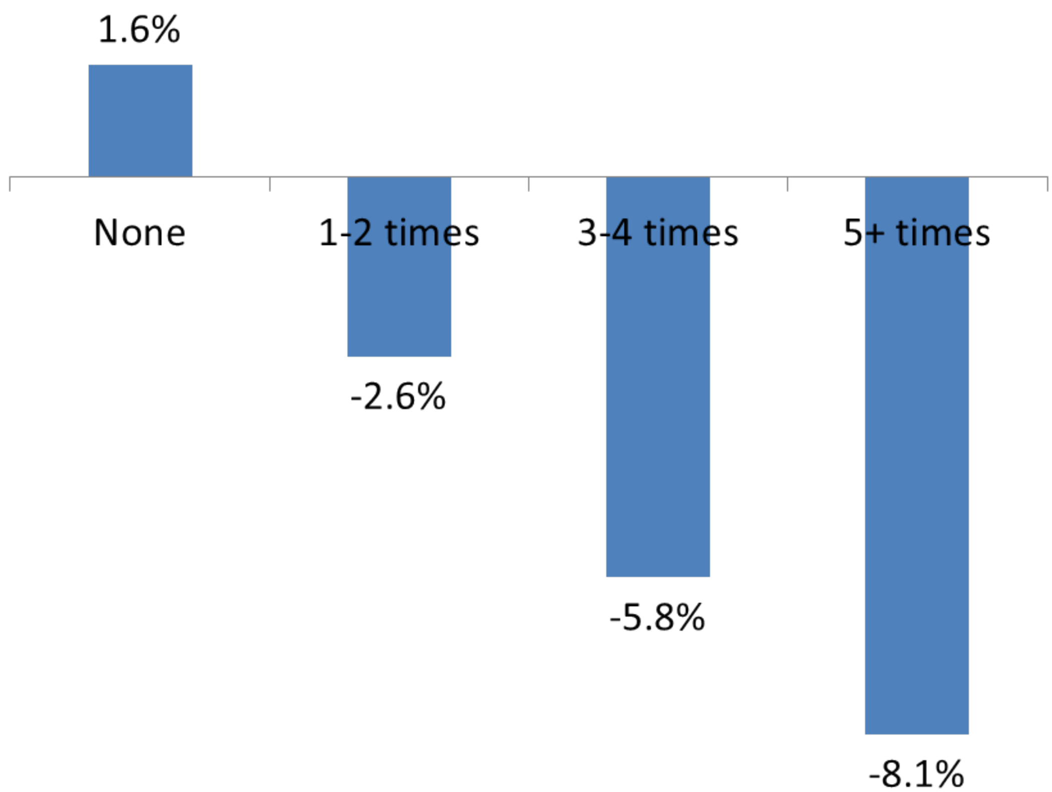

Figure 18: Percent difference versus country average, by classes skipped

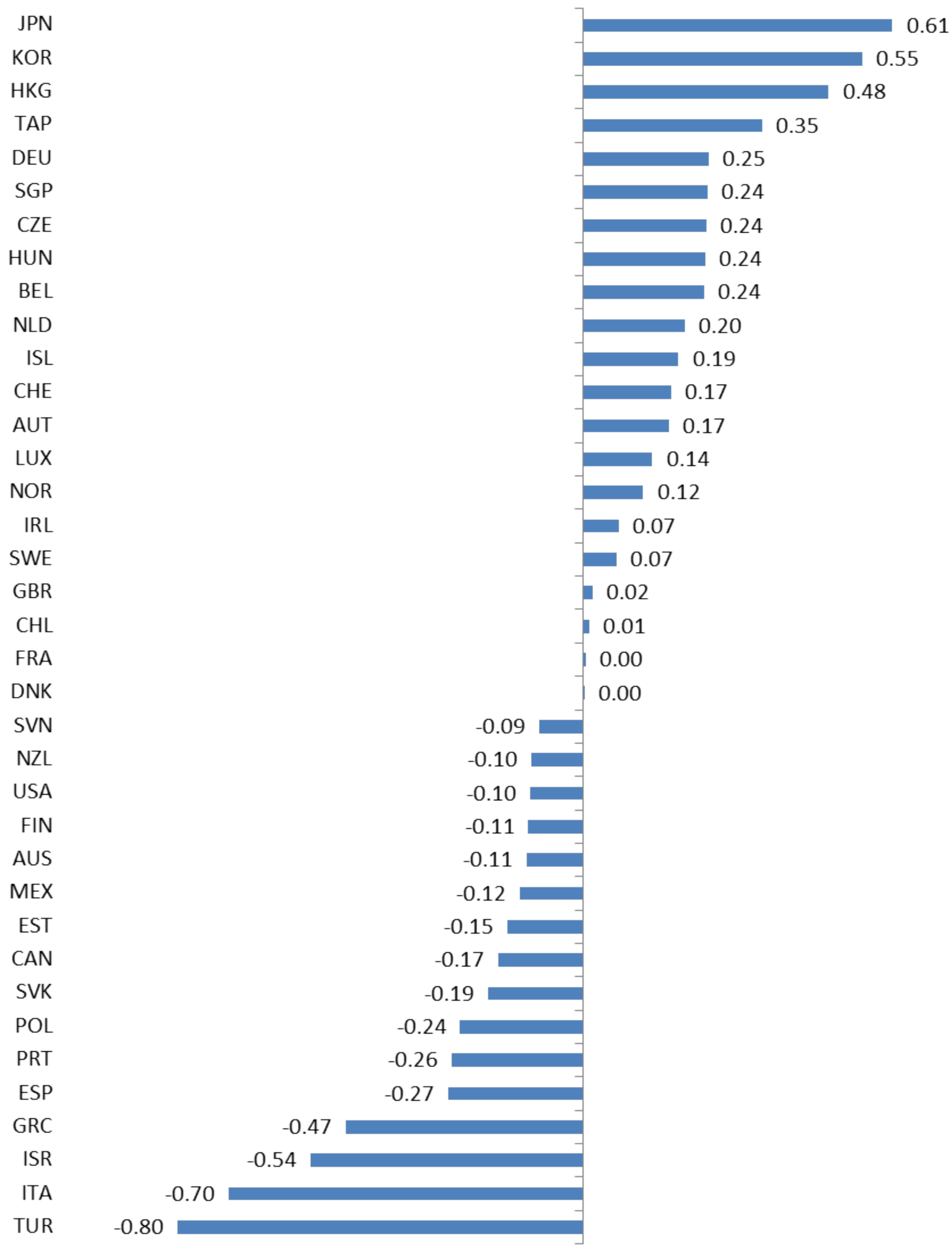

Figure 19: Objective discipline index (2012, 2015 and 2018 round averages)