

Bartolini, Stefano; O'Connor, Kelsey J.

Working Paper

Effects of teaching practices on life satisfaction and test scores: evidence from the Program for International Student Assessment (PISA)

GLO Discussion Paper, No. 1009

Provided in Cooperation with:

Global Labor Organization (GLO)

Suggested Citation: Bartolini, Stefano; O'Connor, Kelsey J. (2022) : Effects of teaching practices on life satisfaction and test scores: evidence from the Program for International Student Assessment (PISA), GLO Discussion Paper, No. 1009, Global Labor Organization (GLO), Essen

This Version is available at:

<https://hdl.handle.net/10419/248473>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Effects of teaching practices on life satisfaction and test scores: evidence from the Program for International Student Assessment (PISA)

By STEFANO BARTOLINI* AND KELSEY J. O'CONNOR**

Schools are ripe for policy intervention. We demonstrate that implementing different teaching practices is effective, finding a greater prevalence of group discussion used in schools positively affects students' life satisfaction and noncognitive skills but has no impact on test scores. The benefits do not apply to girls, however, unless they attend all-girl schools. These findings are based on a sample from the 2015 PISA which includes more than 35 thousand students from approximately 1500 schools in 14 countries or regions. We perform regressions of student life satisfaction on the prevalence of group discussion and lecturing used in their school, including a battery of individual, teacher, and school controls, as well as random intercepts by school. For robustness we use instrumental variables and methods to account for school-selection. The average impact of group discussion is not small – a one standard deviation leads to an increase in life satisfaction that is about one-half of the negative association with grade repetition. On the other hand, more or less lecturing does not affect life satisfaction, noncognitive skills, nor test scores. We conclude that teaching practices – group discussion – can be used to improve student life satisfaction, thereby likely positively affecting future economic outcomes and well-being.

Keywords: subjective well-being, teaching, teaching practices, non-cognitive skills, education, PISA, participatory teaching, vertical teaching, horizontal teaching

JEL codes: I21, I31, J24

* Professor of Economics, University of Siena. 7 Piazza San Francesco, 53100 Siena, Italy. Phone: +39 0577 233049. Email: stefano.bartolini@unisi.it

** Economics Researcher, STATEC Research (National Institute of Statistics and Economic Studies); GLO Fellow, Global Labor Organization (GLO); Research Affiliate, Institute of Labor Economics (IZA); Senior Research Associate, School of Economics, University of Johannesburg. 14 Rue Erasme, L-1468 Luxembourg, Luxembourg. Phone: +352 247 84351. Email: Kelsey.OConnor@statec.etat.lu.

Acknowledgements: The authors would like to thank Francesco Sarracino for comments on previous drafts, and also seminar participants for helpful comments. O'Connor gratefully acknowledge the financial support of the Observatoire de la Compétitivité, Ministère de l'Economie, DG Compétitivité, Luxembourg, and STATEC. Views and opinions expressed in this article are those of the authors and do not reflect those of STATEC, or funding partners. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

I. Introduction

You likely recall sitting in a lecture hall. To this day there is a high prevalence of lecturing. It arose in the nineteenth century as society developed mass schooling and many administrators continue to rely on it because it is cost effective or simply because they have not considered alternatives. Did you enjoy sitting in lecture halls? Should we still rely on lecturing as much as we do? It was effective, but the needs of society have changed. We need evidence, evidence on student outcomes that go beyond short-term academic achievement.

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). Schools are ripe for intervention. Teachers are among the most important inputs to education (Hanushek 2009; 1986). Their impacts extend beyond academic outcomes – “The effect of the school [on emotional well-being] is remarkable ... (Clark et al. 2019, 188).” Teaching practices can be adopted. The infrastructure is in place and there is a long history of educational interventions (Heckman and Kautz 2013).

What sets this paper apart is the focus on life satisfaction. We chose life satisfaction as an important outcome in its own right (Graham, Laffan, and Pinto 2018; Frijters et al. 2020; The Global Happiness Council 2018; Kahneman et al. 2004; Helliwell 2019), but also because it promotes social capital (Gruen 2011) and traditional-economic outcomes, including for instance: wages (Graham, Eggers, and Sukhtankar 2004; De Neve and Oswald 2012; Oswald, Proto, and Sgroi 2015), employment (O’Connor 2020a; Krause 2013) consumption and savings (Gruen 2012), and health (Tay, Kuykendall, and Diener 2015), among other positive behavioral outcomes. See De Neve et al. (2013) or Piekalkiewicz (2017) for a summary of articles.¹ What is more, life satisfaction is closely related to noncognitive skills (Specht, Egloff, and Schmukle 2012; Soto 2015), which are likewise important predictors of later life outcomes (Lleras 2008), yet they are not necessarily picked up in test scores (Jackson 2018).

How might teaching practices affect life satisfaction? Lecturing, as a traditional form of teaching used in mass schooling, was *not* intended to promote student well-being or noncognitive skills. Mass schooling was based on the need for social control. Its basic goal was to train future blue-

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, 223).

collar workers and soldiers. Education had to develop in students the ability to obey and be bored, skills highly appreciated by the industrial labor market and the army. Teaching was vertical, stemming from and focused on the teacher. It was based on discipline, sticks and carrots, competition, physical segregation, immobility, passivity, and cognition. Students were vertically taught cognitive notions required by an industrializing society, such as writing, reading and simple calculations. This exclusive focus on cognitive skills was based on positivist ideas dating back to the late 1800s, which saw emotions as completely distinct from intelligence and a hindrance to cognitive activity. These are outmoded ideas. Since then, everything has changed. We know now that cognition requires an active role of emotions and that they are mutually reinforcing (Mayer, Salovey, and Caruso 2000; Damasio 1995).

Vertical teaching has been increasingly challenged for its alleged negative contribution to serious problems in developed countries. The latter are plagued by extensive youth distress, low self-esteem, loneliness, troubled relationships, low cooperative capacity, and poor access to the labor market (Schor 2004; Twenge 2000; 2015; Berguno et al. 2004).²

Group discussion contrasts vertical teaching practices such as lecturing. 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 safe³ 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 senses of autonomy, mastery, and self-efficacy. Students develop an implicit understanding of reciprocity amongst each other.

² Schor (2004) documented the worsening well-being of American kids and teens. Twenge, depicted a disquieting picture of evolution on mental health among American youngsters. Twenge et al. (2010) assessed their evolution over many decades using a survey, administered to high school and college students since the 1930s, including many questions regarding symptoms of depression and anxiety. The authors found that depression and anxiety rose significantly in the period 1938-2007 in the US. Another survey showed that between 1982 and 2013, symptoms of depression and anxiety rose significantly in a nationally representative sample of US 12th graders (Twenge 2015). Using measures of anxiety dating back to (and unchanged since) the 1950s, Twenge (2000) showed that normal American kids and teens in the 1980s scored as high on the anxiety scale as children in psychiatric care in the 1950s.

As for social capital, loneliness and bullying soared among youngsters, bringing the share of young people under 18 years of age reporting to experience loneliness to 80% (Berguno et al. 2004). Difficult access to the labor market by young people became commonplace in the West. As an example, in June 2020 the youth unemployment rate in the euro area was 17%, much higher than the unemployment in the overall workforce (7.8%). Source:

[https://ec.europa.eu/eurostat/statistics-](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Unemployment_statistics#Youth_unemployment_trends)

[explained/index.php?title=Unemployment_statistics#Youth_unemployment_trends](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Unemployment_statistics#Youth_unemployment_trends)

³³ For a discussion of psychological safety in team performance see (Edmondson 1999).

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 where they are expected to perform, e.g., leadership and agreeableness. The economic literature on the various noncognitive skills (e.g., reciprocity) is still somewhat sparse, yet concerning the Big Five personality traits, there is evidence showing a positive (often simultaneous) relation with subjective well-being, including life satisfaction (Specht, Egloff, and Schmukle 2012; Soto 2015). Students may also simply enjoy being in group discussion. For these reasons, we expect a greater prevalence of group discussion to be associated with greater student life satisfaction.

Teaching practices differ from country to country, between schools and within schools. This variability has been exploited to investigate which teaching practices work better (Bietenbeck 2014; Cordero, Cristobal, and Gil, n.d.). Quantitative evidence shows that participatory practices have greater cognitive and noncognitive benefits than vertical methods. For instance, modern-horizontal teaching practices positively relate to teachers ability to affect math scores and noncognitive skills, while traditional-vertical teaching hinders noncognitive skills (Flèche 2017).

The impacts of teaching practices on life satisfaction remains an open question. We found only one study using internationally comparable data on teaching practices and life satisfaction (Brulé and Veenhoven 2014). This study has two main findings. First, the general populations in countries with more horizontal teaching are more satisfied with their lives. Second, *students* living in countries with more horizontal 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 horizontal 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. In another study, Algan et al. (2013) find similar results, demonstrating group discussion increases students' social capital within schools – social capital is closely related to life satisfaction (Helliwell and Akinin 2018).

We add to the literature on teaching practices by assessing their impact on students' life satisfaction, using a large scale cross-country dataset on 15 year-olds. We are the first to show that the prevalence of group discussion used in schools positively affects students' life satisfaction at

the school level. The positive effects on life satisfaction come at no cost in terms of tests' scores – they are not impacted by teaching practices, which is important because the effects could have been positive or negative (Goldhaber and Brewer 1997; Bietenbeck 2014). What is more, the benefits of group discussion extend to noncognitive skills/behavior.

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. Section IV, discusses the effects of teaching practices on test scores and noncognitive skills. Section V concludes.

II. Data and Methods

A. Program for International Student Assessment

We estimate the impact of teaching practices on life satisfaction, test scores, and noncognitive ability using data from 2015 round of PISA. The primary purpose of PISA is to measure 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, which provides the necessary information on teaching practices for our analysis.

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

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

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.

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 survey. We also exclude privately owned or operated schools.

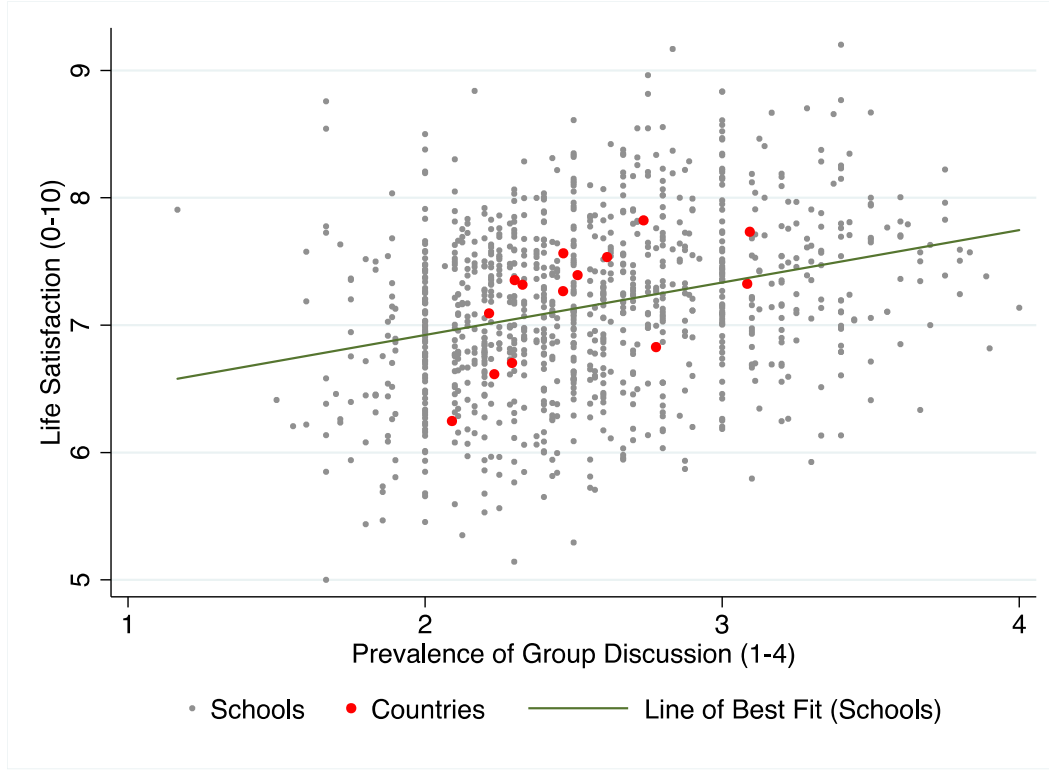
Table 1: Sample observations, schools and students within each country or region

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

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.

⁵ 20 percent was likewise used in (Cordero, Cristobal, and Gil, n.d.).

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



Source: Author calculations using PISA 2015

B. 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:

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

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. \mathbf{Teach}_{sc} , \mathbf{School}_{sc} , and \mathbf{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 school level.

The standard exogeneity conditions are necessary to estimate an unbiased relation for teaching practices. The characteristics of *individual* students, i.e., life satisfaction, should not have much of an impact on teaching practices because the latter are measured at the school level. Likewise, any omitted variables at the individual level should not strongly affect teaching practices. However, factors common to students at the school level are different. Teaching practices may be determined by the administration; teaching practices may proxy for alternative characteristics of the teachers or school; and students may have selected into schools based on the schools' characteristics. Omitted variables at the teacher/school level is addressed first using a battery of pertinent controls, discussed below. Secondly, the random intercept, μ_{sc} , accounts for residual variation between schools. The assumption is that teaching practice is orthogonal to the random intercept. Given this assumption is untestable we run additional robustness analysis 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.

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, 298).

III. Results

A. 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*** (0.062)	0.257*** (0.064)	0.219*** (0.064)	0.186*** (0.066)
Teacher explains	-0.032 (0.078)	-0.028 (0.083)	0.000 (0.082)	-0.002 (0.082)
Female Teacher Ratio		0.174 (0.169)	0.256 (0.175)	0.236 (0.173)
Average Teacher Age		-0.013 (0.017)	-0.015 (0.017)	-0.015 (0.016)
Avg. Teacher Exp. (Yrs.)		0.020 (0.017)	0.023 (0.017)	0.020 (0.016)
Teach. Share < Bach.		0.489* (0.291)	0.170 (0.284)	0.151 (0.283)
Teach. Share > Bach.		-0.135 (0.176)	-0.007 (0.161)	0.022 (0.163)
Teach. Share Fully Cert.		-0.039 (0.130)	-0.040 (0.127)	-0.029 (0.127)
Sch: Socioecon. & Cult.			-0.197*** (0.068)	-0.204*** (0.067)
Student-Teacher Ratio			-0.004 (0.004)	-0.002 (0.004)
Hindered: Lack of Res.			-0.023 (0.030)	-0.011 (0.029)
Hindered: Lack of Staff			-0.048* (0.029)	-0.028 (0.033)
Rural			0.018 (0.065)	0.004 (0.063)
All Male or Female Studs.			-0.013 (0.091)	0.009 (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*** (0.543)	7.948*** (0.705)	7.757*** (0.718)	6.241*** (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., (Clark, Frijters, and Shields 2008; Easterlin 2001; Easterlin and O'Connor 2021)).

B. 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 unbiased under the standard conditions. The limitations of the Lewbel

approach is that it is not very intuitive nor do the instruments 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 (v), (2) then generate the instruments as $(X_j - E(X_j)) * v = Z_j$. There must be heteroskedasticity in v 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 (Sarracino and Fumarco 2018; Le Moglie, Mencarini, and Rapallini 2015; Arampatzi et al. 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 (2020), based on Di Tella et al. (2003), Graham et al. (2004) and Krause (2013).

The third approach addresses selection into schools. Perhaps initially more satisfied students selected schools with more group discussion. To assess, we use a two-step procedure that adds selection controls to equation 1. 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 selection controls similar to Inverse Mills Ratios. In the second step, the modified version of equation 1 including the selection terms is estimated. The results include 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, 553), which 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)	(2)	(3)	(4)	(5)	(6)	(7)
	Rnd. Intcp. Life Sat.	Rnd. Slope Life Sat.	Lewbel Life Sat.	M-Logit Tch Style 4	OLS Life Sat.	Rnd. Intcp. Life Sat.	Rnd. Intcp. Life Sat.
Group discussion	0.249*** (0.062)	0.248*** (0.061)	0.217*** (0.072)				
Low G & T					-0.073 (0.070)	-0.083 (0.074)	-0.065 (0.072)
High G & Low T					0.080 (0.092)	0.074 (0.094)	0.067 (0.091)
High G & T					0.187* (0.096)	0.178* (0.102)	0.188** (0.096)
Teacher explains			-0.014 (0.090)				
Admis-Performance				-0.636** (0.303)			0.400 (0.487)
Admis-Feeder				0.071 (0.414)			0.121 (0.259)
Admis-Ideology				-0.213 (0.352)			-0.113 (0.225)
Admis-Program				-0.200 (0.296)			-0.086 (0.327)
Admis-Legacy				0.665** (0.297)			-0.064 (0.419)
Admis-Residence				0.151 (0.300)			0.027 (0.292)
select1					0.021 (0.022)	0.022 (0.021)	-0.152 (0.250)
select3					-0.001 (0.016)	0.001 (0.017)	0.179 (0.204)
select4					0.007 (0.015)	-0.000 (0.016)	-0.012 (0.296)
Constant	7.903*** (0.483)	7.906*** (0.482)	5.940*** (0.837)	8.249 (5.029)	5.973*** (1.038)	6.050*** (1.068)	12.048 (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

C. 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 poor 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 columns⁷. 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. Poor 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 scoring⁸ 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*** (0.066)	0.280*** (0.086)	0.178** (0.070)	0.297 (0.329)	0.862*** (0.286)	0.170** (0.078)	0.189*** (0.068)	0.287*** (0.087)	0.215*** (0.078)
Teacher explains	-0.002 (0.082)	0.171* (0.095)	0.020 (0.084)	-0.161 (0.470)	-0.174 (0.283)	0.080 (0.094)	-0.008 (0.083)	0.171* (0.095)	-0.057 (0.109)
Female	-0.316*** (0.046)	1.195*** (0.402)	-0.316*** (0.046)	-0.316*** (0.046)	-0.327*** (0.046)	-0.315*** (0.046)	-0.316*** (0.046)	1.315*** (0.409)	-0.316*** (0.046)
All Male or Female Studs.	0.009 (0.089)	0.015 (0.093)	0.004 (0.089)	0.008 (0.089)	-0.005 (0.089)	-0.007 (0.090)	-0.362 (0.814)		0.007 (0.089)
Group X Female		-0.173** (0.087)						-0.198** (0.088)	
Teach Ex. X Female		-0.336*** (0.099)						-0.351*** (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*** (0.863)	5.457*** (0.878)	6.218*** (0.874)	6.458*** (1.542)	5.299*** (1.237)	6.088*** (0.892)	6.256*** (0.864)	5.456*** (0.880)	6.363*** (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

IV. 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, 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, 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, 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, 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 Lleras, 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.

A. Test Scores

Test scores cover math, science, reading, and collaborative problem-solving skills¹⁰. 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) still 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, 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 socioeconomic 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 2017, 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 (4.094)	12.917** (6.555)	-1.104 (4.507)	8.599 (7.839)	2.531 (4.312)	11.905 (7.741)	-2.301 (4.303)	6.567 (7.763)
Teacher explains	-1.504 (4.501)	-1.834 (7.905)	2.617 (5.149)	-3.616 (9.546)	-2.142 (4.476)	-7.419 (7.933)	2.965 (4.652)	5.805 (8.586)
Female Teacher Ratio	4.013 (9.581)	3.521 (9.282)	-0.411 (11.004)	-0.341 (10.801)	-2.558 (10.010)	-3.361 (9.690)	10.961 (10.285)	10.236 (10.109)
Average Teacher Age	-0.950 (0.861)	-1.048 (0.866)	-0.952 (0.869)	-1.027 (0.857)	-0.921 (0.893)	-1.022 (0.904)	-1.598* (0.898)	-1.759** (0.897)
Avg. Teacher Exp. (Yrs.)	0.903 (0.794)	1.019 (0.796)	1.672** (0.822)	1.767** (0.814)	1.490* (0.846)	1.592* (0.842)	1.981** (0.826)	2.156*** (0.830)
Teach. Share < Bach.	-86.217*** (20.133)	-77.378*** (19.515)	-112.679*** (20.512)	-100.062*** (19.961)	-105.509*** (18.805)	-95.189*** (18.084)	-69.411*** (17.586)	-63.372*** (17.093)
Teach. Share > Bach.	9.972 (8.571)	9.940 (8.549)	-0.303 (9.770)	0.441 (9.610)	0.607 (9.020)	0.812 (8.893)	3.339 (11.377)	3.229 (11.283)
Teach. Share Fully Cert.	-4.726 (5.028)	-4.439 (4.998)	-7.186 (4.976)	-6.809 (4.954)	-7.957* (4.361)	-8.067* (4.627)	-9.251* (5.386)	-8.830* (5.184)
Sch: Socioecon. & Cult.	33.894*** (2.905)	34.356*** (2.841)	34.619*** (3.601)	35.531*** (3.515)	34.863*** (3.224)	35.275*** (3.042)	31.707*** (2.980)	31.535*** (2.907)
Student-Teacher Ratio	-0.477** (0.199)	-0.441** (0.195)	-0.400* (0.230)	-0.295 (0.221)	-0.566*** (0.216)	-0.464** (0.200)	-0.030 (0.228)	-0.010 (0.221)
Hindered: Lack of Res.	-1.654 (1.791)	-1.550 (1.740)	-2.003 (1.884)	-1.689 (1.828)	-2.094 (2.006)	-1.877 (1.900)	-2.427 (1.783)	-2.327 (1.729)
Hindered: Lack of Staff	-0.289 (1.643)	0.070 (1.633)	-1.480 (2.014)	-1.252 (1.979)	-2.127 (1.758)	-1.882 (1.732)	-0.585 (1.891)	-0.278 (1.843)
Rural	1.428 (3.730)	2.338 (3.529)	-0.885 (3.893)	-0.314 (3.642)	3.567 (3.514)	3.797 (3.294)	-2.942 (3.786)	-1.707 (3.677)
All Male or Female Studs.	1.258 (5.593)	0.803 (5.699)	2.024 (5.001)	2.254 (5.000)	2.046 (5.344)	1.936 (5.316)	1.677 (4.798)	0.654 (4.837)
Stu. Grouped by Ability	-2.707 (2.493)	-2.935 (2.512)	-0.941 (2.660)	-0.864 (2.690)	0.259 (2.509)	0.334 (2.530)	2.230 (2.385)	2.107 (2.407)
Tests Used to Group	-0.142 (3.176)	-0.280 (3.166)	-3.792 (3.702)	-4.362 (3.603)	-2.691 (3.304)	-3.313 (3.187)	-3.843 (2.888)	-3.866 (2.849)
Tests used to Adapt Teaching	1.543 (4.535)	2.015 (4.532)	3.271 (4.245)	2.664 (4.205)	3.706 (4.545)	2.585 (4.334)	3.518 (4.018)	3.977 (4.063)
Hindered: Teach. Behavior	0.001 (1.824)	0.021 (1.834)	1.739 (1.754)	1.696 (1.739)	1.118 (1.657)	1.030 (1.683)	1.779 (1.489)	1.894 (1.474)
Satisfaction of Teachers	-1.136 (6.949)	-4.826 (7.151)	5.815 (9.332)	3.854 (9.186)	5.196 (7.841)	3.439 (7.846)	-0.967 (7.576)	-3.775 (7.859)
Attend Phys. Ed. (Days)	-1.453** (0.658)	-0.851 (0.622)	-2.653*** (0.750)	-2.548*** (0.751)	-2.548*** (0.653)	-2.312*** (0.651)	-2.505*** (0.870)	-2.488*** (0.823)
Constant	256.161*** (36.316)	247.247*** (42.735)	247.753*** (42.851)	247.506*** (52.242)	250.672*** (38.591)	254.582*** (44.733)	308.669*** (40.393)	295.585*** (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

B. 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 used¹², 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 behavior¹³.

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 character 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* (0.021)	-0.040* (0.023)	-0.032 (0.045)	0.092*** (0.027)	0.102*** (0.030)	0.169*** (0.054)
Teacher explains	0.034 (0.025)	0.037 (0.030)	0.031 (0.065)	-0.005 (0.028)	-0.004 (0.033)	-0.016 (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 (0.026)	0.027 (0.029)	0.034 (0.051)	0.039 (0.056)	0.049 (0.058)	0.137 (0.089)
Teacher explains	-0.016 (0.032)	-0.027 (0.038)	-0.036 (0.072)	0.093 (0.071)	0.085 (0.076)	-0.087 (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** (0.019)	0.040* (0.022)	0.075* (0.041)	0.026 (0.031)	0.027 (0.033)	0.035 (0.067)
Teacher explains	-0.025 (0.024)	-0.019 (0.029)	-0.023 (0.058)	-0.023 (0.036)	-0.031 (0.042)	-0.109 (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 (0.017)	-0.032* (0.018)	-0.059* (0.033)		-0.228 (0.164)	-0.360 (0.239)
Teacher explains	0.018 (0.018)	0.014 (0.021)	0.022 (0.040)		0.020 (0.180)	0.033 (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

V. 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.

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.

The results are based on a sample from PISA 2015 that includes approximately 35 thousand students, in nearly 1500 schools, in 14 countries/regions. Teaching practices are aggregated at the school level from teacher surveys. Regression models include a battery of individual, teacher, and school controls, as well as random intercepts by school. We also use instrumental variable methods and a two-stage selection model for robustness.

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, while the prevalence of traditional-vertical methods does not affect outcomes one way or another.

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

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.

Appendix

Table 7. Table 1 including individual coefficients

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

Table 7. Continued

	(1) Life Sat.	(2) Life Sat.	(3) Life Sat.	(4) Life Sat.
Female Teacher Ratio		0.174 (0.169)	0.256 (0.175)	0.236 (0.173)
Average Teacher Age		-0.013 (0.017)	-0.015 (0.017)	-0.015 (0.016)
Avg. Teacher Exp. (Yrs.)		0.020 (0.017)	0.023 (0.017)	0.020 (0.016)
Teach. Share < Bach.		0.489* (0.291)	0.170 (0.284)	0.151 (0.283)
Teach. Share > Bach.		-0.135 (0.176)	-0.007 (0.161)	0.022 (0.163)
Teach. Share Fully Cert.		-0.039 (0.130)	-0.040 (0.127)	-0.029 (0.127)
Sch: Socioecon. & Cult.			-0.197*** (0.068)	-0.204*** (0.067)
Student-Teacher Ratio			-0.004 (0.004)	-0.002 (0.004)
Hindered: Lack of Res.			-0.023 (0.030)	-0.011 (0.029)
Hindered: Lack of Staff			-0.048* (0.029)	-0.028 (0.033)
Rural			0.018 (0.065)	0.004 (0.063)
All Male or Female Studs.			-0.013 (0.091)	0.009 (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*** (0.543)	7.948*** (0.705)	7.757*** (0.718)	6.241*** (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"

REFERENCES

- Algan, Yann, Pierre Cahuc, and Andrei Shleifer. 2013. "Teaching Practices and Social Capital." *American Economic Journal: Applied Economics* 5 (3): 189–210. <https://doi.org/10.1257/app.5.3.189>.
- Arampatzi, Efstratia, Martijn J. Burger, Elena Ianchovichina, Tim Röhricht, and Ruut Veenhoven. 2018. "Unhappy Development: Dissatisfaction with Life on the Eve of the Arab Spring." *Review of Income and Wealth* 64 (S1): S80–113. <https://doi.org/10.1111/roiw.12388>.
- Baum, Christopher F., Arthur Lewbel, Mark E. Schaffer, and Oleksandr Talavera. 2013. "Instrumental Variables Estimation Using Heteroskedasticity-Based Instruments." Stata Users Group. <https://econpapers.repec.org/RePEc:boc:dsug13:05>.
- Berguno, George, Penny Leroux, Katayoun McAinsh, and Sabera Shaikh. 2004. "Children's Experience of Loneliness at School and Its Relation to Bullying and the Quality of Teacher Interventions." *Qualitative Report* 9 (3): 483–99.
- Bietenbeck, Jan. 2014. "Teaching Practices and Cognitive Skills." *Labour Economics* 30: 143–53. <https://doi.org/10.1016/j.labeco.2014.03.002>.
- Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43 (4): 972–1059. <http://www.jstor.org/stable/40057376>.
- Brulé, Gaël, and Ruut Veenhoven. 2014. "Participatory Teaching and Happiness in Developed Nations." *Advances in Applied Sociology* 04 (11): 235–45. <https://doi.org/10.4236/aasoci.2014.411028>.
- Clark, Andrew E., Sarah Fleche, Richard Layard, Nattavudh Powdthavee, and George Ward. 2019. *The Origins of Happiness: The Science of Well-Being over the Life Course*. Princeton: Princeton University Press.
- Clark, Andrew E., Paul Frijters, and Michael A. Shields. 2008. "Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles." *Journal of Economic Literature* 46 (1): 95–144. <https://doi.org/10.1257/jel.46.1.95>.
- Cordero, Jose Manuel, Victor Cristobal, and Maria Gil. n.d. "Teaching Strategies and Their Effect on Student Achievement: A Cross-Country Study Using Data from PISA 2015."
- Damasio, Antonio. 1995. *Descartes' Error: Emotion Reason, and the Human Brain*. New York: Penguin Books.
- Dubin, Jeffrey A., and Daniel L. McFadden. 1984. "An Econometric Analysis of Residential Electric Appliacne Holdings and Consumption." *Econometrica* 52 (2): 345–62.
- Easterlin, Richard A. 2001. "Income and Happiness, towards a Unified Theory." *Economic Journal* 111: 465–84.
- Easterlin, Richard A., and Kelsey J. O'Connor. 2021. "The Easterlin Paradox." In *Handbook of Labor, Human Resources and Population Economics*, edited by Klaus Zimmermann. Springer.
- Echazarra, Alfonso, Daniel Salinas, Ildefonso Mendez, Vanessa Denis, and Giannina Rech. 2016. "How Teachers Teach and Students Learn: Successful Strategies for School." 130. OECD Education Working Papers. <https://doi.org/10.1787/5jm29kpt0xxx-en>.
- Edmondson, Amy. 1999. "Psychological Safety and Learning Behavior in Work Teams Author." *Administrative Science Quarterly* 44 (2): 350–83. <http://www.jstor.org/stable/2666999>.
- Flèche, Sarah. 2017. "Teacher Quality, Test Scores and Non-Cognitive Skills: Evidence from

- Primary School Teachers in the UK.” 1472. *Centre for Economic Performance*. CEP Discussion Paper No. <http://ezproxy.library.uvic.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED583855&site=ehost-live&scope=site>.
- Frijters, Paul, Andrew E. Clark, Christian Krekel, and Richard Layard. 2020. “A Happy Choice: Wellbeing as the Goal of Government.” *Behavioural Public Policy*, 1–40. <https://doi.org/10.1017/bpp.2019.39>.
- Goldhaber, Dan D., and Dominic J. Brewer. 1997. “Why Don’t Schools and Teachers Seem to Matter?: Assessing the Impact of Unobservables on Educational Productivity.” *Journal of Human Resources* 32 (3): 505–23. <https://doi.org/10.2307/146181>.
- Graham, Carol, Andrew Eggers, and Sandip Sukhtankar. 2004. “Does Happiness Pay? An Exploration Based on Panel Data from Russia.” *Journal of Economic Behavior and Organization* 55 (3): 319–42. <https://doi.org/10.1016/j.jebo.2003.09.002>.
- Graham, Carol, Kate Laffan, and Sergio Pinto. 2018. “Well-Being in Metrics and Policy.” *Science* 362 (6412): 287–88. <https://doi.org/10.1126/science.aau5234>.
- Granovetter, Mark S. 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78 (6): 1360–80.
- Güven, Cahit. 2011. “Are Happier People Better Citizens?” *Kyklos* 64 (2): 178–92. <https://doi.org/10.1111/j.1467-6435.2011.00501.x>.
- . 2012. “Reversing the Question: Does Happiness Affect Consumption and Savings Behavior?” *Journal of Economic Psychology* 33 (4): 701–17. <https://doi.org/10.1016/j.joep.2012.01.002>.
- Hanushek, Eric A. 1986. “The Economics of Schooling: Production and Efficiency in Public Schools.” *Journal of Economic Literature* 24 (3): 1141–77. <https://doi.org/http://www.jstor.org/stable/2725865>.
- . 2009. “Teacher Deselection.” In *Creating A New Teaching Profession*, edited by Dan D. Goldhaber and Jane Hannaway, 165–80. Washington DC: Urban Institute Press.
- Heckman, James J., and Tim Kautz. 2013. “Fostering and Measuring Skills: Interventions That Improve Character and Cognition.” 19656. *NBER Working Paper Series*.
- Helliwell, John F. 2019. “Measuring and Using Happiness to Support Public Policies.” 26529. *NBER Working Paper Series*.
- Helliwell, John F., and Lara B. Aknin. 2018. “Expanding the Social Science of Happiness.” *Nature Human Behaviour* 2 (4): 248–52. <https://doi.org/10.1038/s41562-018-0308-5>.
- Jackson, C. Kirabo. 2018. “What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes.” *Journal of Political Economy* 126 (5): 2072–2107. <https://doi.org/10.1086/699018>.
- Kahneman, Daniel, Alan B. Krueger, David Schkade, Norbert Schwarz, and Arthur A. Stone. 2004. “Toward National Well-Being Accounts.” *American Economic Review* 94 (2): 429–34.
- Krause, Annabelle. 2013. “Don’t Worry, Be Happy? Happiness and Reemployment.” *Journal of Economic Behavior and Organization* 96: 1–20. <https://doi.org/10.1016/j.jebo.2013.09.002>.
- Lewbel, Arthur. 2012. “Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models.” *Journal of Business & Economics Statistics* 30 (1): 67–80. <https://doi.org/10.1080/07350015.2012.643126>.
- Lleras, Christy. 2008. “Do Skills and Behaviors in High School Matter? The Contribution of Noncognitive Factors in Explaining Differences in Educational Attainment and Earnings.” *Social Science Research* 37 (3): 888–902. <https://doi.org/10.1016/j.ssresearch.2008.03.004>.

- Luttmer, Erzo F.P. 2005. "Neighbors as Negatives : Relative Earnings and Well-Being." *Quarterly Journal of Economics* 120 (3): 963–1002.
- Mayer, John D., Peter Salovey, and David Caruso. 2000. "Models of Emotional Intelligence." In *Handbook of Intelligence*, edited by Robert J. Sternberg, 396–420. Cambridge: Cambridge University Press.
- Moglie, Marco Le, Letizia Mencarini, and Chiara Rapallini. 2015. "Is It Just a Matter of Personality? On the Role of Subjective Well-Being in Childbearing Behavior." *Journal of Economic Behavior and Organization* 117: 453–75. <https://doi.org/10.1016/j.jebo.2015.07.006>.
- Mohanty, Madhu S. 2009. "Effects of Positive Attitude on Earnings: Evidence from the US Longitudinal Data." *Journal of Socio-Economics* 38 (2): 357–71. <https://doi.org/10.1016/j.socec.2008.07.012>.
- Neve, Jan-Emmanuel De, Ed Diener, Louis Tay, and Cody Xuereb. 2013. "The Objective Benefits of Subjective Well-Being." In *World Happiness Report 2013*, edited by John F. Helliwell, Richard Layard, and Jeffrey Sachs, 56–79. New York: UN Sustainable Development Solutions Network. <http://worldhappiness.report>.
- Neve, Jan-Emmanuel De, and Andrew J. Oswald. 2012. "Estimating the Influence of Life Satisfaction and Positive Affect on Later Income Using Sibling Fixed Effects." *Proceedings of the National Academy of Sciences of the United States of America* 109 (49): 19953–19958. <https://doi.org/10.1073/pnas.1211437109>.
- O'Connor, Kelsey J. 2020a. "Life Satisfaction and Noncognitive Skills: Effects on the Likelihood of Unemployment." *Kyklos* 73: 568–604. <https://doi.org/10.1111/kykl.12226>.
- . 2020b. "The Effect of Immigration on Natives' Well-Being in the European Union." *Journal of Economic Behavior and Organization* 180. <https://doi.org/10.1016/j.jebo.2020.10.006>.
- O'Connor, Kelsey J., and Carol Graham. 2019. "Longer, More Optimistic, Lives: Historic Optimism and Life Expectancy in the United States." *Journal of Economic Behavior and Organization* 168: 374–92. <https://doi.org/10.1016/j.jebo.2019.10.018>.
- OECD. 2009. "PISA Data Analysis Manual: SPSS Second Edition." Paris.
- . 2016. "PISA 2015 Results (Volume 1): Excellence and Equity in Education." *OECD Publishing*. Paris. <https://doi.org/10.1787/9789264266490-en>.
- . 2017. "PISA 2015 Technical Report." *OECD Publishing*. Paris. <https://doi.org/10.1021/op8002129>.
- Opdenakker, Marie Christine, and Jan Van Damme. 2006. "Teacher Characteristics and Teaching Styles as Effectiveness Enhancing Factors of Classroom Practice." *Teaching and Teacher Education* 22 (1): 1–21. <https://doi.org/10.1016/j.tate.2005.07.008>.
- Origo, Federica, and Laura Pagani. 2009. "Flexicurity and Job Satisfaction in Europe: The Importance of Perceived and Actual Job Stability for Well-Being at Work." *Labour Economics* 16 (5): 547–55. <https://doi.org/10.1016/j.labeco.2009.02.003>.
- Oswald, Andrew J., Eugenio Proto, and Daniel Sgroi. 2015. "Happiness and Productivity." *Journal of Labor Economics* 33 (4): 789–822. <https://doi.org/10.1086/298035>.
- Piekałkiewicz, Marcin. 2017. "Why Do Economists Study Happiness?" *The Economic and Labour Relations Review* 28 (3): 361–77. <https://doi.org/10.1177/1035304617717130>.
- Sarracino, Francesco, and Luca Fumarco. 2018. "Assessing the Non-financial Outcomes of Social Enterprises in Luxembourg." *Journal of Business Ethics*. <https://doi.org/10.1007/s10551-018-4086-9>.

- Schor, Juliet B. 2004. *Born to Buy: The Commercialized Child and the New Consumer Culture*. New York: Scribner.
- Soto, Christopher J. 2015. "Is Happiness Good for Your Personality? Concurrent and Prospective Relations of the Big Five With Subjective Well-Being." *Journal of Personality* 83 (1): 45–55. <https://doi.org/10.1111/jopy.12081>.
- Specht, J., B. Egloff, and S. C. Schmukle. 2012. "Examining Mechanisms of Personality Maturation: The Impact of Life Satisfaction on the Development of the Big Five Personality Traits." *Social Psychological and Personality Science* 4 (2): 181–89. <https://doi.org/10.1177/1948550612448197>.
- Tay, Louis, Lauren Kuykendall, and Ed Diener. 2015. "Satisfaction and Happiness – The Bright Side of Quality of Life." In *Global Handbook of Quality of Life: Exploration of Well-Being of Nations and Continents*, edited by Wolfgang Glatzer, Laura Camfield, Valerie Møller, and Mariano Rojas, 839–53. Springer Netherlands. https://doi.org/10.1007/978-94-017-9178-6_39.
- Tella, Rafael Di, Robert J. MacCulloch, and Andrew J. Oswald. 2003. "The Macroeconomics of Happiness." *The Review of Economics and Statistics* 85 (4): 809–27. <https://doi.org/10.1162/003465303772815745>.
- The Global Happiness Council. 2018. *Global Happiness Policy Report 2018*. New York: Sustainable Development Solutions Network. happinesscouncil.org.
- Twenge, Jean M. 2000. "The Age of Anxiety? Birth Cohort Change in Anxiety and Neuroticism, 1952-1993." *Journal of Personality and Social Psychology* 79 (6): 1007–21. <https://doi.org/10.1037/0022-3514.79.6.1007>.
- . 2015. "Time Period and Birth Cohort Differences in Depressive Symptoms in the U.S., 1982–2013." *Social Indicators Research* 121 (2): 437–54. <https://doi.org/10.1007/s11205-014-0647-1>.
- Twenge, Jean M., Brittany Gentile, C. Nathan DeWall, Debbie Ma, Katharine Lacefield, and David R. Schurtz. 2010. "Birth Cohort Increases in Psychopathology among Young Americans, 1938-2007: A Cross-Temporal Meta-Analysis of the MMPI." *Clinical Psychology Review* 30 (2): 145–54. <https://doi.org/10.1016/j.cpr.2009.10.005>.
- Weel, Bas ter. 2008. "The Noncognitive Determinants of Labor Market and Behavioral Outcomes: Introduction to the Symposium." *Journal of Human Resources* 43 (4): 729–37.