School Discipline and Teaching Quality Across Countries: an Empirical View of Educational Production*

Noam Gruber†

September 23, 2025

Abstract

Formulating an education production function and using objective estimates of school discipline, this paper attempts to identify the relations between discipline, class size, teaching quality and scholastic outcomes. The data shows individual discipline to be a powerful predictor of the PISA math score, while variance in discipline among classmates has a strongly negative effect. Furthermore, school discipline is correlated with larger classes. Controlling for discipline, class size and parents' education, teaching quality is estimated as a latent variable. This analysis allows for the break-down of the contribution of each educational "production factor" to final student achievements.

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

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. I have no relevant or material financial interests that relate to the research described in this paper. The views expressed in this paper are mine alone and do not represent the views of the IMF, its Executive Board, or IMF management. All errors are my own.

†The International Monetary Fund. Email: ngruber@imf.org.

1 Introduction

Much like economic value creation, the educational process can also be formulated as a function, where the central factors of educational production would arguable be individual student ability, effort, peer effect and teaching quality. However, unlike factors of industrial production, the definitions and measurements of these educational production factors are not yet well established. This paper attempts to measure and use discipline as a factor of educational production (a proxy for effort at the individual level and for the peer effect at the class level). In addition, by establishing normative class parameters, teaching quality is estimated as a latent variable.

The role of discipline in explaining the differences in scholastic achievement levels between countries has thus far been under-explored, mostly due to the substantial cultural bias present in surveying 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), to construct a measure of school discipline that is "objective", i.e. valid across countries. This measure is well-correlated with subjective within-country measurements of discipline, and is shown to strongly correlate to student achievements.

Section 2 overviews the relevant literature. Section 3 presents a theoretical model. Section 4 describes the PISA data used in this paper. Section 5 focuses on the estimation of discipline, comparing disciplinary atmosphere, as subjectively reported by students, to a more objective measure of discipline, based on truancy and tardiness data. Section 6 contains econometric analysis of education production function, including an estimation of teaching quality as a latent variable. Section 7 concludes.

2 Literature Review

Previous 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, due to their culturally-biased nature. It is often the case in this literature that when differences across countries cannot be explained away using hard data, unmeasurable cultural differences are called upon (e.g. see Perera and Asadullah (2019) regarding Malaysia and Asadullah et al. (2020) regarding Vietnam).

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 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. In a relevant attempt, Zamarro et al. (2019) construct measures of student effort in answering the PISA test to explain score variation across countries. In Asadullah et al. (2021), student effort is found to play a significant role in explaining score variation among students in Bangladesh.

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, 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.[1] Lazear (2001) breaks this mold by modeling the trade-offs between class sizes and teaching quality, and how they depend on class discipline.

As both classroom discipline and teaching quality are hard to measure quantitatively (on measuring teaching quality, see Hanushek and Rivkin (2006)), empirical work on their co-relations is particularly scarce. Jepsen and Rivkin (2009) argues that higher discipline enables larger classes, thus reducing the number of teachers needed, vacating resources for teaching quality and allowing schools to be more selective vis-à-vis the teachers employed. 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.[2] The causal relationship between country- and school-level discipline and teacher quality are therefore potentially complex and bi-directional.

This paper employs a added-value approach, in which teacher quality is measured as a latent variable, i.e. a residual on the effect of other inputs,[3], with the addition of an explicit measure of discipline.

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

[2] For example, Hanushek et al. (2004) shows that student characteristics are important factors in the decisions of teachers whether to leave specific schools.

[3] See Hanushek et al. (2005) and Hanushek and Rivkin (2006) for more on this approach.

This paper contributes to existing economic literature by further exploring the concept of an education production function, 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, while estimating teaching quality as a latent variable.[4]

3 The Education Production Function

Equation 1 presents a student-level education production function:

$$E_{i,j} = Q_j^{\beta_0} P_i^{\beta_1} P_{-i,j}^{\beta_2} X_i^{\gamma_x} Z_j^{\gamma_z} \tag{1}$$

where $E_{i,j}$ is the educational product of student i in class j, Q_j is the quality of teaching at class j, P_i is student i's level of discipline ($0 \leq P \leq 1$), and $P_{-i,j}$ is the product of the level of discipline of the rest of the students in class j excluding student i, as seen in Equation 2 below. X_i is the set of student i's educational attributes other than discipline (e.g. talent, motivation and parental assistance), and Z_j is the set of such attributes of the other students in class j.

$$P_{-i,j} = \prod_{s=1}^{n_j - 1} P_{s \neq i} \tag{2}$$

where n_j is the number of students in class j.

The multiplicative form of $P_{-i,j}$ is inspired by Lazear (2001), which views $\bar{P}$ as the probability of the representative student not interrupting his/hers classmates, and $\bar{P}^n$ therefore as the probability of uninterrupted study in a classroom with n students. However, the Lazear (2001) does not take into account variation in classmates' discipline levels, unlike $P_{-i,j}$ in Equation 2.

Notice that Q_j, teaching quality in class j, is defined here as orthogonal to $P_{-i,j}$, class

[4]This paper employs a added-value approach in which teacher quality is measured as a residual on the effect of other inputs. See Hanushek et al. (2005) and Hanushek and Rivkin (2006) for discussion of this approach.

discipline. In other words, Q is teaching quality beyond inspiring/enforcing discipline.

The challenges for measuring teaching quality as a latent variable, and for estimating the education production function in general, are therefore to establish a credible measure of individual and class discipline (p_i and $p_{-i,j}$), individual and class attributes (x_i and z_j), and critically, a way to control for the possible endogeneity issues of the explanatory variables. Specifically, class size, class discipline and teaching quality may all be mutually correlated.

It is useful to understand the mechanisms behinds such correlations. Let us assume that school management seeks to maximize educational outcomes using the following two-step optimization procedure:

Firstly, it seeks to equalize discipline levels across classes, such that $P_j \simeq \bar{P}_J \forall j \in J$, where J is the school's entire grade (the 10th grade in the case of the PISA test), and $P_j = \prod_{s=1}^{n_j} P_s$, similar to Equation 2 but without excluding a specific student. Then, much along the lines of Lazear (2001)'s argument, the number of students in the class will be positively correlated with average class discipline ($N_j \propto P_j$).

Secondly, the school has access to discrete teaching resources (i.e. teachers) drawn from a distribution (i.e. $Q \sim (\bar{Q}_J, \sigma_{Q_J})$. After having roughly equalized overall discipline levels across classes, school management assigns higher quality teachers to larger classes. Teaching quality Q_j will then be positively correlated to class size n_j, class discipline P_j and class scholastic outcomes $\bar{E}_j = \frac{\sum_{i=1}^{N_j} E_{i,j}}{N_j}, i \in j$, i.e. $Q \propto N \propto P$.

As the empirical literature has rarely been able to measure teaching quality and/or class discipline, it is often found that class size is counter-intuitively correlated with students' performance. By empirically estimating discipline, this paper seeks disentangle this selection bias and back out Q as a latent variable. Given perfect information, we would have liked to estimate the following model, based on Equation 1 in log form:

$$e_{i,j} = \alpha + \beta_0 q_j + \beta_1 p_i + \beta_2 p_{-i,j} + \gamma_x x_i + \gamma_z z_j + \varepsilon_{i,j} \tag{3}$$

Note that $p_{-i,j}$ is the log form of $P_{-i,j}$, hence $p_{-i,j} = \sum_{s=1}^{n_j - 1} p_{s \neq i} = (n_j - 1)\bar{p}_{-i,j}$.

Then, provided that we have good measures for discipline, and if we assume the model to be well-specified, and (critically) break the endogeneity between q and the rest of the explanatory variables, we can estimate the below model:

$$e_{i,j} = \alpha + \nu_j + \beta_1 p_i + \beta_2 (n_j - 1)\bar{p}_{-i,j} + \gamma_x x_i + \gamma_z z_j + \varepsilon_{i,j} \tag{4}$$

where, iff the above assumptions hold, $\hat{\nu}_j$ is a good estimate of $\beta_0 q_j$, a linear transformation of teaching quality at class j.

4 PISA Data: 2012, 2015, 2018 and 2022 Rounds

The PISA test is administered to 15 year olds[5] every three years (the round planned for 2021 was postponed to 2022 due to the COVID-19 pandemic).[6] This work uses data from the 2012, 2015, 2018 and 2022 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.[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 Item Response Theory (IRT) methodology. In addition to the test questions, PISA respondents, as well as school administrators, also fill out survey questionnaires, which provide many details on the students' family background, school attributes, attitudes etc. The present study uses data for the 34 OECD member countries,[8] as well as Singapore, Hong-

[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.

Kong[9] and Taiwan. For these 37 countries the data comprises 306,420 test-takers in 2012, 257,634 test-takers in 2015, 289,543 test-takers in 2018 and 286,980 test-takes in 2022.

4.1 PISA Performers: top and bottom Countries

Figure 1 presents the math scores by country, averaged across the PISA test 2012, 2015, 2018 and 2022 rounds. The 36 developed countries that appear in the figure are the 33 OECD countries (as of 2012, minus Luxembourg due to its size) 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 last decade of the 20th century. Both Israel and Singapore are small, isolated, multi-ethnic countries. This paper aims to identify the role of educational production factors in the scholastic success of some countries, and the failure of others.

5 Subjective Versus Objective Discipline

5.1 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

[9]Hong-Kong is 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 country.

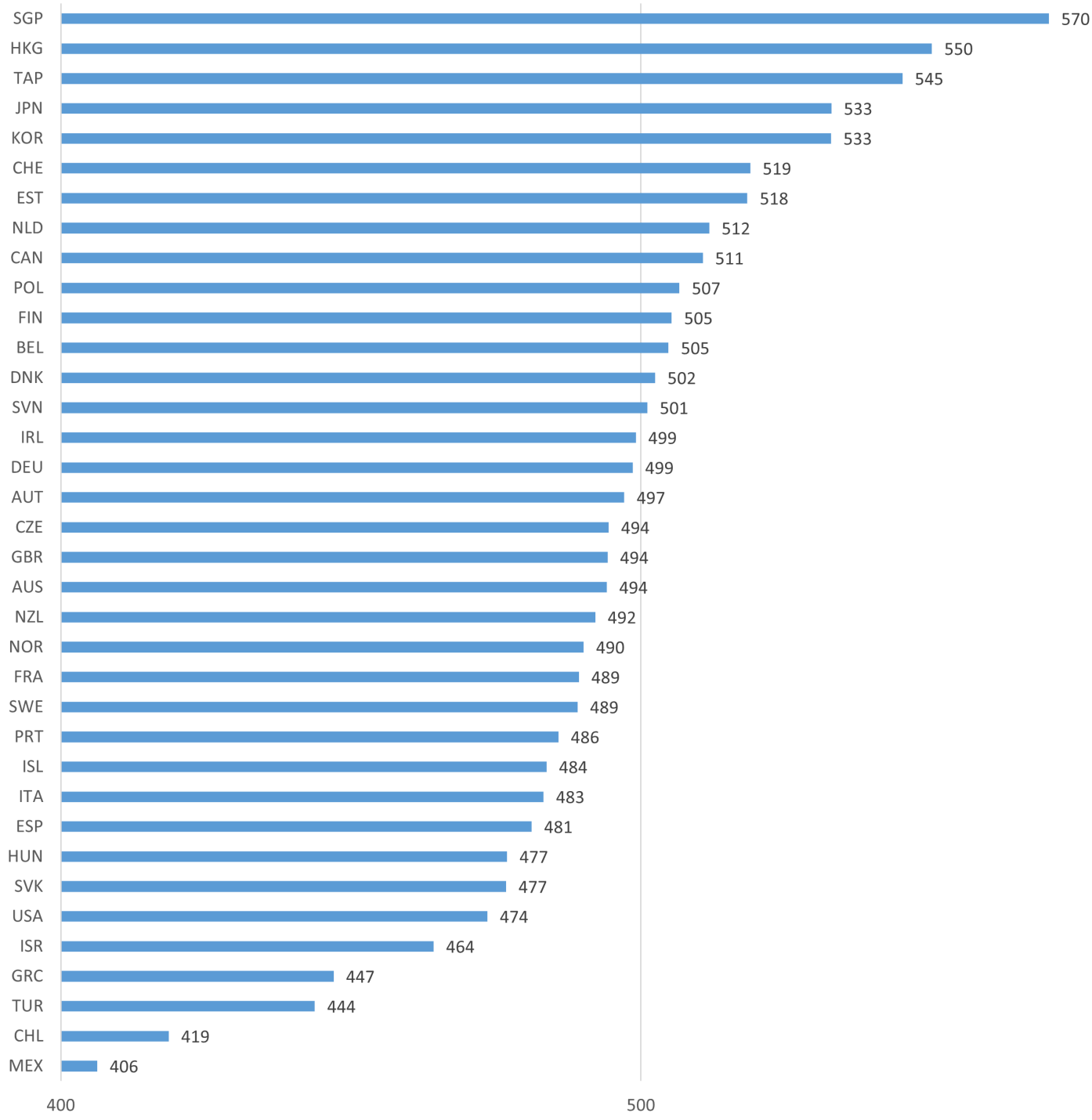


Figure 1: PISA math score by country, averages of the 2012, 2015, 2018 and 2022 test rounds

lesson," "Most lessons," "Some lessons," and "Never or hardly ever."

Figure 2 demonstrates the correlation between the perceived disciplinary climate reflected by the response to these statements 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.

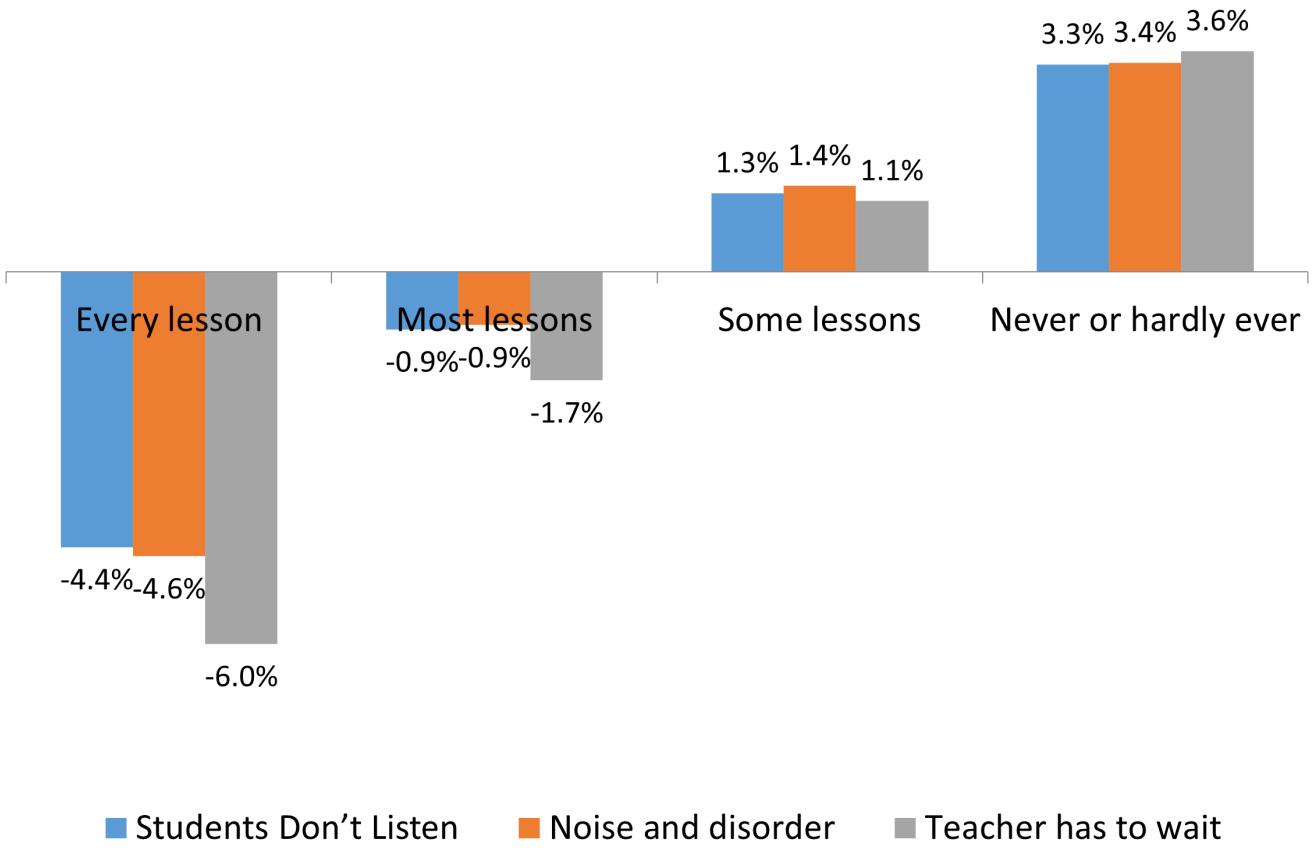


Figure 2: Percent difference from country average math score, by subjective discipline measure

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 10 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, 2018 and 2022 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 and Spain) 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 Belgian students report a weak one.

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

Unlike the subjective discipline measure, based on students' perceptions, the objective discipline measure is based on truancy and tardiness data, which is likely much more comparable across countries.[10] 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 individual and school motivation and discipline levels.[11] The PISA questionnaire questions students about the number of times they arrived late to school, the number

[10] It should be noted that truancy and tardiness are self reported and may still contain some cultural bias.

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

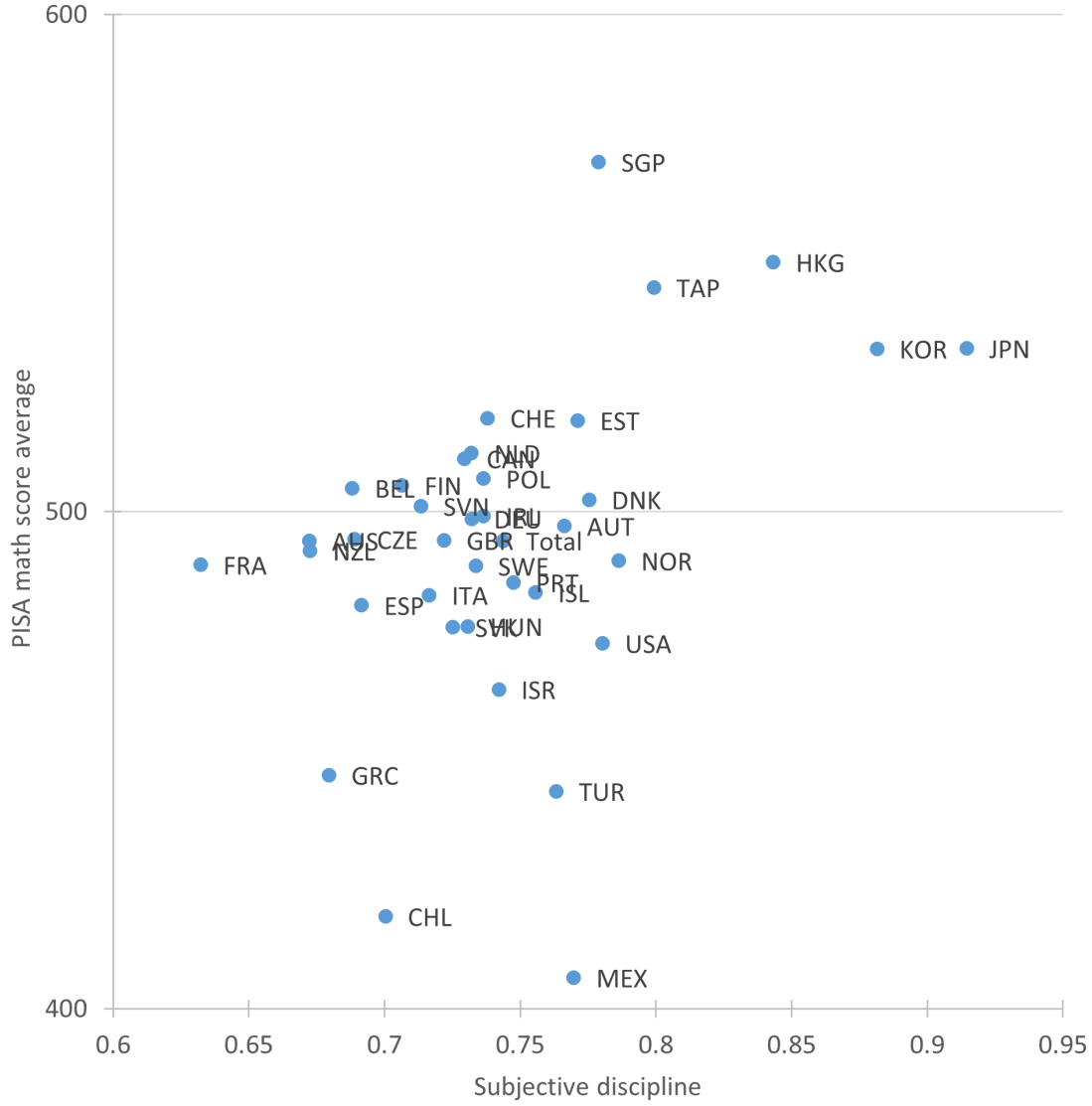


Figure 3: PISA score versus subjective discipline, by coutry (2012, 2015, 2018 and 2022 test rounds average)

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 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.

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 11 in Appendix A.

Figure 5 displays the 'objective discipline' index versus PISA math score (both averaged between the 2012, 2015, 2018 and 2022 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 Poland, 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 normative level of discipline, 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.

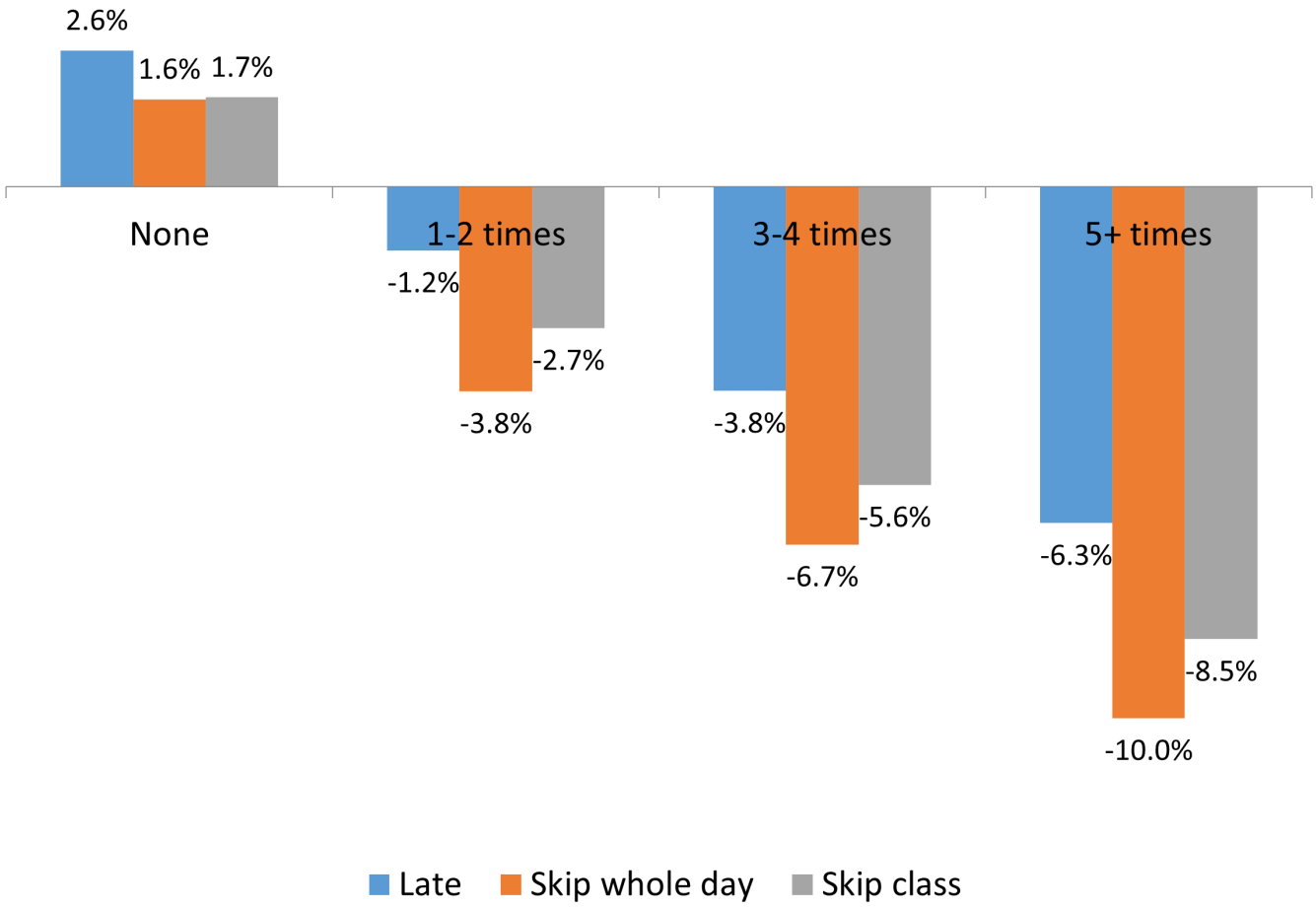


Figure 4: Percent difference from country average math score, by objective discipline measures (during the last two weeks)

5.3 The Relation Between Objective and Subjective Measures of Discipline

As Figure 6 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

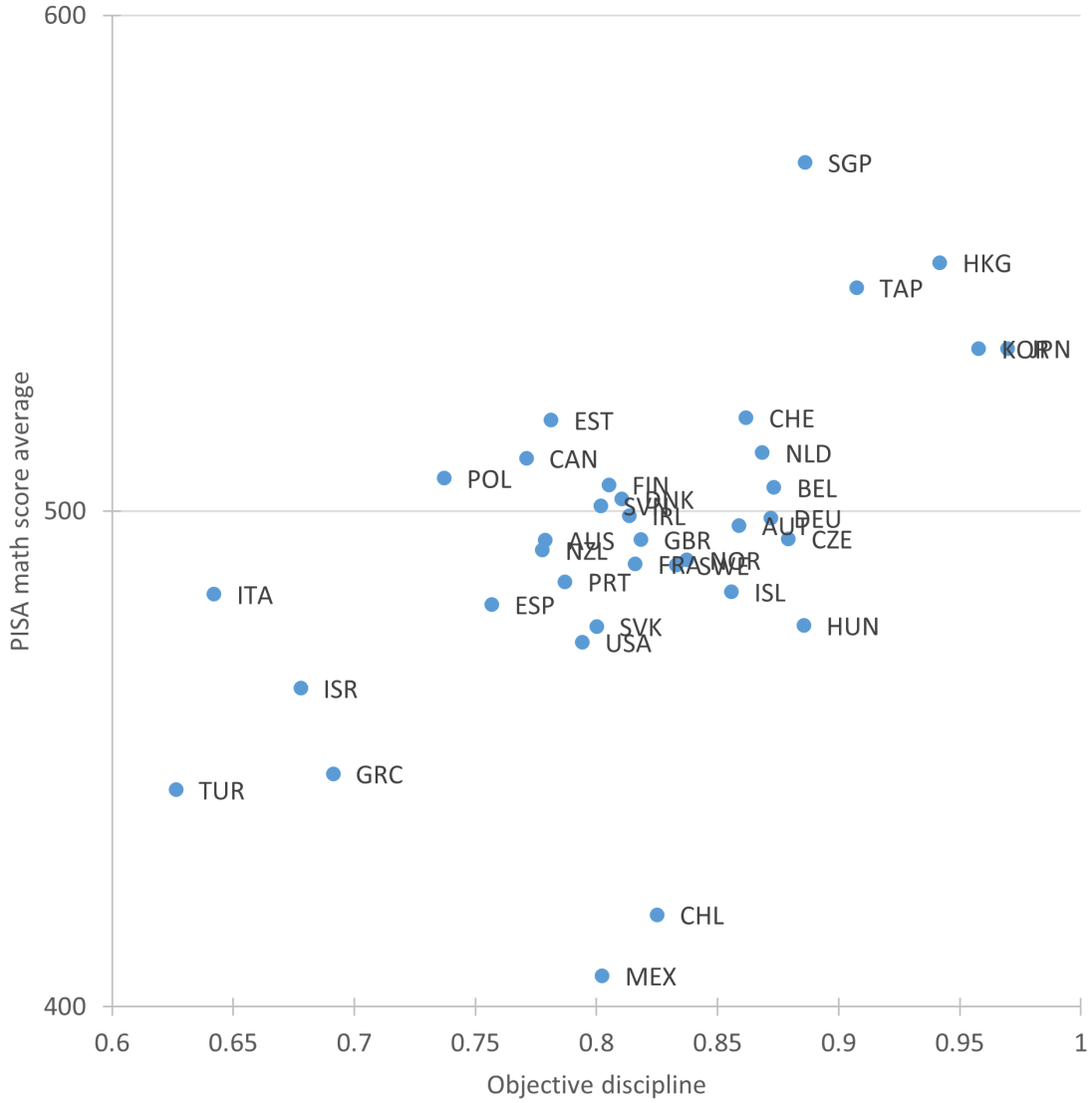


Figure 5: PISA score versus objective discipline (2012, 2015, 2018 and 2022 test rounds average)

Table 1: Subjective versus objective discipline

	(1)	(2)
Obj. Disc.	.337	.327
	(.051)***	(.030)***
Const.	.453	.401
	(.044)***	(.025)***
Country FE	No	Yes
Obs.	42157	42157
R^2	.076	.169

Notes: Errors clustered by country and year fixed effects included for both regressions.
*Significant at 10%; **significant at 5%; ***significant at 1%

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, 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 1 presents regressions of the 'subjective discipline' index, on the left hand side, versus the 'objective discipline' index, on the right hand side, at the school level. As can be seen in Column (1), the objective discipline index coefficient is significant at the 1% level (a t-value of 6.62). However, with $R^2 = 0.076$, the explanatory power is fairly weak. Column (2) adds country fixed effect, to account for the cultural biases in the perception of discipline. The t-value increases substantially, to 10.98, and R^2 jumps to 0.169. Clearly, with cultural bias accounted for by country FE, truancy and tardiness data provide strong indications for perceived disciplinary climates in schools.

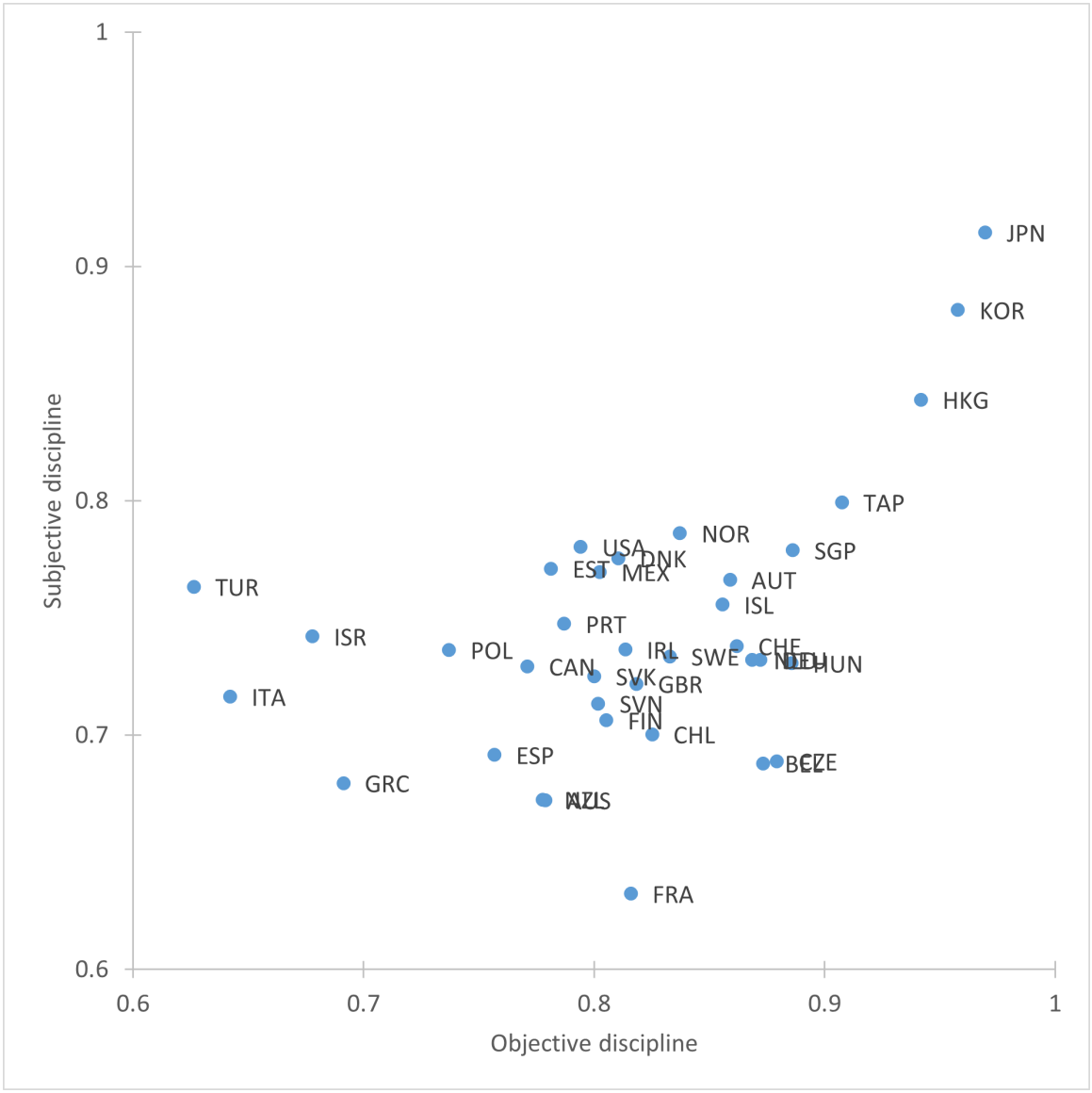


Figure 6: Subjective versus objective discipline, by country (2012, 2015, 2018 and 2022 test rounds average)

5.4 Comparing the Correlation Between PISA Math Scores and Objective and Subjective Discipline

Table 2 uses naïve OLS regressions to show the correlation between PISA math scores (on the left-hand side) and subjective and objective discipline measures (on the right-hand side) at the school level. Columns (1) and (2) include subjective discipline while columns (3) and (4) include objective discipline. The even-numbered columns also control for country fixed effects. As Column (1) indicates, subjective discipline is well correlated with with scores, with a t-value of 7.14. However, at 0.087, the R^2 indicates it does not explain much of the variability. Column (2) adds country fixed effects, addressing cultural bias with regard to the perception of discipline at the country level. Compared to Column (1), the t-value of subjective discipline increases to 10.07, and the R^2 jumps to 0.296. Column (3) examines the correlations between objective discipline and PISA math scores. The t-value is 11.13 and the R^2 is 0.157, both substantially higher than in Column (1), indicating that objective discipline has more explanatory power. Column (4) adds country fixed effects. The t-value increases marginally to 11.32, while the R^2 jumps to 0.327.

The fact that objective discipline performs much better than subjective discipline without country fixed effects (columns 1 and 3), but that much of the gap closes when adding country fixed effect (columns 2 and 4) clearly demonstrates the advantage of using objective discipline when comparing the effect of discipline across different countries and cultures, among which the norms and perceptions of discipline can be very heterogeneous.

5.5 Discipline, Class Size and Student Achievement

The theory presented in Section 3 predicts class size to have a magnifying effect on the relation between discipline and student achievements (see Equation 2. In that sense, for a given level of discipline and teaching quality, we would expect an increase in class size to lower test results. However, the model in Section 3 also assumes that school management

Table 2: PISA score versus objective and subjective discipline measures

	(1)	(2)	(3)	(4)
Subj. Disc.	122.139	117.368		
	(17.116)***	(10.066)***		
Obj. Disc.			204.473	203.046
			(18.377)***	(17.943)***
Const.	395.497	415.906	314.273	327.981
	(10.510)***	(6.941)***	(17.606)***	(15.145)***
Country FE	No	Yes	No	Yes
Obs.	42491	42491	42836	42836
R^2	.087	.296	.157	.327

Notes: Errors clustered by country and year fixed effects included for all regressions.
*Significant at 10%; **significant at 5%; ***significant at 1%

adjusts student number and allocation in order to equalize discipline levels across classes, and then allocates more resources (i.e. better teachers) to the larger classes. Such optimizations would yield a positive relation between discipline and class size, as well as a positive relation between class size, teaching quality and educational outcomes.

Figure 7 presents the objective discipline level versus class size. As can be seen, the 5 leading Asian countries 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 Austria, Belgium, the Czech Republic, Germany, Iceland, the Netherlands, Hungary and Switzerland enjoy relatively high levels of student discipline. Arguably, it could be more optimal for them to have larger classes, which would allow for increasing teaching quality through greater selectivity.

Of the low-performing countries, Chile and Mexico have large classes, but only a medium level of student discipline, and Turkey, Israel, Greece and Italy 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. Of additional concern would be the ability of educational systems with large classes and low levels of discipline to attract and maintain high quality teaching staff.

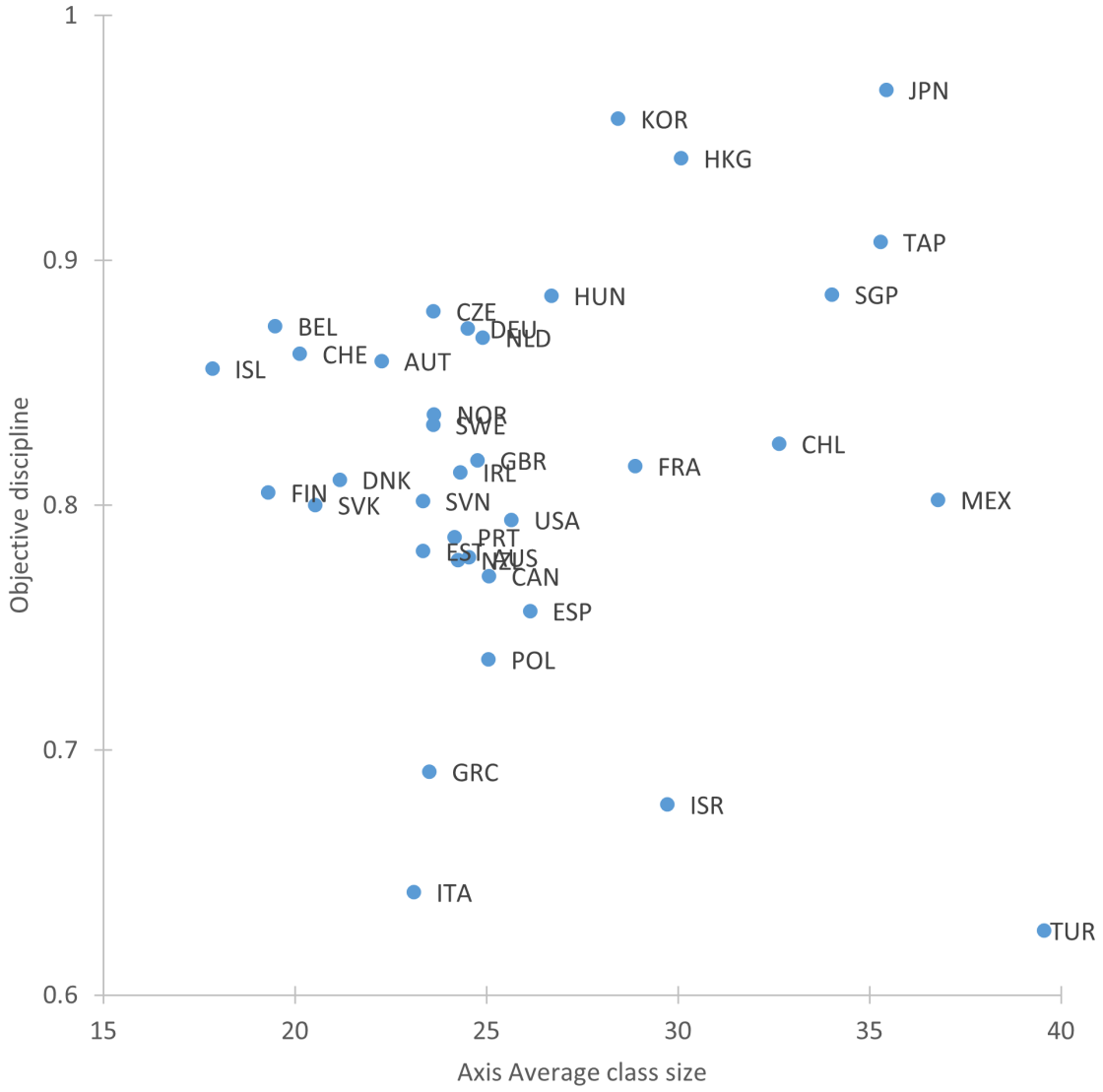


Figure 7: Objective discipline versus class size, by country (2012, 2015, 2018 and 2022 test rounds average)

Table 3: Class size versus objective discipline

	(1)	(2)	(3)	(4)
Obj. Disc.	3.451	10.794	3.201	29.528
	(4.809)	(21.722)	(1.132)***	(9.166)***
Obj. Disc. squared		-4.944		-18.161
		(15.557)		(6.174)***
Const.	24.117	21.488	23.832	14.729
	(4.272)***	(7.367)***	(1.057)***	(3.522)***
Country FE	No	No	Yes	Yes
Obs.	37593	37593	37593	37593
R^2	.01	.01	.346	.348

Notes: Errors clustered by country and year fixed effects included for all regressions.
*Significant at 10%; **significant at 5%; ***significant at 1%

It is clear from Figure 7 that at the country-level class size is far from fully-determined by the level of country-average discipline. It can be argued that budget constraints, demographic trends and cultural preferences provide a country-level variation in class-size that is to a large degree exogenous to country-level discipline levels.

Columns (1) and (2) of Table 3 also reflect this country-level lack of correlation. Without country fixed effects objective discipline (alone or with a squared term) is not significantly correlated with class size. However, when columns (3) and (4) add country fixed effects, making this a within-country estimation, the correlation is very strong, with higher student discipline correlated to larger classes.

Figure 8 presents class sizes and PISA math scores across countries (2012, 2015, 2018 and 2022 averages). Clearly, class size by itself is not a good indicator of PISA scores at the country level. PISA scores in Iceland and France are very close 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 fairly similar class sizes, yet their PISA scores differ a lot. That is also the case of Mexico versus Japan and Taiwan. As shown later, some of these differences be explained by the different levels of discipline in these countries, while some can be attributed to differences in teaching quality.

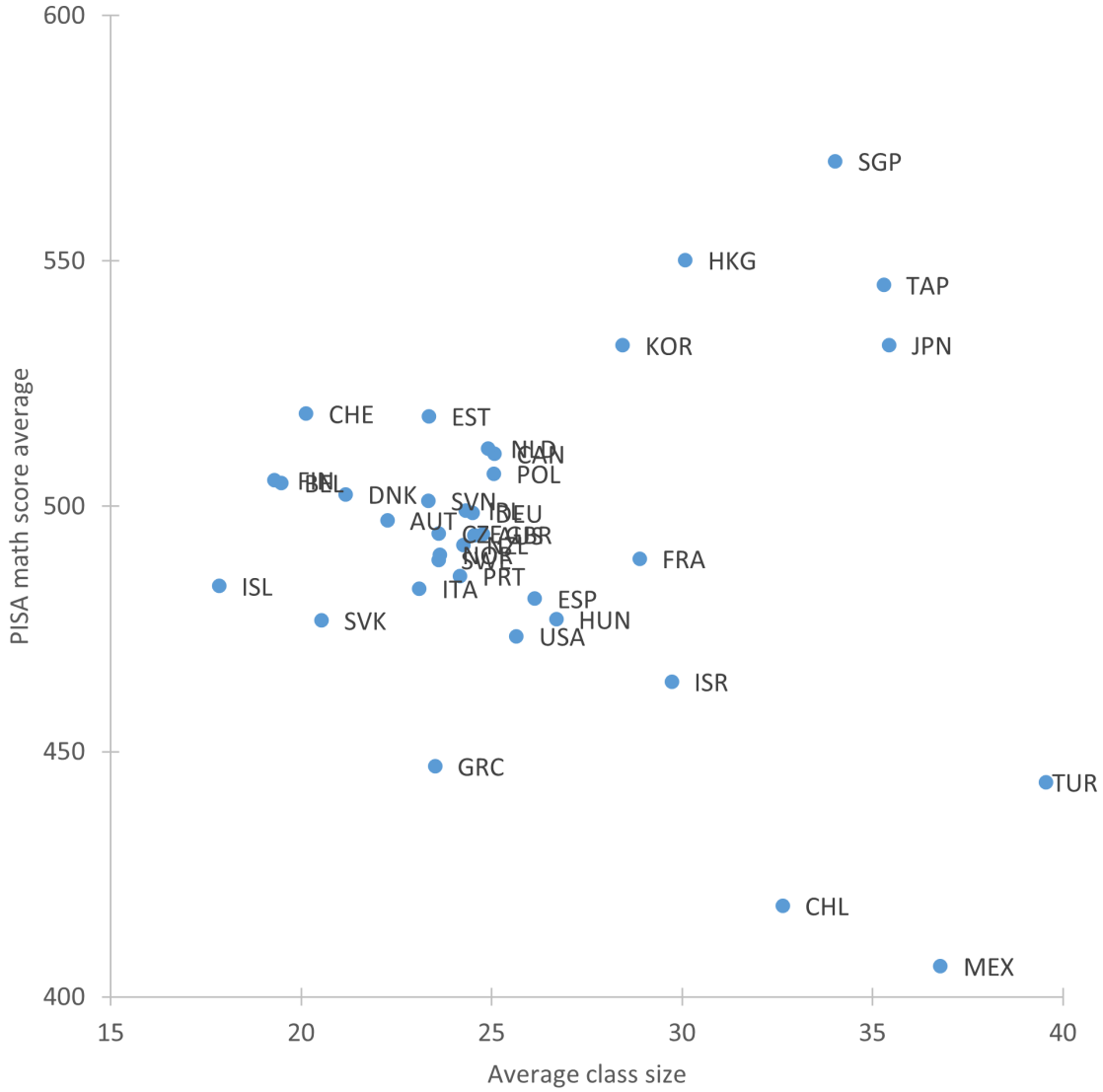


Figure 8: PISA math scores versus class size, by country (2012, 2015, 2018 and 2022 test rounds average)

Table 4: PISA math score versus class size

	(1)	(2)	(3)	(4)
Class size (cat.)				
16-20	24.730	19.385		
	(6.698)***	(4.982)***		
21-25	36.885	35.549		
	(7.648)***	(6.376)***		
26-30	51.026	53.086		
	(8.307)***	(7.854)***		
31-35	55.979	68.904		
	(11.985)***	(11.433)***		
36-40	59.134	63.945		
	(20.081)***	(11.251)***		
41-45	26.225	66.107		
	(24.987)	(11.065)***		
46-50	1.459	58.297		
	(15.306)	(10.612)***		
51+	-.249	46.019		
	(10.335)	(7.834)***		
Class size (cont.)			9.190	7.702
			(1.411)***	(1.160)***
Class size squared			-.140	-.097
			(.023)***	(.016)***
Country FE	No	Yes	No	Yes
Obs.	37993	37993	37993	37993
R^2	.073	.298	.069	.296

Notes: Errors clustered by country and year fixed effects included for all regressions.
*Significant at 10%; **significant at 5%; ***significant at 1%

Table 4 examines the correlation between class size and the PISA math score at the school level. Columns (1) and (2) use class size categories while columns (3) and (4) use a continuous class size polynomial. Columns (1) and (3) do not incorporate country fixed effects while columns (2) and (4) do. Other than for class size categories '41-45', '46-50' and '51+' in Column (1), the correlation between class size and PISA math scores is statistically significant at the 1% range for all specifications. For Column (1), a class size of '36-40' is correlated with highest scores. For Column (2) that is class size '31-35'. For the polynomials in columns (3) and (4) that is 32.8 and 39.7 respectively.

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 if class size is

changed *Ceteris Paribus*. This is a partial equilibrium analysis, as it does not take into account the system-level effect of changes in average class size on budget needs and on the quality of teachers. Also absent from much of the 'class size' literature is the discussion on the role of discipline over the effect of class size on students' achievements.

6 Estimating the Education Production Model with Latent Teaching Quality

The model presented in Section 3 predicts discipline to be correlated with class size, and the two to be correlated with unobserved teaching quality, and therefore with student outcomes. This framework yields a positive relation between discipline and class size, a positive relation between discipline and student outcomes (both directly and through correlation with unobserved teaching quality) and a positive relation between class size and student outcomes (purely through the correlation with teaching quality, whereas the negative effect of class size is offset by higher discipline through selection), all evident in the data. Given that teaching quality is not observed directly, it will be treated here as a latent variable, incorporated into the residual of the estimations, much like in Hanushek et al. (2005) and Hanushek and Rivkin (2006).

As teachers in PISA data are not allocated randomly, or according to any well defined rules, and there are no clear shocks to the allocation process to serve as natural experiments, a different approach is required to disentangle teaching quality from class size. This paper opts for setting the class size constant, i.e. choosing a sample of the data with classes in an identical range. Since the most frequent class size categories in the data are '21-25' (280,456 observations) and '26-30' (236,845), these are the categories used. These are also the most frequent median class sizes across countries (see Table 6 in the Appendix).

The use of this methodology for estimating latent teaching quality has a clear drawback: for countries where these class sizes are not the most 'normative', the estimated latent teach-

ing quality may not be fully representative. To address this concern results are compared to estimations using the full sample, controlling for class size, as well as fixed effects at the country and region (PISA strata) levels.

Table 5 presents the results of 9 OLS regressions estimations of student-level PISA math scores, divided into 3 sets: first using the full student sample while controlling for class size, second using only student in classes of size '21-25', and third using only students classes of size '26-30'. For each set, three regressions are estimated: no location fixed effects, country fixed effects and region (PISA strata) fixed effects.[12] All regression control for the students' maternal and paternal education levels (corresponding to x_i in Equation 4), average post-secondary education of schoolmates' parents (corresponding to z_j in Equation 4) and test year. The main explanatory variables in all 9 regressions are the student's own level of objective discipline, schoolmates average objective discipline (not including the student), and the school's standard deviation of discipline.

In all 9 regressions the student's own level of discipline is statistically significant at the 1% level and positively correlated with the student's PISA math score, with a coefficient ranging from 51 in Column (1) to 55 in Column (9) (recall that the discipline score is normalized to the 0-1 range). This result is easy to interpret: not only do students who are late to classes or skip them altogether miss out on study time, but this is also highly indicative of their level of commitment and effort.

Average school discipline is positive on all 9 regressions, though the scale and the level of statistical significance differ. It is statistically significant at the 1% level in Columns (3), (6) and (9), indicating average school discipline is well correlated with test scores within school regions, but less so between them.

[12] 792 strata in 2012, 608 strata in 2015, 588 strata in 2018, and 446 strata in 2022.

Table 5: Regressions on student-level PISA math scores

	All Class Sizes			**21-25 Class Size**			**26-30 Class Size**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Obj. Disc. (student)	51.342	52.192	52.597	53.115	54.267	54.759	54.717	54.717	55.001
	(4.608)***	(3.832)***	(3.862)***	(4.208)***	(3.102)***	(3.366)***	(3.999)***	(3.576)***	(3.628)***
Obj. Disc. (school)	42.798	58.373	68.347	48.556	76.273	85.908	43.426	36.707	45.325
	(27.085)	(32.452)*	(25.315)***	(28.284)*	(40.905)*	(29.037)***	(32.693)	(26.652)	(19.848)**
Obj. Disc. SD (school)	-136.901	-101.345	-70.569	-100.959	-61.666	-33.127	-143.602	-96.712	-69.799
	(37.647)***	(32.961)***	(24.644)***	(31.154)***	(36.607)*	(22.931)	(39.129)***	(30.249)***	(23.370)***
Class size									
16-20	7.836	6.784							
	(3.104)***	(2.134)***	(1.534)***						
21-25	18.794	13.989	12.472						
	(4.020)***	(2.931)***	(2.586)***						
26-30	25.560	21.803	17.228						
	(4.668)***	(3.565)***	(2.845)***						
31-35	25.378	29.058	20.908						
	(8.865)***	(4.462)***	(3.380)***						
36-40	27.122	32.241	23.636						
	(13.705)**	(4.791)***	(4.740)***						
41-45	6.265	33.569	29.433						
	(15.091)	(5.290)***	(5.377)***						
46-50	-6.172	24.927	20.194						
	(11.558)	(7.638)***	(5.670)***						
51+	6.697	14.426	16.804						
	(10.419)	(6.226)**	(3.437)***						
Const.	326.541	310.591	298.107	331.584	292.906	279.234	360.895	352.327	334.848
	(29.746)***	(35.506)***	(28.054)***	(26.844)***	(41.712)***	(27.787)***	(36.415)***	(27.969)***	(20.169)***
Location FE	None	Country	Region	None	Country	Region	None	Country	Region
Obs.	885131	885131	885131	280456	280456	280456	236845	236845	236845
R^2	.226	.312	.363	.184	.25	.311	.18	.268	.331

Notes: Errors clustered by country and year fixed effects included for all regressions. Controls (variables used but not displayed): maternal and paternal education and share of schoolmates whose parents have post-secondary education.
*Significant at 10%; **significant at 5%; ***significant at 1%

Perhaps of greater interest than average school discipline is it's distribution, as represented by the standard deviation. As Equation 2 argues, class discipline is multiplicative. In other words, a student with an extremely low level of discipline (e.g. close to zero) would have an out-sized effect on class atmosphere (all it takes is one student constantly interrupting). The school discipline standard deviation is consistently negative across all 9 regressions, in line with theory (the greater the variance the larger the chance of a very-low discipline student in the class). Scale-wise, it is largest for the specifications with no location fixed effects - columns (1), (4) and (7), smaller for the specifications with country fixed effects - columns (2), (5) and (8), and smallest for the specifications with regional fixed effects - columns (3), (6) and (9). This indicates that the second moment of discipline is well correlated with test scores across countries and regions. The school discipline standard deviation is significant at the 1% level for all specifications except Column (5) - 21-25 class size with country fixed effects - where it is significant at the 10% level, and Column (6), where it is not significant even at that level.

For the first set of regression - columns (1)-(3) - which control for class size, it is interesting to note that the coefficients of the class size dummies are much smaller (about half the size) compared to Table 4. This is clearly the result of including discipline in the regression, as class size is correlated with discipline within locations (see Table 3). As argued in Section 3, the positive correlation between class size and scores is a result of the correlation between class size, discipline levels and teaching quality. Including discipline in the regression therefore controls for some of this bias.

The regressions in Table 5 enable estimating teaching quality as a latent variable, as outlined in Equation 4. Figure 9 presents teaching quality by country, extracted as country fixed effects from the regressions in columns (2), (5) and (8). Australia (the first country alphabetically) serves as the zero baseline.

While there are some exceptions, the teaching quality estimate is highly correlated between the three specifications. It is interesting to note that while for some of the countries

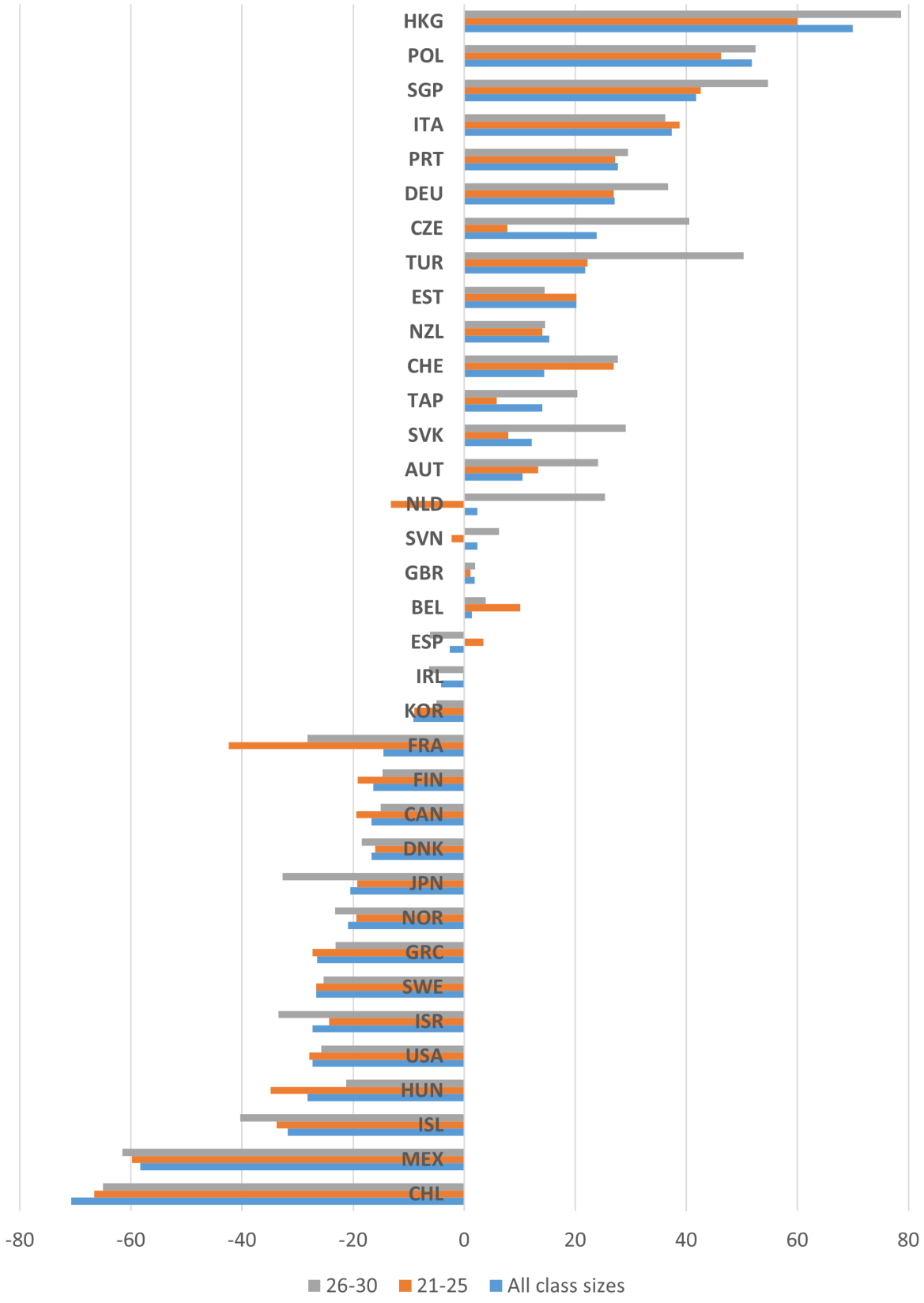


Figure 9: Estimates of teaching quality by country (Australia = 0 as baseline)

leading in PISA math scores teaching quality is very high (Hong Kong and Singapore), for some it is quite low (Korea and Japan). In other words, since teaching quality is estimated as the country average of the residual, Hong Kong and Singapore score well above what the discipline and parents' education levels would predict, while Korea and Japan score below. Likewise, the lagging countries also differ: Chile and Mexico, as well as Israel and Greece, are at the bottom in terms of teaching quality (scoring below their potential according to discipline and parents' education), while Turkey and Italy enjoy high levels of teaching quality.

This form of estimation is susceptible to multiple potential problems: first, as discussed earlier, while the '21-25' and '26-30' class size categories are generally the most common, that is not true for all countries. As such limiting the sample to these class sizes might bias the results where these sizes are not normative. In addition, parents' education, which is controlled for, might have different effect across locations. For example, where academic education is rare it might be more meaningful to have a parent with such an education. Finally, while the objective discipline measure was constructed in order to avoid the cultural bias associated with subjective student impressions, it cannot be argued to be perfectly culture-neutral. For example, in an environment where being late to class is normative, a student can be sometimes late but still maintain high motivation and high work ethic. That is less likely in an environment where being late is not normative and is thus more indicative of attitude problems.

7 Discussion and Conclusion

This paper starts out by establishing a theoretical 'education production' framework, largely based on Lazear (2001), which predicts student discipline, class size and teaching quality to be positively correlated. The paper then proceeds to use truancy and tardiness data from the PISA (the OECD's Programme for International Student Assessment) to construct an index

of school discipline that is far more objective (less culturally biased and more comparable across countries) than students' reports on classroom atmosphere. This measure of objective discipline proves to be a vital element in education production, enabling to estimate the previously theorized cross-relations between discipline, class size and teaching quality.

School discipline is shown to be strongly correlated with student achievement, as measured by the PISA math score. Beyond that, school discipline is also correlated with class size (within country), indicating that high discipline enables larger classes. Large classes (up to a certain size) are also correlated with higher average scores, indicating that large, normative classes get allocated better teaching resources, as hypothesized. On a systemic level, high discipline can enables efficient learning in large classes, enabling schools to be more selective and attract a higher caliber of teachers.

Using individual student discipline and school discipline (average and standard deviation), while controlling for parents education (at the student and school level) and holding class size constant, further enables this paper to estimate teaching quality as a latent variable. This methodology allows for the breakdown of student achievement - at the student, school, region and country levels - to its basic components.

In conclusion, this paper sheds light on how discipline, class size, and teacher quality interact to produce student achievement. It suggests that policies which improve classroom discipline could yield substantial dividends by enabling optimal class sizes and a stronger teaching force. Ultimately, understanding these interactions is essential for efficient educational resource allocation. While high school discipline is by no means the only determinant of success, this study highlights it as an integral part of the production function that merits explicit attention in both research and policy discussions.

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 regarding this topic, and in particular regarding the interaction between discipline, class size and teacher quality. Future research could

further our understanding of these links by exploiting exogenous school regulations affecting discipline, class-size and/or teacher allocation. Furthermore, it is regrettable that PISA does not test teachers alongside their students, as such data could prove to be pertinent.

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.

Asadullah, M. Niaz, Alain Trannoy, Sandy Tubeuf, and Gaston Yalonetzky, "Measuring educational inequality of opportunity: pupils effort matters," *World Development*, 2021, *138.*

_ **, Liyanage Devangi H. Perera, and Saizi Xiao**, "Vietnams extraordinary performance in the PISA assessment: A cultural explanation of an education paradox," *Journal of Policy Modeling*, 2020, *42* (5), 913–932.

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).

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, and Steven G. Rivkin**, "Why Public Schools Lose Teachers," *Journal of Human Resources*, 2004, *39* (2), 326–354.

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

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.

Perera, Liyanage Devangi H. and M. Niaz Asadullah, "Mind the gap: What explains Malaysias underperformance in Pisa?," *International Journal of Educational Development*, 2019, *65* (C), 254–263.

Zamarro, Gema, Collin Hitt, and Ildefonso Mendez, "When Students Dont Care: Reexamining International Differences in Achievement and Student Effort," *Journal of Human Capital*, 2019, *13* (4), 519–552.

A Additional Data

Table 6: Median Class Size, by Country and Year

Country	2012	2015	2018	2022	Total
AUS	26-30	26-30	21-25	21-25	21-25
AUT	21-25	21-25	21-25	21-25	21-25
BEL	21-25	16-20	16-20	16-20	16-20
CAN	26-30	21-25			26-30
CHE	16-20	16-20	16-20	16-20	16-20
CHL	36-40	31-35	36-40	36-40	36-40
CZE	21-25	21-25	21-25	21-25	21-25
DEU	26-30	26-30	26-30	21-25	26-30
DNK	21-25	21-25	21-25	21-25	21-25
ESP	26-30	26-30	26-30	21-25	26-30
EST	26-30	21-25	21-25	21-25	21-25
FIN	16-20	16-20	16-20	16-20	16-20
FRA	31-35	31-35	31-35	31-35	31-35
GBR	26-30	21-25	26-30	26-30	26-30
GRC	21-25	21-25	21-25	21-25	21-25
HKG	31-35	31-35	26-30	26-30	31-35
HUN	31-35	26-30	26-30	26-30	26-30
IRL	26-30	26-30	21-25	21-25	26-30
ISL	21-25	21-25	16-20	16-20	16-20
ISR	31-35	31-35	31-35	31-35	31-35
ITA	21-25	21-25	21-25	21-25	21-25
JPN	36-40	36-40	36-40	36-40	36-40
KOR	36-40	31-35	26-30	21-25	26-30
MEX	41-45	41-45	46-50	41-45	41-45
NLD	26-30	21-25	26-30	26-30	26-30
NOR	26-30	21-25	21-25		21-25
NZL	21-25	26-30	26-30	26-30	26-30
POL	21-25	21-25	21-25	26-30	21-25
PRT	21-25	26-30	26-30	21-25	26-30
SGP	36-40	36-40	36-40	31-35	36-40
SVK	21-25	21-25	21-25	21-25	21-25
SVN	26-30	21-25	26-30	26-30	26-30
SWE	21-25	21-25		21-25	21-25
TAP	36-40	36-40	36-40	31-35	36-40
TUR	51+	51+	51+	31-35	51+
USA	26-30	21-25	26-30	26-30	26-30
Total	26-30	21-25	26-30	21-25	26-30

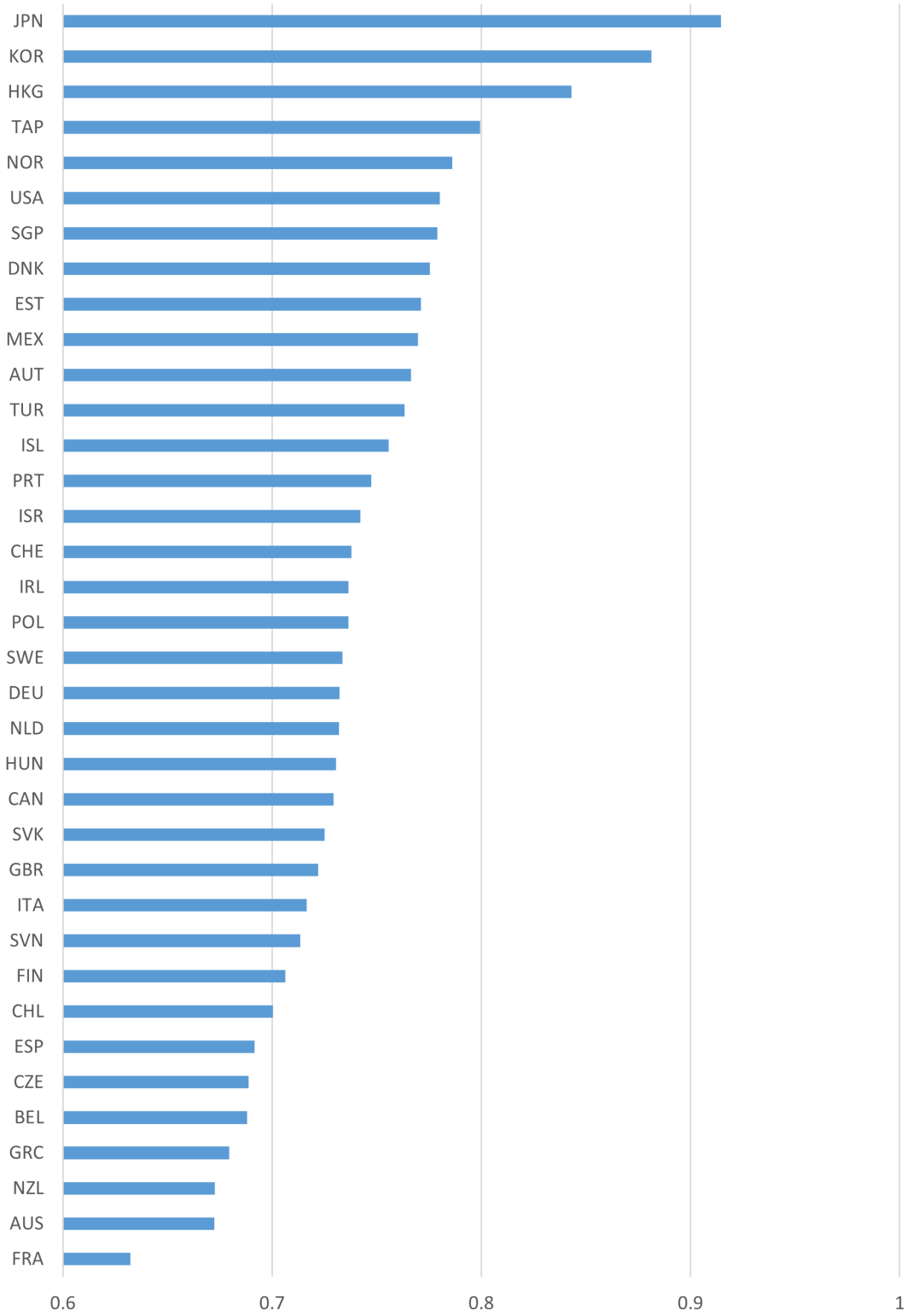


Figure 10: Subjective discipline index (2012, 2015, 2018 and 2022 round averages)

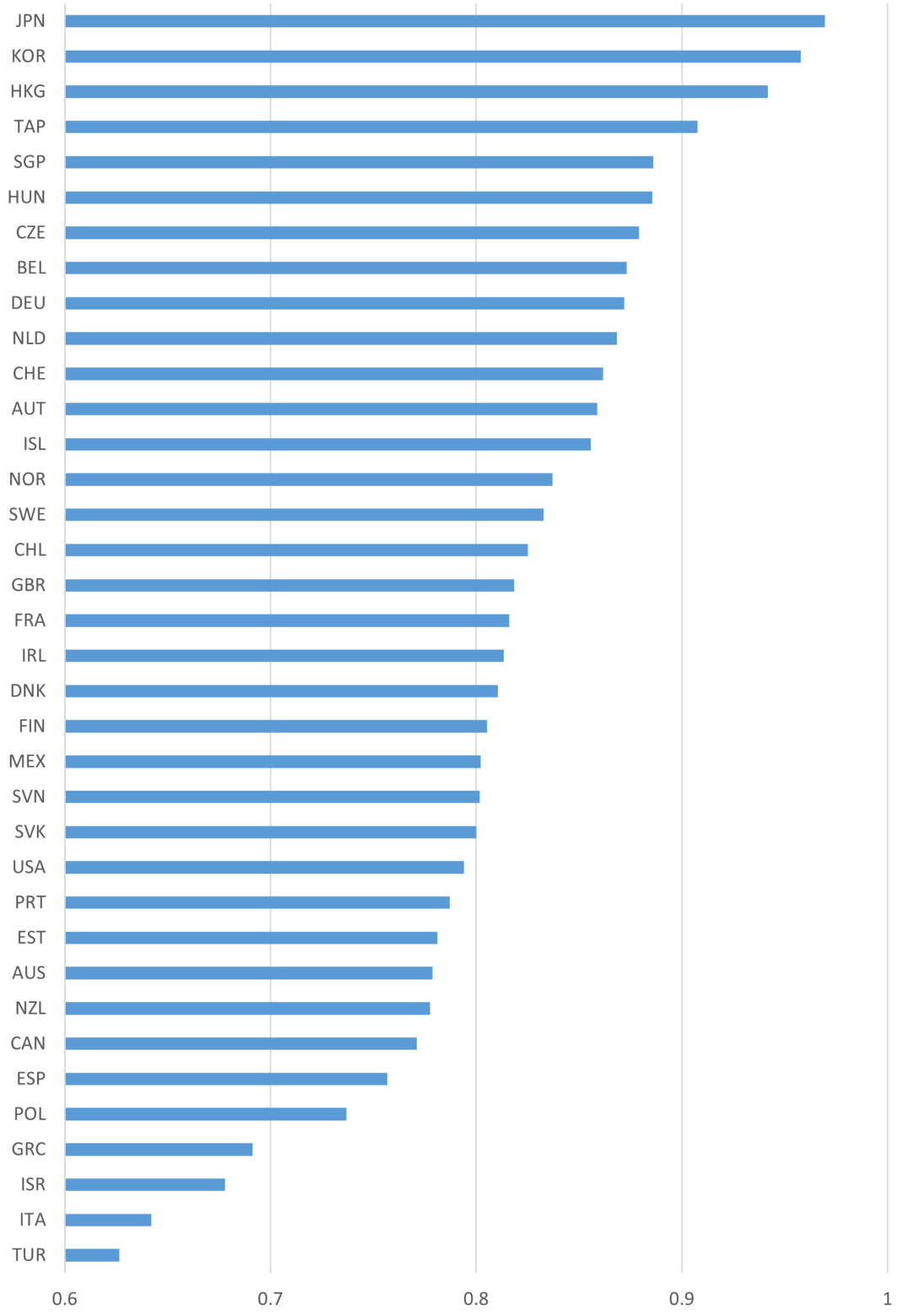


Figure 11: Objective discipline index (2012, 2015, 2018 and 2022 round averages)