I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17145

Effects of Teaching Practices on Life Satisfaction and Test Scores: Evidence from the Program for International Student Assessment (Pisa)

Kelsey J. O'Connor
Stefano Bartolini

JULY 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17145

Effects of Teaching Practices on Life Satisfaction and Test Scores: Evidence from the Program for International Student Assessment (Pisa)

Kelsey J. O'Connor
STATEC Research, GLO, IZA and University of Johannesburg

Stefano Bartolini
University of Siena

JULY 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Effects of Teaching Practices on Life Satisfaction and Test Scores: Evidence from the Program for International Student Assessment (Pisa)*

Schools are ripe for policy intervention. We demonstrate that a greater prevalence of group discussion used in schools positively affects students' life satisfaction and noncognitive skills but has no impact on test scores, based on a sample from the 2015 PISA which includes more than 35 thousand students from approximately 1500 schools in 14 countries. We perform regressions of student life satisfaction on school-level group discussion and lecturing, including a battery of controls and random intercepts by school. For robustness we use instrumental variables and methods to account for school-selection. The impact of group discussion is meaningful – a one-standard-deviation increase leads to an increase in life satisfaction that is about one-half of the negative-association with grade repetition. In contrast, lecturing does not have any effects. We are the first to show group discussion improves student life satisfaction and noncognitive skills, and thereby likely positively affects later-life outcomes.

JEL Classification: I21, I31, J24

Keywords: subjective well-being, teaching practices, noncognitive skills, test scores, participatory teaching, horizontal teaching

Corresponding author:
Kelsey J. O'Connor
School of Economics
University of Johannesburg
14 Rue Erasme
L-1468 Luxembourg
Luxembourg

E-mail: Kelsey.OConnor@statec.etat.lu

* The authors would like to thank Francesco Sarracino and Petra Thiemann for comments on previous drafts, and also seminar participants for helpful comments. O'Connor gratefully acknowledge the financial support of STATEC. Views and opinions expressed in this article are those of the authors and do not reflect those of STATEC. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare they have no competing interests.

1. Introduction

Schools are ripe for intervention. They have infrastructure in place and there is a long history of research to inform implementation (Heckman and Kautz 2013). One promising avenue for intervention is on teaching practices. Teachers are among the most important inputs to education (Hanushek 1986, 2009; Chetty et al. 2014a, b) and teaching practices can easily be adopted. The purpose of this paper is to assess the roles of lecturing and group discussion in determining student outcomes, in particular life satisfaction, test scores, and noncognitive skills using the 2015 Program for Individual Student Assessment (PISA).

Lecturing and group work are mainstays of, respectively, traditional and modern teaching methods (the latter also called horizontal, participatory, student-centered, or progressive). Traditional teaching was repeatedly challenged throughout the twentieth century. Several reform movements promoted more student-centered teaching practices, wherein emphasis is placed on small group work and discussion among students (Dewey 1944; Montessori 2004). International organizations, such as the (OECD 2017a) and (UNICEF 2009), and many national educational authorities have advocated for increased use of such practices (NCTM 1989; 1991; Capps, Crawford, and Constas 2012; National Research Council (U.S.) 1996). Despite these efforts however, traditional teaching practices still predominate in many countries (Brulé and Veenhoven 2014).

Some quantitative studies have found a positive and significant effect of traditional teaching methods on test scores, while the impact of modern methods is weak or non-existent (Cordero et al.; Schwerdt and Wuppermann 2011; Bietenbeck 2014). Other studies found mixed results, depending on students' gender and socio-economic background (Lavy 2016). However, test scores account for only a modest portion of the variability in later-life success. Adolescent test scores predict approximately 15% of the variance in later-life earnings (Heckman and Kautz 2012).

An increasing body of empirical studies indicates that noncognitive skills rival IQ in predicting educational and labor market outcomes, health, and criminality (Roberts et al. 2007; Borghans et al. 2008; Lleras 2008; Almlund et al. 2011; Heckman and Kautz 2012, 2013). Noncognitive skills are often defined and measured in terms of personality traits, social skills, and work habits (effort, discipline, and determination) and they are not necessarily picked up by test scores (Jackson 2018).

The limited quantitative evidence suggests that modern practices have greater noncognitive benefits than traditional methods. For instance, modern teaching practices positively relate to

teachers' ability to affect math scores and noncognitive skills, while traditional teaching hinders noncognitive skills (Flèche 2017). Group discussion also fosters students' social capital (Algan et al. 2013), which is related to the social skills dimension of noncognitive skills. In another study, Bietenbeck (2014) finds modern teaching better promotes reasoning skills than traditional methods. Reasoning skills are similar to noncognitive skills in that they are both not measured well by standardized tests (Bietenbeck 2014).

What sets our paper apart is the focus on life satisfaction. Life satisfaction is an important outcome on its own (Kahneman et al. 2004; Graham et al. 2018; The Global Happiness Council 2018; Helliwell 2019; Frijters et al. 2020; O'Connor 2023) and a predictor of numerous positive outcomes, including for instance: social capital (Guven 2011), wages (Graham et al. 2004; De Neve and Oswald 2012; Oswald et al. 2015), employment (Krause 2013; O'Connor 2020a), consumption and savings (Guven 2012), and health (Tay et al. 2015). See (De Neve et al. 2013) or (Piekałkiewicz 2017) for a summary of articles.[1]

We hypothesize that modern teaching practices are positively related to life satisfaction. We formulated this hypothesis based first on the above-evidence that modern practices relate to noncognitive skills and social capital, which are in turn closely related to life satisfaction (Specht et al. 2012; Soto 2015; Helliwell and Aknin 2018). Second and more fundamentally, we base our hypothesis on the nature of modern teaching practices. In group discussion, students' attentions shift horizontally, from the teacher towards each other. Yet, unlike settings outside the classroom, group discussion is supervised. Students thus have a more psychologically safe[2] environment to socialize and communicate. Shy and diverse students also contribute, facilitating otherwise difficult inclusion. They have the opportunity to express themselves to their peers and the teacher, which, with constructive feedback, may lead to greater senses of autonomy, competence, and relatedness – the three basic psychological needs according to Self-Determination Theory (Ryan et al. 1996). Students also develop an implicit understanding of reciprocity amongst each other. Exposure-to-diversity improves the relations between groups, potentially forging what is referred to as bridging social capital (Granovetter 1973). Moreover, they develop skills in an environment

[1] Although these articles focus on the influence of life satisfaction in adults, the best predictor of adult life satisfaction is their emotional health in childhood (Clark et al. 2019, p. 223).

[2] For a discussion of psychological safety in team performance see (Edmondson 1999).

where they are expected to perform, e.g., leadership and agreeableness. Third, students may simply enjoy themselves more and experience more positive affect in group discussion.

Ultimately, the relationship between modern teaching practices and life satisfaction is an open question. We found one related study (Brulé and Veenhoven 2014), which has two main findings. First, the *general populations* in countries with more modern teaching are more satisfied with their lives. Second, *students* living in countries with more modern teaching are no more or less satisfied with their lives. While informative, it is difficult to generalize from this study. The authors offer the explanation that modern teaching practices likely increase psychological freedom or autonomy in a country, but there is little difference between students' perceptions of freedom. Alternatively, teaching practices are associated with other positive country characteristics, while the connection between teaching practices at the country level is not strong enough to explain differences in student satisfaction. In the present study, the connection is more immediate; we connect life satisfaction of students to the teaching practices within a school.

We contribute by rigorously assessing the relations between the teaching practices, lecturing and group discussion, and students' life satisfaction, test scores, and noncognitive skills using a large-scale cross-country dataset on 15-year-olds around the world. We regress individual student outcomes on school-level teaching practices, a battery of controls, and school-level random effects. Unfortunately, the PISA data only allow us to match students and teachers to their schools – we cannot directly match students and teachers. However, exploiting school-level variation means we only need to address school-level endogeneity. In addition to random school effects, we conduct robustness tests using instrumental variables and a two-stage model for school selection.

We are the first to show that the prevalence of group discussion used in schools is positively related to students' life satisfaction at the school level. The benefits do not apply to girls, however, unless they attend all-girl schools. The positive average life-satisfaction relation comes at no cost in terms of tests' scores – they are not impacted by different levels of teacher explains or group discussion, which is important as the relations could have been positive or negative (Goldhaber and Brewer 1997; Schwerdt and Wuppermann 2011; Bietenbeck 2014). What is more, the apparent benefits of group discussion extend to noncognitive skills.

The paper is organized as follows. Section II discusses the data and methods, Section III, the results concerning the effects of lecturing and group discussion on life satisfaction, including robustness and heterogeneity. In Section IV, we first discuss the importance of noncognitive skills

for later life outcomes and then discuss the relations between teaching practices and test scores and noncognitive skills. Section V concludes.

2. Data and Methods

2.1. Program for International Student Assessment

We estimate the relations between teaching practices and life satisfaction, test scores, and noncognitive skills using data from 2015 round of PISA.[3] The primary purpose of PISA is to measures the science, reading, and mathematics knowledge and skills of 15-year-olds in OECD and partner countries. However, it also conducts questionnaires for students, parents, teachers, and school administrators. In 2015, a teacher questionnaire was administered that provides the necessary information on teaching practices for our analysis. However, this questionnaire has two important limitations. First, only the science teachers were asked, as the 2015 round was focused on science. Second, the teacher questionnaire is only matched to schools and not to students. Together these limitations mean the analysis can only exploit individual-student outcomes as a function of school-level variation in science teaching practices.

We operationalize teaching practices with an emphasis on the use of group discussion and lecturing.[4] The prevalence of group discussion and lecturing within a school are calculated as the average of teachers' responses within a school to the two questions: "How often do these things happen in your <school science> lessons?" (1) "I explain scientific ideas." (2) "A small group discussion between students takes place." Possible answers include: (1) Never or almost never (2) Some lessons (3) Many lessons (4) Every lesson or almost every lesson. Using this approach we implicitly assume that teachers share a common "teaching culture" (as discussed in (Echazarra et al. 2016)). To improve reliability, schools were dropped with fewer than four responding teachers or with a 20 percent or smaller response rate.[5] Schools with a low level of teacher cooperation were also dropped, based on the responses to a set of questions relating to cooperation.

[3] Unfortunately, not all of the relevant variables are available in earlier or subsequent waves.

[4] Although there are various alternative approaches, including creating indices for modern and traditional teaching practices (Cordero et al.; Bietenbeck 2014). We feel our approach is intuitive and limits subjectivity in forming the index. (Algan et al. 2013; Brulé and Veenhoven 2014) likewise use group discussion and lecturing, although the latter subtracts lecturing from group discussion to create one index.

[5] 20 percent was likewise used in (Cordero et al.).

Life satisfaction is measured using the response to the question "The following question asks how satisfied you feel about your life, on a scale from "0" to "10". Zero means you feel 'not at all satisfied' and "10" means 'completely satisfied'. Overall, how satisfied are you with your life as a whole these days?" (Please move the slider to the appropriate number.)" The survey is conducted on a computer.

Table 1 presents details of our final sample, which includes more than 35 thousand students, across nearly 1500 schools, in 14 different countries or regions. PISA 2015 includes a larger number of countries, but we are constrained to those that voluntarily administered the teacher questionnaire. We also exclude privately owned or operated schools.

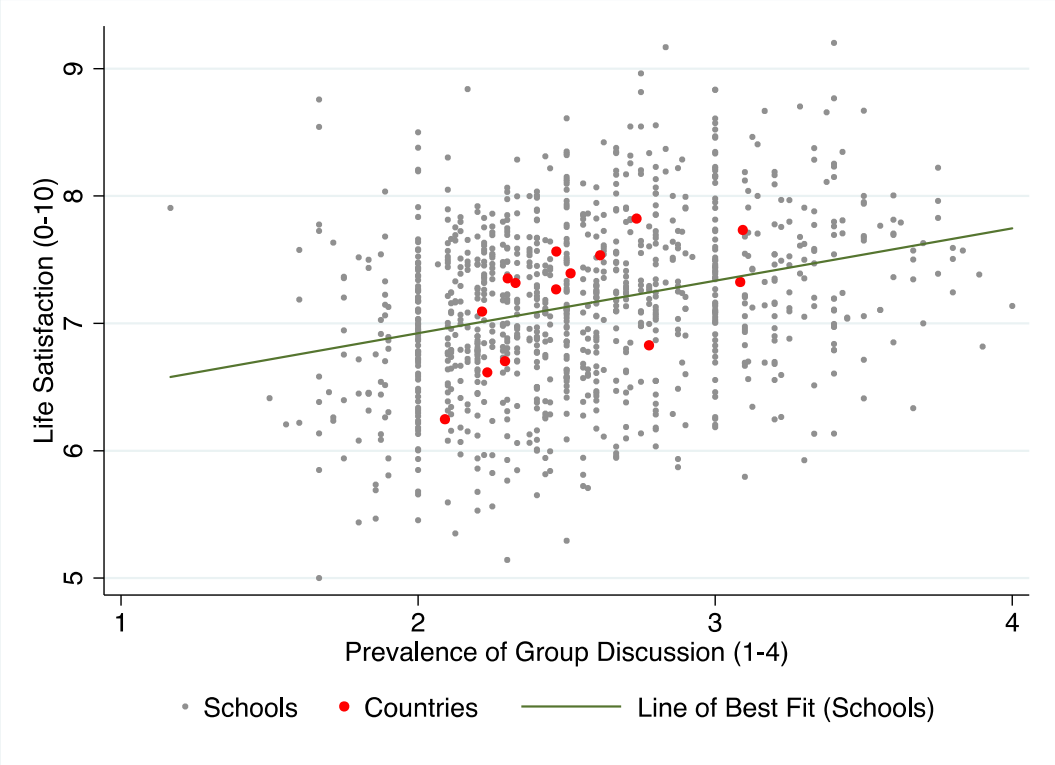
Table 1: Sample observations, schools and students within each country or region

Country/Region	# of schools	# of students
Brazil	156	1,786
Chile	21	471
Chinese Taipei	117	3,871
Colombia	88	2,366
Czech Republic	222	4,032
Germany	106	1,142
Hong Kong	9	295
Korea	66	2,048
Peru	73	1,540
Portugal	127	3,447
Spain	97	2,740
United Arab Emirates	97	2,233
United States	88	2,136
B-S-J-G (China)	222	7,019
Total	1,489	35,126

Note: B-S-J-G (China) refers to the four China provinces: Beijing, Shanghai, Jiangsu and Guangdong.

Figure 1 illustrates the data, in particular average life satisfaction and group discussion separately by school and country. The scatter plot suggests that students in both schools and countries with a greater prevalence of group discussion are more satisfied with their life.

Figure 1 Life satisfaction and prevalence of group discussion within schools and countries



Source: Author calculations using PISA 2015

2.2. Methods

Students' life satisfaction is assumed to depend on teaching practices, individual characteristics, and the characteristics of their teachers, school, and country. Equation 1 specifies this data generating process:

(1) $Y_{isc} = \alpha + Discussion_{sc}\beta_1 + Lecture_{sc}\beta_2 + \boldsymbol{Teach'_{sc}\gamma_1} + \boldsymbol{School'_{sc}\gamma_2} + \boldsymbol{X'_{isc}\gamma_3} + \boldsymbol{d'_c\gamma_3} + \mu_{sc} + \boldsymbol{\epsilon_{isc}}$

Y_{isc} is the outcome for student i in school s and country c. Initially life satisfaction is used as the outcome, but more are assessed later. $Discussion_{sc}$ is the prevalence of group discussions in a school, while $lecture_{sc}$, is the prevalence of teacher explains. $\boldsymbol{Teach_{sc}}$, $\boldsymbol{School_{sc}}$, and $\boldsymbol{X_{isc}}$ are vectors of teacher, school, and individual characteristics. $\mathbf{d_c}$ is a vector of country dummies and μ_{sc}

represents random intercepts by school in a hierarchical linear model, estimated using maximum likelihood with standard errors clustered at the school level.

Equation 1 is analogous to a reduced form regression in an instrumental variable approach where school-level teaching practices is used as the excluded instrument for classroom teaching practices. The analogy is helpful to understand potential threats to identification. Any idiosyncratic student characteristics, observed or unobserved, are not likely to cause any endogeneity in teacher-practices at the school-level. On the other hand, factors common to students at the school-level could be a problem. Teaching practices may be determined by the administration, or correlate with alternative characteristics of the teachers or school, and students may have selected into schools based on the schools' characteristics. To address this concern, omitted variables at the teacher/school level are first accounted for using a battery of pertinent controls, discussed below. Secondly, the random intercept, μ_{sc}, accounts for residual variation between schools. This approach is fairly stringent, as random intercepts vary directly with our variable of interest (which is why fixed effects cannot be used); however, it also requires the assumption that the random intercepts are orthogonal to teaching practices. This assumption is not testable, but we relax it using alternative methods described below.

To interpret our main results casually, we must ultimately rely on conditional independence. This assumption is strengthened by the fact that we use two variables for teaching practices that are defined in the same way. Potential threats to identification would need to affect one variable and not the other, which we can safely rule out at least in the case of measurement error, for instance. We also run robustness tests in Section III.B, using models that allow for heterogeneous effects of group discussion by school (random slopes), instrumental variables, and selection into the schools. The latter two approaches also do not make the orthogonality assumption behind random intercepts.

We include a battery of controls in different specifications – student characteristics: gender; relative age (months); grade; whether ever repeated a grade; nativity; parents' education and nativity, socio-economic and cultural status; whether there is a computer and number of books at home. Socio-economic and cultural status is an index prepared by PISA based on parents' occupation, education, and home possessions as a proxy for income. Teacher characteristics include: the ratio of female to male teachers; average teacher age; average teacher experience; education of teachers; share of teachers that are fully certified to teach; and teacher job satisfaction.

Teacher characteristics are based on responses to the teacher survey and averaged over teachers within a school. School characteristics include: the student-teacher ratio; average socio-economic and cultural status of students; whether the students are all male or female; rural or urban location; and additional characteristics based on the responses to a survey of school principals. These include whether: students are grouped by ability, tests are used to adopt teaching practices, and the principal thought education was hindered separately by: (1) lack of resources (2) lack of staff or (3) teacher behavior.

PISA samples students in two stages: schools are first sampled and then students are sampled in the participating schools. Student responses are weighted using the final student weights provided by PISA (w_fstuwt). Additional school weights were used in the random intercept, hierarchical, models. They are calculated as the sum of final student weights within a school (OECD 2016, p. 298).

3. Results

3.1. Life Satisfaction Main Results

Table 2 presents the main results. A greater prevalence of group discussion within schools is associated with higher life satisfaction. This result holds when including characteristics of the students, teachers, and schools. In column 1, a school with one standard deviation more group discussion (0.48) is associated with 0.12 greater life satisfaction points (0.48 * 0.25 = 0.12), which is small relative to the standard deviation of life satisfaction (2.26), but not small compared to the relations for other variables – more than one third of that for girls, who are approximately 0.33 points less satisfied with their life and one half of those who repeat a grade; they are approximately 0.23 points less satisfied. The coefficients on individual characteristics are omitted from Table 1 for brevity but included in Appendix Table 7.

Table 2. Regressions of Life Satisfaction on Teaching practices

	(1) Life Sat.	(2) Life Sat.	(3) Life Sat.	(4) Life Sat.
Group discussion	0.248***	0.257***	0.219***	0.186***
	(0.062)	(0.064)	(0.064)	(0.066)
Teacher explains	-0.032	-0.028	0.000	-0.002
	(0.078)	(0.083)	(0.082)	(0.082)
Female Teacher Ratio		0.174	0.256	0.236
		(0.169)	(0.175)	(0.173)
Average Teacher Age		-0.013	-0.015	-0.015
		(0.017)	(0.017)	(0.016)
Avg. Teacher Exp. (Yrs.)		0.020	0.023	0.020
		(0.017)	(0.017)	(0.016)
Teach. Share < Bach.		0.489*	0.170	0.151
		(0.291)	(0.284)	(0.283)
Teach. Share > Bach.		-0.135	-0.007	0.022
		(0.176)	(0.161)	(0.163)
Teach. Share Fully Cert.		-0.039	-0.040	-0.029
		(0.130)	(0.127)	(0.127)
Sch: Socioecon. & Cult.			-0.197***	-0.204***
			(0.068)	(0.067)
Student-Teacher Ratio			-0.004	-0.002
			(0.004)	(0.004)
Hindered: Lack of Res.			-0.023	-0.011
			(0.030)	(0.029)
Hindered: Lack of Staff			-0.048*	-0.028
			(0.029)	(0.033)
Rural			0.018	0.004
			(0.065)	(0.063)
All Male or Female Studs.			-0.013	0.009
			(0.091)	(0.089)
Tests Used to Group				0.037
				(0.054)
Tests used to Adapt Teaching				0.202***
				(0.060)
Hindered: Teach. Behavior				-0.022
				(0.027)
Satisfaction of Teachers				0.356**
				(0.147)
Attend Phys. Ed. (Days)				0.047**
				(0.021)
Constant	8.001***	7.948***	7.757***	6.241***
	(0.543)	(0.705)	(0.718)	(0.863)
Individual Controls	yes	yes	yes	yes
Random School Intercept	yes	yes	yes	yes
Country Fixed Effects	yes	yes	yes	yes
Total Obs.	35126	35126	35126	35126
Number of Schools	1489	1489	1489	1489

Notes: Life satisfaction has a mean of 7.13 and standard deviation of 2.26. Group discussion has a mean of 2.75 and standard deviation of 0.48.
Clustered standard errors in parentheses (by school); * p<0.10 ** p<0.05 *** p<0.01

Each column successively adds controls. Few are consistently related to student life satisfaction. At the school level, perhaps unsurprisingly teachers that are more satisfied with their job lead to positive outcomes (consistent with (Opdenakker and Van Damme 2006)), in this case more satisfied students. Students' satisfaction also benefits from physical education and adaptive teaching based on tests. The average socio-economic and cultural status across the school is negatively related to life satisfaction, confirming similar findings for adults on the negative effects of relative income (Luttmer 2005). Among the individual characteristics presented in the extended Table 6, girls and those who repeated a grade are less satisfied, while those with computers, fathers with moderate education (ISCED 1), and higher socio-economic and cultural status parents are more satisfied. Interestingly, the individual and school-average coefficients on cultural status are nearly the same magnitude and in opposite directions – indicating that equally increasing every students' socio-economic status, e.g., through economic growth, would not have a direct effect on students' life satisfaction, but could if some students gained more than others (consistent with explanations of the Easterlin Paradox e.g., (Easterlin 2001; Clark et al. 2008; Easterlin and O'Connor 2021)).

3.2. Robustness of teaching practice effects on life satisfaction

The previous analysis controls for numerous sources of potential bias; to address further concerns we utilize three additional approaches. First, there may be significant heterogeneity in the impacts of group discussion on life satisfaction across schools and the estimated common-relation could be driven by a small number of outliers. To address this concern, we utilize a strength of hierarchical linear models, random slopes. Thus, we allow the relation for group discussion to vary across schools. Although additional variables could be allowed to take random slopes, the maximum likelihood function becomes more complicated and poses computational difficulties. Indeed this was a problem in our case. As a consequence, we endeavored to simplify the model as much as possible without sacrificing important controls. We discuss this further with the results below.

The second approach is two-stage least squares using excluded instruments that are generated using the (Lewbel 2012) approach. Using this approach, both teacher explains and group discussion are allowed to be endogenous and predicted in a first stage, allowing for second stage estimates that are considered to be unbiased and causally interpretable. The primary limitation of

the Lewbel approach is that it is not very intuitive and that the instruments do not have an explicit economic meaning. Stated briefly, to generate instruments, two steps are taken: (1) regress the endogenous variable on the exogenous ones (X) and save the residuals (ν), (2) then generate the instruments as $(X_j - E(X_j))*\nu = Z_j$. There must be heteroskedasticity in ν for the instruments Z_j to take meaningful values, which is testable using the standard Breusch-Pagan test. The key assumption (untestable) is that X is unrelated to the *product* of the first stage errors with the second stage errors. However, the standard instrumental variable (IV) diagnostics can be used (i.e., first stage F stat and an overidentification test) to assess the relevance and validity of the generated instruments. For a detailed description of the Lewbel method, see (Lewbel 2012); for implementation in STATA, see (Baum et al. 2013). The approach has been used in a number of publications now (as documented in (Lewbel 2012) and more recently by (Le Moglie et al. 2015; Arampatzi et al. 2018; Sarracino and Fumarco 2018; O'Connor and Graham 2019; O'Connor 2020b). It is particularly relevant in research on subjective well-being, due to the difficulty of identifying plausible instruments in such settings, as stated in (O'Connor 2020a), based on (Di Tella et al. 2003; Graham et al. 2004; Krause 2013).

The third approach addresses selection into schools using a two-step procedure that adds selection controls similar to Inverse Mills Ratios to equation 1. The selection controls account for unobserved heterogeneity that would drive students to choose one type of school over another, thus strengthening the conditional independence assumption. To facilitate this approach, we characterized schools as belonging to one of four types: (1) below medians of both group discussion and teacher explains, (2) below median group discussion and above median teacher explains, (3) above median group discussion and below median teacher explains, and (4) above medians of both group discussion and teacher explains. The first step of the procedure is to estimate the probability of being in a particular school type, using a multinomial logit model. The probabilities are used to create the selection controls, and in the second step, the modified version of equation 1, including the selection terms, is estimated. The results yield unbiased estimates of the relationship between school practices and student life satisfaction. For identification, the two-step procedure relies on the functional form of the selection terms and any exclusion restrictions. In order to have exclusion restrictions, we included controls for the prevalence of certain factors used in determining student admission solely in the first step (e.g., academic performance, recommendations of feeder schools, religion, special programme, family, geographic area, other).

For further details of the method see (Origo and Pagani 2009, p. 553), which is in turn is based on (Dubin and McFadden 1984).

Results of the robustness tests are presented in Table 3. Regardless of approach, more group discussion is associated with greater life satisfaction. Allowing for heterogeneous effects of group discussion across schools had no impact. Column 1 replicates the main analysis with random intercepts but with fewer controls and excluding teacher explains. This limited model compares with column 2, which uses random slopes for teacher explains. It was limited because the more complex models did not converge. The Lewbel method yields similar results, with a larger magnitude than the baseline results with the same controls except the random intercept (compare with 0.186 in Table 2 column 4). A larger magnitude makes sense when there is classical measurement error. The IV diagnostics indicate the Lewbel instruments are relevant and valid (high F and Hansen J p-value).

Results from the selection model (columns 5-7) also suggest group discussion led to higher life satisfaction. Students in the fourth group, comprised of above median prevalence of group discussion and teacher explains, are more satisfied than students in the omitted-second group, comprised of below median group discussion and above median teacher explains. The only difference between these two groups is whether the school is above or below the median in group discussion and students in the group above are more satisfied. The relation is statistically significant in each specification, while the selection terms are not statistically significant (individually or jointly as indicated by the F or Chi Squared stat at the bottom of the table). Together these results suggest teaching practices, not selection into one teaching-practice type school or another, affects life satisfaction. Admission criteria is important in determining which type of school students select into, but not life satisfaction, suggesting it indeed works well as an excluded variable for the selection model. Column 4, shows the probabilities associated with going to a school in the fourth group, compared with the second group. Students in schools that admit based on past performance are 64 percent less likely to go to the fourth.

Table 3 Robustness regressions using different models

	(1) Rnd. Intcp. Life Sat.	(2) Rnd. Slope Life Sat.	(3) Lewbel Life Sat.	(4) M-Logit Tch Style 4	(5) OLS Life Sat.	(6) Rnd. Intcp. Life Sat.	(7) Rnd. Intcp. Life Sat.
Group discussion	0.249***	0.248***	0.217***				
	(0.062)	(0.061)	(0.072)				
Low G & T					-0.073	-0.083	-0.065
					(0.070)	(0.074)	(0.072)
High G & Low T					0.080	0.074	0.067
					(0.092)	(0.094)	(0.091)
High G & T					0.187*	0.178*	0.188**
					(0.096)	(0.102)	(0.096)
Teacher explains			-0.014				
			(0.090)				
Admis-Performance				-0.636**			0.400
				(0.303)			(0.487)
Admis-Feeder				0.071			0.121
				(0.414)			(0.259)
Admis-Ideology				-0.213			-0.113
				(0.352)			(0.225)
Admis-Program				-0.200			-0.086
				(0.296)			(0.327)
Admis-Legacy				0.665**			-0.064
				(0.297)			(0.419)
Admis-Residence				0.151			0.027
				(0.300)			(0.292)
select1					0.021	0.022	-0.152
					(0.022)	(0.021)	(0.250)
select3					-0.001	0.001	0.179
					(0.016)	(0.017)	(0.204)
select4					0.007	-0.000	-0.012
					(0.015)	(0.016)	(0.296)
Constant	7.903***	7.906***	5.940***	8.249	5.973***	6.050***	12.048
	(0.483)	(0.482)	(0.837)	(5.029)	(1.038)	(1.068)	(14.840)
Individual Controls	yes	yes	yes	yes	yes	yes	yes
Teacher Controls	-	-	yes	yes	yes	yes	yes
School Controls	-	-	yes	yes	yes	yes	yes
Rnd. School Intercept	yes	yes	-	-	-	yes	yes
Country Fixed Effects	yes	yes	yes	yes	yes	yes	yes
Total Obs.	35126	35126	35126	35126	35126	35126	35126
Number of Schools	1489	1489	1489	1489	1489	1489	1489
Hansen J p-value			0.551				
Kleibergen-Paap F			210.243				
Selection Joint F or Chi Stat.					0.621	1.407	1.351

Notes: Clustered standard errors in parentheses (by school); * p<0.10 ** p<0.05 *** p<0.01

3.3. Heterogeneity of teaching practice effects on life satisfaction

Teaching practices may have different impacts on different students, in particular: boys and girls, wealthy/non-wealthy students, natives/immigrants, high/low test-scoring students, or students in different environments, that is, wealthy versus poor schools, coeducation compared to single gender schools, and rich versus lower-income countries.[6] We tested contrasting groups sequentially using interactions with the group of interest. Indeed, important differences are observed for group discussion across different population groups. Teacher explains generally maintains a statistically insignificant relation. Table 4 presents the results. Column 1 repeats the baseline results from Table 2 column 4.

Gender plays a substantive role. Group discussion is positively related to life satisfaction for boys but not girls when they are educated together; however, when girls are in all-girl schools, group discussion is again positively associated with life satisfaction. The first relation is observed in column 2: the group discussion – female interaction term is statistically significant, negative, and large enough to offset the positive main effect. The correlation of group discussion for girls when in coeducation schools is 0.280 - 0.173 = 0.107, which is statistically insignificant (unreported). In column 8, the female interaction is used along with a dummy and interaction for all-girl schools. In all-girl schools, the correlation of group discussion on girls is positive (0.287 - 0.198 + 0.552 = 0.641), which is statistically significant. Teacher explains also has a more negative correlation for girls; indeed, statistically negative in column 2 (0.171 - 0.336 = -0.165), but the relation turns significantly positive again in all girls schools (0.171 - 0.351 + 0.369 = 0.189). Being in an all-girl school is statistically negatively related generally (-2.657, column 9), however, only all-girl schools – single gender schools are not statistically more or less satisfied across columns[7]. The coefficient on gender also changes substantively when including interaction terms by gender; indeed girls report higher life satisfaction than boys (cols. 2 and 9), but only when the interaction terms play no role, that is, when the prevalence of both group discussion and teacher explains are zero. Accounting for the interactions with teaching practices at their mean values, girls report

[6] Rich countries include: Chinese Taipei, Germany, Hong Kong, Korea, Spain, United Arab Emirates, and the United States. Lower income countries include: Brazil, Chile, Colombia, Czech Republic, Peru, Portugal, and B-S-J-G (China).

[7] The control for single-gender is dropped in column 8 to due to collinearity with all-female schools.

approximately 0.323 life satisfaction points less than boys, which is quite similar to the estimates in other columns.

Test performance, nativity, and school socio-economic status also moderate the relations. Higher scoring[8] students benefit less from group discussion. In column 5 the interaction term is statistically significant and negative. However, the total group-discussion – life satisfaction relationship does not turn negative. The high score in the sample is 854.03, meaning the relationship is still positive (0.862-0.001*854.03). At the mean test score of 502.37, group discussion is related to 0.360 (0.862-0.001*502.37) more life satisfaction points per unit increase in group discussion. In column 4, natives seem to benefit less from group discussion, but not significantly. For the foreign born, the correlation of group discussion (the main effect as they are the omitted category) increases in magnitude but becomes statistically insignificant, presumably because there is a smaller number of immigrants, which reduces precision. School socio-economic status (col. 6) plays a small moderating role, slightly reducing the magnitude of group discussion, and increasing the correlation of teacher explains (significant at ten percent). The other interaction terms are not statistically significant and the main effects are similar. The relationship between group discussion and student life satisfaction does not significantly vary by student SES, country income, or single-gender schools, except all-girl schools as discussed above.

[8] We use the average of test scores across math, reading, and science.

Table 4 Regressions of life satisfaction on teaching practices, heterogeneous relations by group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Group discussion	0.186***	0.280***	0.178**	0.297	0.862***	0.170**	0.189***	0.287***	0.215***
	(0.066)	(0.086)	(0.070)	(0.329)	(0.286)	(0.078)	(0.068)	(0.087)	(0.078)
Teacher explains	-0.002	0.171*	0.020	-0.161	-0.174	0.080	-0.008	0.171*	-0.057
	(0.082)	(0.095)	(0.084)	(0.470)	(0.283)	(0.094)	(0.083)	(0.095)	(0.109)
Female	-0.316***	1.195***	-0.316***	-0.316***	-0.327***	-0.315***	-0.316***	1.315***	-0.316***
	(0.046)	(0.402)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.409)	(0.046)
All Male or Female Studs.	0.009	0.015	0.004	0.008	-0.005	-0.007	-0.362		0.007
	(0.089)	(0.093)	(0.089)	(0.089)	(0.089)	(0.090)	(0.814)		(0.089)
Group X Female		-0.173**						-0.198**	
		(0.087)						(0.088)	
Teach Ex. X Female		-0.336***						-0.351***	
		(0.099)						(0.100)	
Group X SES			-0.015						
			(0.050)						
Teach Ex. X SES			0.045						
			(0.048)						
Group X Native				-0.114					
				(0.316)					
Teach Ex. X Native				0.164					
				(0.461)					
Mean Score					0.002				
					(0.002)				
Group X Test Score					-0.001**				
					(0.001)				
Teach Ex. X Test Score					0.000				
					(0.001)				
Group X Sch. SES						-0.024			
						(0.077)			
Teach Ex. X Sch. SES						0.154*			
						(0.094)			
Group X Sch. 1 Gender							-0.061		
							(0.175)		
Teach Ex. X Sch. 1 Gender							0.170		
							(0.226)		
All Girl School								-2.657***	
								(0.859)	
Group X All Girl								0.552***	
								(0.199)	
Teach Ex X All Girl								0.369	
								(0.355)	
Group X Rich Cntry									-0.057
									(0.128)
Teach Ex. X Rich Cntry									0.107
									(0.161)
Constant	6.241***	5.457***	6.218***	6.458***	5.299***	6.088***	6.256***	5.456***	6.363***
	(0.863)	(0.878)	(0.874)	(1.542)	(1.237)	(0.892)	(0.864)	(0.880)	(0.938)
Individual Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Teacher Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
School Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Random School Intercept	yes	yes	yes	yes	yes	yes	yes	yes	yes
Country Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Total Obs.	35126	35126	35126	35126	35126	35126	35126	34622	35126
Number of Schools	1489	1489	1489	1489	1489	1489	1489	1466	1489

Notes: Test scores use the average across math, reading, and science.
Clustered standard errors in parentheses (by school); * p<0.10 ** p<0.05 *** p<0.01

4. Effects on Tests Scores and Noncognitive Skills

Perhaps group discussion is conducive not only to student life satisfaction, but also tests scores and noncognitive skills. Many readers are naturally interested in test scores. They are the primary output of PISA and contribute to academic placement. For this reason, we assess the impact of teaching practices on tests scores. We also assess noncognitive skills or behavior. We know test scores miss important outcomes for later life achievement (Jackson 2018), which may be better predicted using noncognitive skills.

The importance of noncognitive skills for later life outcomes can be illustrated by the Perry Preschool Study and further supported by the literature. The study assessed the impacts of an experimental intervention that targeted children with low intelligence quotients (IQs) and found the intervention did not improve IQs but the treatment group nonetheless showed greater success in later life (educational achievement and social skills). This improvement was attributed to improvements in personality and motivation (Borghans et al. 2008, p. 973). The results are not surprising in light of the noncognitive skills literature. In a formal model, (Borghans et al. 2008) represents noncognitive skills as capacities, similar to cognitive or physical ability; capacities in turn affect productivity and therefore the energy or time necessary to achieve tasks. Empirically, noncognitive skills have been shown to be important for a variety of tasks or outcomes. For support, see the special issue on noncognitive skills in the *Journal of Human Resources* (ter Weel 2008) and the following quotes: (1) "For many outcomes, the predictive power of character [noncognitive] skills rivals that of measures of cognitive ability (Heckman and Kautz 2013, p. 23)," and (2) "…the effects of psychological capital variables in predicting wages are even stronger than the effects of traditional human capital variables (Mohanty 2009, p. 357)."

Noncognitive skills have multiple definitions and operationalizations; they "are often defined and measured in terms of work habits, such as effort, discipline, and determination, or in terms of behavioral traits, such as self-confidence, sociability, and emotional stability (ter Weel 2008, p. 729)." (Heckman and Kautz 2013) refer to them as character skills instead of noncognitive *traits* to emphasize they are not strictly free from cognition and they are trainable. By either name, they are most frequently measured in economics using the Big Five Personality traits, because the Big Five are well understood and frequently measured. However, while the Big Five are relatively more widely available, few cross-country surveys include them.

Some researchers have used what might be called noncognitive *behavior*[9], which (Lleras 2008), summarizing from the literature, describes as "[1] being hard working and conscientiousness, [2] having the ability to get along well with others and work effectively in groups, [3] being polite, and following the rules (pg. 893)." The author measures the first using teachers' perceptions of students' work habits, including, for example, whether the student completes assignments or is late to class. The second, sociability and cooperativeness, was measured using the teacher perceptions of whether the student "related well with other students" and the number of extracurricular and sports activities the student participated in. Lastly, (Lleras 2008) used teachers' responses regarding whether the student was "disruptive in class". Similarly, (Jackson 2018) uses absences, course grades, grade repetition, and suspensions. The first three would generally fall into the first category described by Llears, while suspensions fall more into the third category.

The PISA surveys include behaviors that fit within the broad definitions of noncognitive skills. In particular, we use: (1) self-reports on how often they skipped class, (2) an index for perceptions of whether other students pay attention in class, (3) whether the student studies or does homework before or after school, (4) self-reports of expected level of completed education (5) self-reports of whether they make friends easily (1 strongly disagree to 4 strongly agree), (6) self-reports on sense of (6) belonging socially or (7) being bullied at school (PISA constructed), and (8) whether the principal believed student behavior hindered learning. The first three are similar to conscientiousness; the fourth is similar to a measure of optimism; five and six are about sociability; and 7 and 8 are about being polite and following the rules.

4.1. Test Scores

Test scores cover math, science, reading, and collaborative problem-solving skills[10]. To assess the impacts of teaching practices on test scores, we replicate the main analysis in Table 2, column 4, and Lewbel robustness test in Table 3, column 3, using the test scores as outcomes.

Presented in Table 5, the prevalence of group discussion and teacher explains each have no robust impact on test scores, from science to collaborative problem solving. If teaching style were

[9] (Heckman and Kautz 2013) refer to the behaviors used in (Lleras 2008) as character skills.

[10] PISA provides a sophisticated set of test outcomes based on Item Response Theory for each subject. Although using all ten outcomes, referred to as plausible values, is theoretically correct, we solely use the first plausible value, because using more does not substantially alter the results in large samples (OECD 2009, p. 46).

to impact any subject, we would expect them to affect science in the 2015 PISA data because the questions are about the styles used in science lessons. Collaborative problem solving, too, is more likely to be affected by group discussion, because as indicated by its name, it involves working together to problem solve.[11]

Only a couple of the teacher or school variables correlate significantly to test scores. More: teachers without a bachelor's degree, students per teacher, and physical education, each generally relate to lower scores – the student to teacher ratio does not affect collaborative problem solving. Years of teacher experience is generally positively related to scores, while the proportion of teachers that are fully certified often correlates negatively (unexpectedly). The average socio-economic and cultural status of students within schools correlates positively. A presentation of the relations for individual controls is omitted for brevity.

[11] "Collaborative problem solving is the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution (OECD 2017b, p. 43)."

Table 5. Regressions of test scores on teaching practices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Math		Reading		Science		Collab. Probl.	
	Rnd. Intcp.	Lew. Ltd.	Rnd. Intcp.	Lew. Ltd.	Rnd. Intcp.	Lew. Ltd.	Rnd. Intcp.	Lew. Ltd.
Group discussion	2.526	12.917**	-1.104	8.599	2.531	11.905	-2.301	6.567
	(4.094)	(6.555)	(4.507)	(7.839)	(4.312)	(7.741)	(4.303)	(7.763)
Teacher explains	-1.504	-1.834	2.617	-3.616	-2.142	-7.419	2.965	5.805
	(4.501)	(7.905)	(5.149)	(9.546)	(4.476)	(7.933)	(4.652)	(8.586)
Female Teacher Ratio	4.013	3.521	-0.411	-0.341	-2.558	-3.361	10.961	10.236
	(9.581)	(9.282)	(11.004)	(10.801)	(10.010)	(9.690)	(10.285)	(10.109)
Average Teacher Age	-0.950	-1.048	-0.952	-1.027	-0.921	-1.022	-1.598*	-1.759**
	(0.861)	(0.866)	(0.869)	(0.857)	(0.893)	(0.904)	(0.898)	(0.897)
Avg. Teacher Exp. (Yrs.)	0.903	1.019	1.672**	1.767**	1.490*	1.592*	1.981**	2.156***
	(0.794)	(0.796)	(0.822)	(0.814)	(0.846)	(0.842)	(0.826)	(0.830)
Teach. Share < Bach.	-86.217***	-77.378***	-112.679***	-100.062***	-105.509***	-95.189***	-69.411***	-63.372***
	(20.133)	(19.515)	(20.512)	(19.961)	(18.805)	(18.084)	(17.586)	(17.093)
Teach. Share > Bach.	9.972	9.940	-0.303	0.441	0.607	0.812	3.339	3.229
	(8.571)	(8.549)	(9.770)	(9.610)	(9.020)	(8.893)	(11.377)	(11.283)
Teach. Share Fully Cert.	-4.726	-4.439	-7.186	-6.809	-7.957*	-8.067*	-9.251*	-8.830*
	(5.028)	(4.998)	(4.976)	(4.954)	(4.361)	(4.627)	(5.386)	(5.184)
Sch: Socioecon. & Cult.	33.894***	34.356***	34.619***	35.531***	34.863***	35.275***	31.707***	31.535***
	(2.905)	(2.841)	(3.601)	(3.515)	(3.224)	(3.042)	(2.980)	(2.907)
Student-Teacher Ratio	-0.477**	-0.441**	-0.400*	-0.295	-0.566***	-0.464**	-0.030	-0.010
	(0.199)	(0.195)	(0.230)	(0.221)	(0.216)	(0.200)	(0.228)	(0.221)
Hindered: Lack of Res.	-1.654	-1.550	-2.003	-1.689	-2.094	-1.877	-2.427	-2.327
	(1.791)	(1.740)	(1.884)	(1.828)	(2.006)	(1.900)	(1.783)	(1.729)
Hindered: Lack of Staff	-0.289	0.070	-1.480	-1.252	-2.127	-1.882	-0.585	-0.278
	(1.643)	(1.633)	(2.014)	(1.979)	(1.758)	(1.732)	(1.891)	(1.843)
Rural	1.428	2.338	-0.885	-0.314	3.567	3.797	-2.942	-1.707
	(3.730)	(3.529)	(3.893)	(3.642)	(3.514)	(3.294)	(3.786)	(3.677)
All Male or Female Studs.	1.258	0.803	2.024	2.254	2.046	1.936	1.677	0.654
	(5.593)	(5.699)	(5.001)	(5.000)	(5.344)	(5.316)	(4.798)	(4.837)
Stu. Grouped by Ability	-2.707	-2.935	-0.941	-0.864	0.259	0.334	2.230	2.107
	(2.493)	(2.512)	(2.660)	(2.690)	(2.509)	(2.530)	(2.385)	(2.407)
Tests Used to Group	-0.142	-0.280	-3.792	-4.362	-2.691	-3.313	-3.843	-3.866
	(3.176)	(3.166)	(3.702)	(3.603)	(3.304)	(3.187)	(2.888)	(2.849)
Tests used to Adapt Teaching	1.543	2.015	3.271	2.664	3.706	2.585	3.518	3.977
	(4.535)	(4.532)	(4.245)	(4.205)	(4.545)	(4.334)	(4.018)	(4.063)
Hindered: Teach. Behavior	0.001	0.021	1.739	1.696	1.118	1.030	1.779	1.894
	(1.824)	(1.834)	(1.754)	(1.739)	(1.657)	(1.683)	(1.489)	(1.474)
Satisfaction of Teachers	-1.136	-4.826	5.815	3.854	5.196	3.439	-0.967	-3.775
	(6.949)	(7.151)	(9.332)	(9.186)	(7.841)	(7.846)	(7.576)	(7.859)
Attend Phys. Ed. (Days)	-1.453**	-0.851	-2.653***	-2.548***	-2.548***	-2.312***	-2.505***	-2.488***
	(0.658)	(0.622)	(0.750)	(0.751)	(0.653)	(0.651)	(0.870)	(0.823)
Constant	256.161***	247.247***	247.753***	247.506***	250.672***	254.582***	308.669***	295.585***
	(36.316)	(42.735)	(42.851)	(52.242)	(38.591)	(44.733)	(40.393)	(45.141)
Individual Controls	yes	yes	yes	yes	yes	yes	yes	yes
Random School Intercept	yes	-	yes	-	yes	-	yes	-
Country Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Total Obs.	35126	35126	35126	35126	35126	35126	35126	35126
Number of Schools	1489	1489	1489	1489	1489	1489	1489	1489
Hansen J p-value		0.217		0.106		0.224		0.193
Kleibergen-Paap F		18.794		18.794		18.794		18.794

Notes:
Clustered standard errors in parentheses (by school); * p<0.10 ** p<0.05 *** p<0.01

4.2. Noncognitive Skills

Table 6 presents the results for noncognitive skills. Group discussion contributes to positive outcomes, but not across each. Teacher explains does not affect any of the outcomes. In the first panel, more group discussion leads to students skipping class less often and paying more attention when there. For each outcome there are three columns, the first shows the same random intercept specification as before, including all of the control variables, while the second and third columns present results based on the Lewbel method. The first Lewbel specification uses the full set of generated instruments, which are generally likely to be valid based on the Hansen J overidentification test. However, they do not strictly pass the overid test (e.g., Column 5 for Classroom Attention). To overcome this limitation, a restricted set of instruments were used[12], which, presented in Lew. Ltd. columns, appear to be valid for each outcome (i.e., we fail to reject their excludability).

There is qualified support for classroom discussion positively affecting two other outcomes - students feel better able to make friends and are less likely to be bullied. However, classroom discussion does not affect studying, student expectations of future achievement, sense of belonging, or bad student behavior[13].

Accounting for the multiple outcomes / hypotheses, it is warranted to consider a correction. The typical Bonferroni correction divides the critical value by the number of hypotheses. Thus, with eight outcomes, the 5-percent significance levels becomes 0.00625. Using this standard, the improvement of classroom attention remains significant, but not the others.

Thus group discussion leads to positive classroom behavior, at least greater attention. It does not lead to any bad behavior, which might be expected if teachers have more difficulty maintaining discipline when not the center of attention. Positive effects on noncognitive skills support the view that group discussion promotes a classroom and school environment that leads to positive outcomes that in turn support positive later life outcomes.

[12] The limited set of instruments are based solely on individual characteristics, excluding teacher and school covariates.

[13] The first column of bad student behavior is missing because the model did not converge in expectation-maximization, which is used before switching to gradient-based optimization.

Table 6. Regressions of noncognitive skills on teaching practices

	(1) Rnd. Intcp. Skipped Class	(2) Lewbel Skipped Class	(3) Lew. Ltd. Skipped Class	(4) Rnd. Intcp. Class. Attention	(5) Lewbel Class. Attention	(6) Lew. Ltd. Class. Attention
Group discussion	-0.038*	-0.040*	-0.032	0.092***	0.102***	0.169***
	(0.021)	(0.023)	(0.045)	(0.027)	(0.030)	(0.054)
Teacher explains	0.034	0.037	0.031	-0.005	-0.004	-0.016
	(0.025)	(0.030)	(0.065)	(0.028)	(0.033)	(0.068)
Total Obs.	34896	34896	34896	31773	31773	31773
Hansen J p-value		0.489	0.183		0.060	0.147
Kleibergen-Paap F		208.805	18.505		195.171	17.949
Converged	yes			yes		
	Study	Study	Study	Exp. Educ.	Exp. Educ.	Exp. Educ.
Group discussion	0.023	0.027	0.034	0.039	0.049	0.137
	(0.026)	(0.029)	(0.051)	(0.056)	(0.058)	(0.089)
Teacher explains	-0.016	-0.027	-0.036	0.093	0.085	-0.087
	(0.032)	(0.038)	(0.072)	(0.071)	(0.076)	(0.116)
Total Obs.	34909	34909	34909	35086	35086	35086
Hansen J p-value		0.382	0.521		0.116	0.613
Kleibergen-Paap F		208.989	18.785		209.398	18.850
Converged	yes			yes		
	Friends	Friends	Friends	Belonging	Belonging	Belonging
Group discussion	0.038**	0.040*	0.075*	0.026	0.027	0.035
	(0.019)	(0.022)	(0.041)	(0.031)	(0.033)	(0.067)
Teacher explains	-0.025	-0.019	-0.023	-0.023	-0.031	-0.109
	(0.024)	(0.029)	(0.058)	(0.036)	(0.042)	(0.090)
Total Obs.	34984	34984	34984	35099	35099	35099
Hansen J p-value		0.561	0.515		0.179	0.238
Kleibergen-Paap F		209.809	18.811		211.282	18.846
Converged	yes			yes		
	Bullied	Bullied	Bullied		Bad Stud. Beh.	Bad Stud. Beh.
Group discussion	-0.024	-0.032*	-0.059*		-0.228	-0.360
	(0.017)	(0.018)	(0.033)		(0.164)	(0.239)
Teacher explains	0.018	0.014	0.022		0.020	0.033
	(0.018)	(0.021)	(0.040)		(0.180)	(0.237)
Total Obs.	34508	34508	34508		35126	35126
Hansen J p-value		0.285	0.663		0.266	0.176
Kleibergen-Paap F		203.403	18.611		210.243	18.794
Converged	yes					

Notes: Each regression includes individual, teacher, and school controls.
Clustered standard errors in parentheses (by school); * p<0.10 ** p<0.05 *** p<0.01

5. Conclusion

Students in schools that more frequently use the teaching practice *group discussion* are more satisfied with their lives, they also pay more attention in class. Group discussion does not

contribute to tests scores in the short run, either positively or negatively; however, in the literature, adult life satisfaction contributes positively to both economic and non-economic outcomes. For instance, more satisfied people earn higher wages, are less likely to be unemployed, and live longer.[14] This evidence, when combined with the evidence that the best predictor of adult well-being is their well-being as children (Clark et al. 2019), suggests group discussion may have long-run positive effects on student outcomes. What is more, the results suggest group discussion leads to the development of noncognitive skills/behavior, which also contribute to long-run positive outcomes (ter Weel 2008). The prevalence of lecturing ("teacher explains"), on the other hand, does not affect life satisfaction, test scores, nor noncognitive skills.

We add to the literature, for the first time rigorously demonstrating the potential for positive impacts of teaching practices on life satisfaction. The most similar paper, (Brulé and Veenhoven 2014) did not find group discussion to be positively related to life satisfaction, but their null result could be due to the imprecision of their cross-country empirical design. Instead, the next closest paper, (Algan et al. 2013) offers broadly consistent evidence, showing group discussion is positively related to social capital, which in turn is strongly related to life satisfaction.

The average relation does not apply to girls however – the group discussion - life satisfaction relation is statistically insignificant. It is not clear why girls respond differently. It seems the benefits of classroom groups do not equally accrue to girls when boys are present. Indeed, further analysis shows that girls do benefit from group discussion when in all-girl schools, and the relationship is even larger than for boys. The presence of boys seems to affect the psychological safety that girls feel in groups. At the same time, girls in all-girl schools are less satisfied with their lives generally, while students in single-gender schools are not statistically more or less satisfied. Also, the lecturing - life satisfaction relation turns negative for girls in coeducation schools. It is clear that gender dynamics play a significant role in determining student outcomes and should not be ignored when designing research or interventions.

Group discussion represents one teaching practice among many modern, participatory, or horizontal practices. It contrasts lecturing, a vertical method prominently used in the mass education systems developed in the nineteenth century. Broadly speaking, our results suggest a greater prevalence of modern-horizontal methods contributes positively to student outcomes,

[14] See (De Neve et al. 2013) for a summary of supporting articles.

while the prevalence of traditional-vertical methods does not affect outcomes one way or another. However, different scholars use different definitions. Our results are best understood in terms of our variable definitions, though they may apply more broadly.

More group discussion could lead to long-lasting positive effects on society. If policymakers are interested in the well-being of their citizens, then simply instructing teachers to use more group discussion could help them achieve that end. Such an intervention is cost effective. It requires minimal investment and the impact is not small. A one standard deviation increase in the prevalence of group discussion has an impact that is about one quarter of the negative relation for grade repetition. Future research should delve further into the characteristics of group discussion, especially assessing the gender dynamics.

REFERENCES

Algan Y, Cahuc P, Shleifer A (2013) Teaching practices and social capital. Am Econ J Appl Econ 5:189–210. https://doi.org/10.1257/app.5.3.189

Almlund M, Duckworth AL, Heckman J, Kautz T (2011) Personality Psychology and Economics. In: Handbook of the Economics of Education. pp 1–181

Arampatzi E, Burger MJ, Ianchovichina E, et al (2018) Unhappy development: dissatisfaction with life on the eve of the Arab Spring. Review of Income and Wealth 64:S80–S113. https://doi.org/10.1111/roiw.12388

Baum CF, Lewbel A, Schaffer ME, Talavera O (2013) Instrumental variables estimation using heteroskedasticity-based instruments. Stata Users Group

Bietenbeck J (2014) Teaching practices and cognitive skills. Labour Econ 30:143–153. https://doi.org/10.1016/j.labeco.2014.03.002

Borghans L, Duckworth AL, Heckman JJ, Weel B (2008) The Economics and Psychology of Personality Traits. Journal of Human Resources 43:972–1059

Brulé G, Veenhoven R (2014) Participatory Teaching and Happiness in Developed Nations. Adv Appl Sociol 04:235–245. https://doi.org/10.4236/aasoci.2014.411028

Capps DK, Crawford BA, Constas MA (2012) A Review of Empirical Literature on Inquiry Professional Development: Alignment with Best Practices and a Critique of the Findings. J Sci Teacher Educ 23:291–318. https://doi.org/10.1007/s10972-012-9275-2

Chetty R, Friedman JN, Rockoff JE (2014a) Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. American Economic Review 104:2593–2632. https://doi.org/10.1257/aer.104.9.2593

Chetty R, Friedman JN, Rockoff JE (2014b) Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. American Economic Review 104:2633–2679. https://doi.org/10.1257/aer.104.9.2633

Clark AE, Fleche S, Layard R, et al (2019) The Origins of Happiness: The Science of Well-Being over the Life Course. Princeton University Press, Princeton

Clark AE, Frijters P, Shields MA (2008) Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. J Econ Lit 46:95–144. https://doi.org/10.1257/jel.46.1.95

Cordero JM, Cristobal V, Gil M Teaching strategies and their effect on student achievement: A cross-country study using data from PISA 2015

De Neve J-E, Diener E, Tay L, Xuereb C (2013) The Objective Benefits of Subjective Well-Being. In: Helliwell JF, Layard R, Sachs J (eds) World Happiness Report 2013. UN Sustainable Development Solutions Network, New York, pp 56–79

De Neve J-E, Oswald AJ (2012) Estimating the influence of life satisfaction and positive affect on later income using sibling fixed effects. Proc Natl Acad Sci U S A 109:19953–19958. https://doi.org/10.1073/pnas.1211437109

Dewey J (1944) Democracy and Education. Dover, New York

Di Tella R, MacCulloch RJ, Oswald AJ (2003) The Macroeconomics of Happiness. Rev Econ Stat 85:809–827. https://doi.org/10.1162/003465303772815745

Dubin JA, McFadden DL (1984) An Econometric Analysis of Residential Electric Appliacne Holdings and Consumption. Econometrica 52:345–362

Easterlin RA (2001) Income and Happiness, towards a unified theory. Economic Journal 111:465–484

Easterlin RA, O'Connor KJ (2021) The Easterlin Paradox. In: Zimmermann K (ed) Handbook of Labor, Human Resources and Population Economics. Springer

Echazarra A, Salinas D, Mendez I, et al (2016) How teachers teach and students learn: Successful strategies for school

Edmondson A (1999) Psychological Safety and Learning Behavior in Work Teams Author. Adm Sci Q 44:350–383

Flèche S (2017) Teacher Quality, Test Scores and Non-Cognitive Skills: Evidence from Primary School Teachers in the UK

Frijters P, Clark AE, Krekel C, Layard R (2020) A happy choice: wellbeing as the goal of government. Behavioural Public Policy 1–40. https://doi.org/10.1017/bpp.2019.39

Goldhaber DD, Brewer DJ (1997) Why don't schools and teachers seem to matter?: Assessing the impact of unobservables on educational productivity. Journal of Human Resources 32:505–523. https://doi.org/10.2307/146181

Graham C, Eggers A, Sukhtankar S (2004) Does happiness pay? An exploration based on panel data from Russia. J Econ Behav Organ 55:319–342. https://doi.org/10.1016/j.jebo.2003.09.002

Graham C, Laffan K, Pinto S (2018) Well-being in metrics and policy. Science (1979) 362:287–288. https://doi.org/10.1126/science.aau5234

Granovetter MS (1973) The Strength of Weak Ties. American Journal of Sociology 78:1360–1380

Guven C (2011) Are Happier People Better Citizens? Kyklos 64:178–192. https://doi.org/10.1111/j.1467-6435.2011.00501.x

Guven C (2012) Reversing the question: Does happiness affect consumption and savings behavior? J Econ Psychol 33:701–717. https://doi.org/10.1016/j.joep.2012.01.002

Hanushek EA (1986) The Economics of Schooling: Production and Efficiency in Public Schools. J Econ Lit 24:1141–1177. https://doi.org/http://www.jstor.org/stable/2725865

Hanushek EA (2009) Teacher Deselection. In: Goldhaber DD, Hannaway J (eds) Creating A New Teaching Profession. Urban Institute Press, Washington DC, pp 165–180

Heckman JJ, Kautz T (2013) Fostering and Measuring Skills: Interventions that Improve Character and Cognition

Heckman JJ, Kautz T (2012) Hard evidence on soft skills. Labour Econ 19:451–464. https://doi.org/10.1016/j.labeco.2012.05.014

Helliwell JF (2019) Measuring and using happiness to support public policies

Helliwell JF, Aknin LB (2018) Expanding the social science of happiness. Nat Hum Behav 2:248–252. https://doi.org/10.1038/s41562-018-0308-5

Jackson CK (2018) What do test scores miss? The importance of teacher effects on non–test score outcomes. Journal of Political Economy 126:2072–2107. https://doi.org/10.1086/699018

Kahneman D, Krueger AB, Schkade D, et al (2004) Toward National Well-Being Accounts. American Economic Review 94:429–434

Krause A (2013) Don't worry, be happy? Happiness and reemployment. J Econ Behav Organ 96:1–20. https://doi.org/10.1016/j.jebo.2013.09.002

Lavy V (2016) What makes an effective teacher? Quasi-experimental evidence. CESifo Econ Stud 62:88–125. https://doi.org/10.1093/cesifo/ifv001

Le Moglie M, Mencarini L, Rapallini C (2015) Is it just a matter of personality? On the role of subjective well-being in childbearing behavior. J Econ Behav Organ 117:453–475. https://doi.org/10.1016/j.jebo.2015.07.006

Lewbel A (2012) Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. Journal of Business & Economics Statistics 30:67–80. https://doi.org/10.1080/07350015.2012.643126
Lleras C (2008) Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. Soc Sci Res 37:888–902. https://doi.org/10.1016/j.ssresearch.2008.03.004
Luttmer EFP (2005) Neighbors as Negatives : Relative Earnings and Well-Being. Quarterly Journal of Economics 120:963–1002
Mohanty MS (2009) Effects of positive attitude on earnings: Evidence from the US longitudinal data. Journal of Socio-Economics 38:357–371. https://doi.org/10.1016/j.socec.2008.07.012
Montessori M (2004) The Discovery of the Child. Aakar Books
National Research Council (U.S.) (1996) National Science Education Standards. National Academies Press, Washington, D.C.
NCTM (1989) Curriculum and Evaluation Standards for School Mathematics. Reston, VA
NCTM (1991) Professional Standards for Teaching Mathematics. Reston, VA.
O'Connor KJ (2023) Measuring Societal Progress
O'Connor KJ (2020a) Life satisfaction and noncognitive skills: effects on the likelihood of unemployment. Kyklos 73:568–604. https://doi.org/10.1111/kykl.12226
O'Connor KJ (2020b) The effect of immigration on natives' well-being in the European Union. J Econ Behav Organ 180:. https://doi.org/10.1016/j.jebo.2020.10.006
O'Connor KJ, Graham C (2019) Longer, More Optimistic, Lives: Historic Optimism and Life Expectancy in the United States. J Econ Behav Organ 168:374–392. https://doi.org/10.1016/j.jebo.2019.10.018
OECD (2017a) PISA 2015 Results (Volume III): Students' Well-Being. OECD Publishing, Paris
OECD (2016) PISA 2015 Results (Volume 1): Excellence and Equity in Education. Paris
OECD (2009) PISA Data Analysis Manual: SPSS Second Edition. Paris
OECD (2017b) PISA 2015 Technical Report. Paris
Opdenakker MC, Van Damme J (2006) Teacher characteristics and teaching styles as effectiveness enhancing factors of classroom practice. Teach Teach Educ 22:1–21. https://doi.org/10.1016/j.tate.2005.07.008
Origo F, Pagani L (2009) Flexicurity and job satisfaction in Europe: The importance of perceived and actual job stability for well-being at work. Labour Econ 16:547–555. https://doi.org/10.1016/j.labeco.2009.02.003
Oswald AJ, Proto E, Sgroi D (2015) Happiness and Productivity. J Labor Econ 33:789–822. https://doi.org/10.1086/298035
Piekałkiewicz M (2017) Why do economists study happiness? The Economic and Labour Relations Review 28:361–377. https://doi.org/10.1177/1035304617717130
Roberts BW, Kuncel NR, Shiner R, et al (2007) The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. Perspectives on Psychological Science 2:313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x
Ryan RM, Sheldon KM, Kasser T, Deci EL (1996) All goals are not created equal: An organismic perspective on the nature of goals and their regulation. In: Gollwitzer PM, Bargh JA (eds) The psychology of action: Linking cognition and motivation to behavior. The Guilford Press, pp 7–26

Sarracino F, Fumarco L (2018) Assessing the Non-financial Outcomes of Social Enterprises in Luxembourg. Journal of Business Ethics. https://doi.org/10.1007/s10551-018-4086-9
Schwerdt G, Wuppermann AC (2011) Is traditional teaching really all that bad? A within-student between-subject approach. Econ Educ Rev 30:365–379. https://doi.org/10.1016/j.econedurev.2010.11.005
Soto CJ (2015) Is Happiness Good for Your Personality? Concurrent and Prospective Relations of the Big Five With Subjective Well-Being. J Pers 83:45–55. https://doi.org/10.1111/jopy.12081
Specht J, Egloff B, Schmukle SC (2012) Examining Mechanisms of Personality Maturation: The Impact of Life Satisfaction on the Development of the Big Five Personality Traits. Soc Psychol Personal Sci 4:181–189. https://doi.org/10.1177/1948550612448197
Tay L, Kuykendall L, Diener E (2015) Satisfaction and Happiness – The Bright Side of Quality of Life. In: Glatzer W, Camfield L, MØller V, Rojas M (eds) Global Handbook of Quality of Life: Exploration of Well-Being of Nations and Continents. Springer Netherlands, pp 839–853
ter Weel B (2008) The Noncognitive Determinants of Labor Market and Behavioral Outcomes: Introduction to the Symposium. Journal of Human Resources 43:729–737
The Global Happiness Council (2018) Global Happiness Policy Report 2018. Sustainable Development Solutions Network, New York
UNICEF (2009) Child-Friendly Schools Manual. UNICEF, New York

Appendix

Table 7. Table 1 including individual coefficients

	(1) Life Sat.	(2) Life Sat.	(3) Life Sat.	(4) Life Sat.
Female	-0.327***	-0.327***	-0.324***	-0.316***
	(0.047)	(0.047)	(0.046)	(0.046)
Younger Rel. Age (mths)	-0.009	-0.010	-0.010	-0.009
	(0.008)	(0.008)	(0.008)	(0.008)
Repeated Grade	-0.234***	-0.229***	-0.232***	-0.225***
	(0.069)	(0.069)	(0.069)	(0.069)
Grade	-0.093**	-0.073	-0.065	-0.048
	(0.045)	(0.046)	(0.047)	(0.047)
Native	0.170	0.164	0.161	0.135
	(0.193)	(0.192)	(0.191)	(0.189)
Mother ISCED 1	-0.172	-0.163	-0.163	-0.165
	(0.115)	(0.115)	(0.115)	(0.115)
Mother ISCED 2	-0.074	-0.068	-0.066	-0.071
	(0.110)	(0.111)	(0.109)	(0.109)
Mother ISCED 3	-0.151	-0.144	-0.131	-0.129
	(0.113)	(0.114)	(0.112)	(0.113)
Mother ISCED 4	-0.157	-0.152	-0.145	-0.147
	(0.131)	(0.132)	(0.131)	(0.132)
Mother ISCED 5	-0.117	-0.114	-0.100	-0.100
	(0.115)	(0.116)	(0.115)	(0.116)
Mother ISCED 6	-0.108	-0.108	-0.095	-0.097
	(0.151)	(0.152)	(0.151)	(0.153)
Father ISCED 1	0.274*	0.268*	0.269*	0.271*
	(0.151)	(0.151)	(0.150)	(0.149)
Father ISCED 2	0.143	0.139	0.141	0.142
	(0.129)	(0.129)	(0.128)	(0.127)
Father ISCED 3	0.007	0.005	0.013	0.015
	(0.133)	(0.132)	(0.132)	(0.130)
Father ISCED 4	-0.036	-0.037	-0.028	-0.032
	(0.152)	(0.152)	(0.151)	(0.150)
Father ISCED 5	0.005	-0.000	0.014	0.016
	(0.149)	(0.148)	(0.148)	(0.146)
Father ISCED 6	-0.116	-0.124	-0.106	-0.102
	(0.188)	(0.187)	(0.187)	(0.186)
Father Native	0.089	0.084	0.083	0.093
	(0.120)	(0.121)	(0.121)	(0.120)
Mother Native	0.025	0.023	0.028	0.034
	(0.117)	(0.116)	(0.116)	(0.115)
Socioecon. & Cult. Status	0.166***	0.173***	0.195***	0.192***
	(0.045)	(0.045)	(0.047)	(0.047)
Books	0.034	0.033	0.036	0.038*
	(0.022)	(0.022)	(0.022)	(0.022)
Computer Available	0.321***	0.327***	0.334***	0.337***
	(0.064)	(0.064)	(0.064)	(0.064)
Group discussion	0.248***	0.257***	0.219***	0.186***
	(0.062)	(0.064)	(0.064)	(0.066)
Teacher explains	-0.032	-0.028	0.000	-0.002
	(0.078)	(0.083)	(0.082)	(0.082)

Table 7. Continued

	(1) Life Sat.	(2) Life Sat.	(3) Life Sat.	(4) Life Sat.
Female Teacher Ratio		0.174	0.256	0.236
		(0.169)	(0.175)	(0.173)
Average Teacher Age		-0.013	-0.015	-0.015
		(0.017)	(0.017)	(0.016)
Avg. Teacher Exp. (Yrs.)		0.020	0.023	0.020
		(0.017)	(0.017)	(0.016)
Teach. Share < Bach.		0.489*	0.170	0.151
		(0.291)	(0.284)	(0.283)
Teach. Share > Bach.		-0.135	-0.007	0.022
		(0.176)	(0.161)	(0.163)
Teach. Share Fully Cert.		-0.039	-0.040	-0.029
		(0.130)	(0.127)	(0.127)
Sch: Socioecon. & Cult.			-0.197***	-0.204***
			(0.068)	(0.067)
Student-Teacher Ratio			-0.004	-0.002
			(0.004)	(0.004)
Hindered: Lack of Res.			-0.023	-0.011
			(0.030)	(0.029)
Hindered: Lack of Staff			-0.048*	-0.028
			(0.029)	(0.033)
Rural			0.018	0.004
			(0.065)	(0.063)
All Male or Female Studs.			-0.013	0.009
			(0.091)	(0.089)
Tests Used to Group				0.037
				(0.054)
Tests used to Adapt Teaching				0.202***
				(0.060)
Hindered: Teach. Behavior				-0.022
				(0.027)
Satisfaction of Teachers				0.356**
				(0.147)
Attend Phys. Ed. (Days)				0.047**
				(0.021)
Constant	8.001***	7.948***	7.757***	6.241***
	(0.543)	(0.705)	(0.718)	(0.863)
Individual Controls	yes	yes	yes	yes
Random School Intercept	yes	yes	yes	yes
Country Fixed Effects	yes	yes	yes	yes
Total Obs.	35126	35126	35126	35126
Number of Schools	1489	1489	1489	1489

Notes: Life satisfaction has a mean of 7.13 and standard deviation of 2.26. Group discussion has a mean of 2.75 and standard deviation of 0.48.
Clustered standard errors in parentheses (by school); * p<0.10 ** p<0.05 *** p<0.01

Question Details:
Class Attention. Index is the average over the following 5 questions.

Students don’t listen to what the teacher says.
There is noise and disorder.
The teacher has to wait a long time for students to quiet down.
Students cannot work well.
Students don’t start working for a long time after the lesson begins.

The scale is in reverse from every lesson (1) to never or hardly ever (5). Greater index scores indicate students pay more attention in class.

Study takes the values "Study, 1: morning or night, 2: both"