Working Paper Series 591
(ISSN 1211-3298)

Do Teaching Practices Impact Socio-Emotional Skills?

Václav Korbel
Michal Paulus

CERGE-EI
Prague, June 2017

ISBN 978-80-7343-398-7 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-427-3 (Národohospodářský ústav AV ČR, v. v. i.)

Do Teaching Practices Impact Socio-Emotional Skills?*

Václav Korbel†
IES FSV UK‡ & CERGE-EI§

Michal Paulus¶
IES FSV UK

June 2017

Abstract

Recent studies emphasize the importance of socio-emotional skills for the success in school as well as for later economic outcomes. However, little is known about how everyday classroom practices impact development of these skills. Using data from the Czech Republic, we show that modern practices such as working in small groups improve these skills. Intrinsic motivation and self-confidence are particularly positively affected. Moreover, modern practices have no adverse effects on test scores. On the other hand, standard practices such as lecturing or requiring memorizing have no impact on socio-emotional skills and test scores. Our results highlight that test score measures do not capture all skills developed in schools and suggest that changing slightly the composition of teaching practices can have a substantial positive impact on socio-emotional skills.

Keywords: Teaching practices, socio-emotional skills, between-subject variation

JEL classification: I21, C23

*We would like to thank Aline Bütikofer, Štěpán Jurajda, Barbara Pertold-Gebicka, Daniel Münich and Katrine Vellesen Løken for valuable comments. We also thank participants of the CSE conference, the Lisbon Workshop on Economics, Statistics and Econometrics of Education and the KU Workshop on Educations Economics for their comments. All errors remaining in this text are our own. The research was financially supported by GA UK grant No. 910794, by the H2020-MSCA-RISE project GEMCLIME-2020 GA No. 681228 and by grant number P402/12/G130 from the Grant Agency of the Czech Republic.

†Email: vaclav.korbel@fsv.cuni.cz

‡Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, Prague, 11000, Czech Republic

§CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 111 21 Prague, Czech Republic

¶Email: michal.paulus@fsv.cuni.cz

1 Introduction

Socio-emotional skills are important determinants for a variety of socioeconomic outcomes. They actuate children's grades, probability of high-school graduation and their future earnings (Cunha and Heckman, 2006, Wigfield et al., 2009). Several studies support also their predictive ability towards risky behavior, health and public safety (Golsteyn et al., 2014, Heckman et al., 2006, Moffitt et al., 2011). Crucially, socio-emotional skills are malleable primarily during childhood and adolescence (for a review see Heckman et al. (2010)) and can therefore be influenced in the educational process (Koch et al., 2015). However, many countries around the world struggle to set up educational policies which encourage engagement, determination, and self-confidence. Interestingly, little is known about the role of teachers and their teaching methods.

In this paper, we examine the impact of teaching practices on socio-emotional skills (intrinsic motivation, extrinsic motivation and self-confidence). This is of high importance since improvement in socio-emotional skills can have wider effects across populations than a singular focus on content knowledge and cognitive abilities measured by IQ, which is more relevant for later occupations with higher complexity (Kautz et al., 2014). We focus specifically on teaching practices because, though it has been known for a long time that teachers are critical to academic achievement, researchers thus far have been unable to identify what determines teacher quality (Rockoff, 2004, Staiger and Rockoff, 2010). Therefore, the focus has recently shifted to what teachers actually do in classes - to their teaching practices.

So far, evidence on the effects of teaching practices is scarce. Researchers usually look at the impact of standard (traditional) practices like lecturing in front of the classroom or memorization of facts and formulas, and modern practices like group projects or making content more applicable for real life on test scores. Evidence suggests that students benefit more from standard practices, but it is far from conclusive. Studies using within-student between-subject methodology, similarly to this paper, find both a positive effect of standard teaching practices (Bietenbeck, 2014, Schwerdt and Wuppermann, 2011) and a null effect of both traditional and modern practices (Klaveren, 2011) on student performance. Lavy (2015) shows, using longitudinal data from Israel, a positive effect of standard teaching practices on students from low socioeconomic backgrounds, but positive effects of modern teaching practices on students from educated families.

However, it could be argued that test scores capture skills taught by standard practices like content knowledge, whereas modern teaching practices promote other skills. Bietenbeck (2014) supports this claim by showing that modern teaching practices improve reasoning skills and application of knowledge (fluid intelligence). Algan et al. (2013), in a cross-country analysis, find that modern teaching practices positively influence social capital. We hypothesize that modern teaching methods promote socio-emotional skills. In case the positive effects on socio-emotional skills are accompanied by a drop in test scores, policies supporting modern practices would be subject to criticism. To investigate this, we also look at the effects of modern and standard teaching practices on test scores.

We use data from the 2007 wave of the Trends in International Mathematics and Science Study (TIMSS). We focus on the Czech Republic, where the use of modern practices is relatively new and not so widespread. Moreover, students have been shown to experience a sharp drop in motivation between grade 4 and grade 8 (Mullis et al., 2012b). The dataset contains test scores and self-reported answers on motivation and self-confidence from five subjects (math, physics, biology, chemistry, earth science). Furthermore, data on teaching practices from student questionnaires allows us to construct class aggregated indices for standard and modern teaching practices. The index can be interpreted as an effective share of lesson taught using standard, modern and other teaching practices in each subject. Our empirical strategy relies on within-student between-subject variation which controls for most of the selection effects. Including a rich set of teacher characteristics and class variables further limits the problem that effects are driven by unobserved teacher characteristics.

We find that modern teaching practices have a significant and sizeable impact on socio-emotional skills, especially on intrinsic motivation and self-confidence. A 10 percentage point rise (6 minutes in a 60 minutes lesson) in modern teaching practices increases intrinsic motivation by 0.24 standard deviation and confidence by 0.11 standard deviation. The effects on social-emotional skills are approximately five to ten times higher than those found on test scores and cognitive abilities (Bietenbeck, 2014, Schwerdt and Wuppermann, 2011). Standard teaching practices have only weak effects on socio-emotional skills and neither standard nor modern teaching practices improve test scores. This evidence supports our initial hypothesis that modern teaching practices improve socio-emotional skills without harming test scores. Interestingly, standard practices reduce socio-emotional outcomes for boys and boost them for girls. The effect is primarily driven by high-achieving girls. On the other hand, even for high-achieving girls, the standard practices coefficients are smaller than for modern practices.

The results appear to be highly robust. Our identification strategy relies on the assumption that selection into teaching methods is not correlated with unobservable teacher characteristics, and several checks support the validity of this assumption. First, the results are invariant to adding a large set of teacher characteristics. Second, following Oster (2013), we try to quantify the unobservable selection needed to generate our findings spuriously under an assumption of proportional selection. The tests suggest very high levels of selection on unobservables causes a spurious results. In a similar vein, analysis of subsamples of classes oriented on math and science subjects provides evidence that sorting of students into teaching practices in a subject-specific way is very unlikely to drive our results.

Our results contribute to the existing literature in three ways. First, to our best knowledge this is the first paper analysing the relationship between teaching practices and socio-emotional skills. The only similar papers that study outcomes other than test scores are Bietenbeck (2014), who exploits various domains of cognitive skills and Algan et al. (2013), who look at social capital. We show that higher emphases on application of modern practices in classrooms may be potentially an easy and scalable intervention to promote socio-emotional skills. It is important to note that the effects may differ for countries with a long tradition and high usage of modern practices. Second, our results suggest that more widespread usage of standard practices may give girls (mainly high-achieving) an advantage in socio-emotional skills which may contribute to gender gaps in educational outcomes (Cornwell et al., 2013, Jacob, 2002). An increase in modern practices might help both girls and boys in development of socio-emotional skills. Third, our results support the evidence that test scores do not capture all skills developed in school.

The remainder of the paper is structured as follows. Section 2 discusses the dataset and descriptive statistics. Section 3 describes our empirical strategy. Results are presented in Section 4 together with robustness checks. Section 5 concludes.

2 Data

We use a representative sample of Czech students from the 2007 wave of TIMSS testing.[1] This is the last wave in which students reported teaching practices. The test was conducted with fourth- and eight-grade students. We limit our analysis to eight-graders because fourth-graders in the Czech Republic are typically taught by a single teacher in all subjects, so that within-student between-subject approach could not be employed.

Students were tested in five subjects - mathematics, physics, biology, chemistry and earth science. The key element of the dataset is the possibility to link students to teachers and their practices in each subject.[2] TIMSS collects data using a two-stage cluster sampling design. First, schools are chosen and then one or two classes are randomly drawn from within the school. Therefore, sampling weights and re-sampling techniques for variance estimation are used throughout the entire analysis.

The impact of teaching practices is measured on motivation, self-confidence and

[1] In total, 59 countries participated in the 2007 wave.

[2] Not all students are taught by five different teachers, 44% of students are taught by 4 different teachers and 13% by three or by two teachers.

test scores. Motivation and self-confidence are derived from student self-reported values ranging from strongly agree (4) to strongly disagree (1). All variables including test scores are standardized to have mean at zero and standard deviation equal to one.[3] We further divide motivation into intrinsic and extrinsic. Intrinsic motivation can be defined as a curiosity and joy of learning. In our analysis we use the question "I enjoy learning *subject*". Extrinsic motivation refers to a situation when a student is motivated to learn a subject because of an external goal. Economic literature studies extrinsic incentives mostly in the form of monetary rewards or grades (Dubey and Geanakoplos, 2010, Gneezy et al., 2011), but we focus on future prospects in education and job career. This is expressed by two questions: "I need to do well in *subject* to get into the university of my choice" and "I need to do well in *subject* to get the job I want". Self-confidence is derived from the question "*Subject* is more difficult for me than for many of my classmates" which is rescaled so that higher value indicates an increase in self-confidence.[4]

The student questionnaires contain questions on teaching practices, examining how often a given practice is used in lesson. We classify three teaching practices as standard (We listen to the teacher giving a lecture-style presentation, We memorize formulas and procedures, We work out problems on our own) and three as modern (We explain our answers, We relate what we are learning in *subject* to our daily lives, We work together in small groups/We work in small groups on an experiment or investigation).[5] Answers were rescaled in the first step so that "never" is equal to 0, "some lessons" to 0.25, "about half the lessons" to 0.5, and "every or almost every lesson" to 1. Rescaled variables represent the effective share of a lesson taught using standard and modern teaching practices at the individual level. Following

[3]Section 4.2 compares linear estimation with a range of alternative fixed-effects ordered logit model estimators. Those models use an original four point scale.

[4]Results are qualitatively and quantitatively similar when using alternative definitions of intrinsic motivation ("I find *subject* boring", "I like *subject*") and self-confidence ("*Subject* is not my strength").

[5]Modern teaching practices in math and science subjects slightly differ from each other - "We work together in small groups" in math and "We work in small groups on an experiment or investigation" in science subjects. However, both practices represent the same activity, namely, working in a small group.

Bietenbeck (2014), we aggregate the indices on a class level, leaving out the value for the observed pupil for each observation.[6]

Table 1 summarizes standard and modern teaching practices across subjects. Standard practices are more widespread among Czech teachers. On average, teachers spend 64% of lesson teaching using standard practices and 44% using modern practices. Looking at the distribution of teaching practices, the majority of teachers are within 15 pp of the mean (10th and 90th percentiles: 52% and 75% (SP); 32% and 55% (MP)) and only 1.2% of teachers use modern practices more often than standard practices. Other activities not fitting into either category (reviewing or doing homework, writing tests or quizzes and using computers) take up on average 28% of lessons. It is important to note that the sum of all activities does not have to add up to one. Imagine a situation when a teacher relates the content to real life (modern teaching practice) while giving a lecture (standard teaching practice), or when a student explains an answer (modern teaching practice) while reviewing homework (other activity). Our claim that standard and modern practices do not crowd-out each other is supported by their positive correlation, which reaches 0.45.

The final dataset consists of 22,633 observations representing 4,528 students in 212 classes taught by 711 teachers. Not to further decrease the number of observations, missing values in all control variables are imputed with 0 and indicators for imputed values are used in all regressions.[7] TIMSS questionnaires contain a rich set of teacher characteristics including their motivation and professional development. Class characteristics are matched from school questionnaires. Means and standard deviations are reported in Table 2. The age composition and share of female teachers significantly differ across subjects. Biology teachers participate the most in professional development courses, while earth sciences teachers participate least often. Importantly, teaching time is significantly higher for mathematics, which may influence both motivation and test scores of students (Joyce et al., 2015).

[6]Results are robust when indices are created from simple means.

[7]The original dataset of 24,225 observations is restricted for missing values in all outcome variables and teaching practices containing 93 % of observations. Results are robust for dropping missing values in all control variables.

To control for confounding factors, we include the variables presented in Table 2 in all regressions.

3 Empirical strategy

A natural way to estimate the impact of teaching practices on other outcomes would be to use the standard education production function and to regress the variables of our interest on school characteristics, teacher characteristics and student characteristics. This approach would, however, neglect selection problems common in schools. First, if teaching practices are determined based on an unobserved school rule or teacher characteristic, then our estimates would be biased. For example, teachers who prefer modern teaching practices can self-select into schools with emphases on modern style of teaching. In the same vein, teachers preferring modern practices can be assigned within a school to more motivated or able classes. Second, students may choose schools or classes based on the teaching style used.

To avoid problems with selection, we use a within-student between-subject approach following Aslam and Kingdon (2011). Comparing socio-emotional skills or test scores across subjects eliminates any difference due to between school or between class differences.[8] Nevertheless, we have to make an assumption that school and student characteristics influence our outcome variables similarly across subjects. Our student fixed-effects model can be written in the following way:

$$Y_{ijt} = \alpha_j + \beta_1 STP_{ijt} + \beta_2 MTP_{ijt} + \gamma T_{jt} + \delta_i + \mu_{ijs}. \tag{1}$$

The outcome variable (motivation, self-confidence and test score) Y_{ijt} of student i in subject j taught by teacher t is regressed on standard and modern teaching practices (STP_{ijt}, MTP_{ijt}) and on a vector of teacher characteristics T_{jt}. δ_i stands for the student fixed-effects and μ_{ijt} is the error term.

[8]In our sample, almost all classes (99%) remain the same across subjects. The fact that Czech children are exposed in all classes to the same peers alleviates the chance that students benefit differently when exposed to better or worse peers (Hoxby, 2000).

The within-student between-subject approach has two caveats. First, there could still be unobserved teacher characteristics influencing both the outcome variable and selection of teaching practices. For example, more motivated or able teachers may use modern practices more frequently. If unobserved teacher characteristics promote the outcome variables of student in another way, then our estimates would be biased. In the analysis, we control for a rich set of teacher characteristics including proxies for motivation and effort which should minimize the problem. Moreover, we calculate the selection on unobservables in Section 4.2.2 borrowing a procedure developed by Oster (2013).

The second concern stems from potentially subject-specific selection of students to teaching practices. In other words, students may sort into schools that emphasize a certain types of teaching practices in some of their subjects. Even though subject-specific sorting is not common in the Czech Republic,[9] some schools focus on specific subjects (e.g. languages or math). This could potentially be related to the choice of teaching practices. Section 4.2.3 investigates the robustness of results for classes with and without special subject focus.

4 Results

We estimate the effects of standard (SP) and modern teaching practice (MP) indices on several outcome variables: test scores, intrinsic motivation, two types of extrinsic motivation and self-confidence. The results of our main specification are summarized in Table 3. All estimations are based on the student fixed-effect model and adjusted for the complex sampling design.

The estimated coefficients of teaching practices are interpreted in the following way. A X percentage point increase in a teaching variable (time spent on selected practices) increases the outcome variable by $X * \beta$ of a standard deviation. For

[9] In grades 6 and 8, students sort into grammar schools. Grammar schools are focused on preparing students to enter a university. Students (approximately 8% of our sample) typically sort there based on their skills and motivation through admission exams. However, the selection is not subject-specific.

example, a 10 percentage point increase in modern practices accounts for a 0.24 increase of standard deviation in intrinsic motivation. A 10 percentage point change can also be expressed as 6 minutes in a 60-minute lesson. The resulting effect holds when standard practices are constant and vice versa. The change in time spent on modern or standard practices are at the expense of other teaching practices such as writing tests, reviewing homework or classroom management.[10]

In general, the results reveal a strong positive impact of modern practices on socio-emotional skills. Importantly, the coefficients of the modern teaching practices are, in most cases, significantly higher than those of the standard practices (Table 3). The highest impact is found for the intrinsic motivation, at 0.24 of SD (column 2). For extrinsic motivation (column 3-4), we observe a positive influence of both practices, regardless whether extrinsic motivation is related to future studies or job. In case of the first type of extrinsic motivation, the hypothesis that SP and MP coefficients are equal cannot be rejected (p-value=0.988). Their effect is about 0.04 SD, with an increase of a teaching practice type by 10 percentage points. The coefficient on the students' motivation to be accepted to their desired university is slightly higher than in the case of their future job prospects (effect of 0.07 SD compared to 0.04 SD). Interestingly, additional time devoted to modern teaching practices strengthens the self-confidence of students, while standard practices have the opposite effect (column 5).[11,12]

In addition to the effects of modern practices on socio-emotional skills, we are also interested in their influence on test scores, to check for potential adverse effects.

[10]It could be argued that modern and standard practices are substitutes and so an increase in one decreases the other. Even though a positive correlation between both practices (0.45) suggests the opposite, we estimated equation 1 separately for each variable. Results are reported in Table A4, Panel A and B. The coefficients of the modern practice index remain almost identical, but the coefficients of the standard practice index increase for all socio-emotional skills. This indicates that both variables are interrelated and omitting one from the estimated equation would lead to bias in the coefficients.

[11]Table A3 estimates heterogeneity in effects across subjects. Math is the only subject where gains from the standard practices are significantly higher than in other subjects and significantly lower from the modern practices for motivation. Coefficients of other subjects are in accordance with the overall findings. It suggests that the modern practices are especially important for socio-emotional skills in science subjects.

[12]The results are identical, with clustering on the school level.

Our estimation confirms no negative effects of MP on test-scores, as the coefficients of both practices are insignificant (column 1).[13] Our findings are in line with Klaveren (2011), who found no statistically significant relationship for students in the Netherlands, and in contradiction to Schwerdt and Wuppermann (2011) and Bietenbeck (2014), who estimated positive effects of standard practices for US students. Positive effects of both practices on students in Israel was found by Lavy (2015). Mixed evidence hinders generalisation for policy makers and more international evidence is needed to explore what causes differences across countries.[14,15]

Surprisingly, teacher characteristics have no or very little impact on test scores and socio-emotional skills. The first few years of experience positively affect intrinsic motivation, but no other outcome variables. Teachers with a university diploma motivate students less than those without a diploma. However, a majority of teachers (95%) hold a diploma, and the effect may be driven by a few teachers without a diploma who compensate for it by developing other characteristics such as motivation or effort. Looking at all variables, teaching practices are constantly the most important for socio-emotional skills. What is even more striking, adding a set of controls barely shifts $R-squared$ in all regressions (change below 0.01) but basically no movement in the explanatory power is in line with previous studies using within-student between-subject strategy (Bietenbeck, 2014, Schwerdt and Wuppermann, 2011).

[13]Table A2 documents the relationship between teaching practices, socio-emotional skills and test scores. Not surprisingly, higher test scores are associated with higher socio-emotional skills. The size of the coefficients for teaching practices remains almost identical when test scores are added into the socio-emotional skills regressions. It seems implausible that the effects of teaching practices are driven via effects on test scores. However, it is impossible to clearly disentangle the relationship among aforementioned variables with the dataset at hand.

[14]Table A4, Panel C investigates non-linearities in the relationship between outcome variables and teaching practices. Squared terms of teaching practices in the regressions turned out to be insignificant. We therefore cannot claim that the relationship is concave. Similar findings were obtained for other functional forms. Results are available upon request.

[15]A potential issue with measurement error in teaching practice indices is investigated in Table A5. We create a variable called *Consensus* which measures agreement on teaching practices within a classroom. It is constructed as a class average of the difference between individual answers and the mean class value of given teaching practices in a given subject. The final variable is an average across teaching practices and subjects ($mean = 0.17$). Even though lower consensus on teaching practices is correlated with lower value of socio-emotional skills, the coefficients are insignificant and do not change the value of the TP coefficients.

4.1 Gender differences

Girls have begun to outperform boys in many countries both in math and science (Mullis et al., 2012a,b). Recent studies try to understand why boys tend to be more often lower achievers, more often drop out of school, and have more behavioral problems (Bertrand and Pan, 2013, OECD, 2015). Some authors argue that differences in socio-emotional skills explain a large share of the gap (Cornwell et al., 2013, Jacob, 2002). Education literature supports this claim showing that boy's problems are correlated with low intrinsic motivation and less interest in school (Gorard et al., 1999, Houtte, 2004).

To explore these questions, we split the dataset by gender and performance (Table 4). The results show that modern teaching practices have significant and positive effects for both genders, but girls benefit significantly more from standard practices than boys (Panel A, B). In respect to intrinsic motivation and self-confidence, boys are even harmed by the usage of standard practices (Panel B). For girls, the coefficients of the standard practice indices are smaller or the same as the coefficients of the modern practice indices. When we look at girls divided by the mean test score, the effects are driven solely by high-achieving girls. When we divide boys in the same way we see no difference. This suggests that high-achieving girls receive a socio-emotional "boost" from standard practices. Therefore, increases in the use of modern teaching practices may help boys without hurting girls. On the other hand, this result may be specific for science subjects and math and may not hold in other subjects.

4.2 Robustness checks

This subsection tests the sensitivity of our results: first, to alternative definitions of teaching practices, second, to sorting of teachers to teaching practices on unobservables and third, to subject-specific selection of students. These are the main issues with a causal interpretation of our result. Lastly, we compare our results from linear and ordered-logit models for socio-emotional outcomes.

4.2.1 Alternative definitions of teaching practices

The impacts of standard and modern teaching practices on other outcomes may hinge on the exact composition of the teaching indices. Since students are asked about 16 practices, we further explore the robustness of our results by estimating our main specifications with two alternative teaching practice indices. They are presented in Table A1. The first adds one standard and one modern practice. Second, the teaching practice "work alone" could potentially be considered both standard and modern, depending on the context. Therefore, the second definition replaces it with another practice, "writing equations and functions" for math and "reading textbooks" for science. Panels A1 and A2 of Table 5 corroborate our original results with no effects on test scores and positive effects of modern practices on socio-emotional skills.

4.2.2 Sorting into teaching practices on unobservable characteristics

We have shown that the results are robust to inclusion of controls on teacher and class characteristics. We use controls which are typically discussed in the literature. However, there may still be unobserved teacher characteristics which could drive both selection into modern teaching practices and the development socio-emotional skills of students. To asses potential bias, we follow a procedure developed by Oster (2013) who argues that the robustness of coefficients upon the addition of controls hinges on the explanatory power of the controls. Using an assumption that the selection on observables and unobservables is proportional, we can test if unobservables could explain our results.[16] The derived coefficient δ can be interpreted as the degree of correlated unobservable selection necessary to cancel out the estimates. In other words, $delta > 1$ means that selection on unobservables would have to be stronger than selection on observables to explain the results.

[16] Oster (2013) builds on the previous work of Altonji et al. (2005). The assumption on proportionality of selection can be written as $\frac{Cov(X,W_2)}{Var(W_2)} = \delta \frac{Cov(X,W_1)}{Var(W_1)}$, where X is the treatment, W_1 is a vector of observables and W_2 of unobservables. Then, δ represents the strength of selection on unobservables to observables.

The most conservative assumption for this exercise is that all remaining variation could be explained by unobservable selection. However, as we have discussed in the previous section, including controls increases R-squared only marginally (less than 0.01, the value of $R-squared$ ranges from 0.4 to 0.5 for all socio-emotional outcomes). Thus, it seems very unlikely that the rest of variation could be explained by unobserved teacher characteristics, and some may be due to measurement error in the outcome variables.[17] We derive δ first unrestricted ($R-squared$ max $= 1$) and then with the more realistic but still conservative assumption of $R-squared = 0.7$. Obviously, we calculate the value of δ only for significant coefficients of standard and modern teaching practices.

The results are reported at the bottom of Table 3. Under the stricter assumption of $R-squared$ max $= 1$, values of δ start from 0.273 and only modern practices on self-confidence exceed one (2.049). Under the more realistic assumption of $R-squared$ max $= 0.7$, all values of δ are higher than one except for standard practices in extrinsic motivation - uni (0.731) and are close to one for modern practices for the same outcome variable (1.041). This means that selection on unobservables would have to be unrealistically high for self-confidence (4.334) and very high for instrinsic motivation and extrinsic motivation - job (around 1.2) to explain the coefficients. Results on extrinsic motivation - uni turn out to be slightly less robust. Overall, the selection on unobservables should not be a serious issue for interpretation of our results.

4.2.3 Subject-specific sorting

The within-student between-subject approach relies on the assumption that students do not sort to teaching practices in a subject-specific way. This would be violated if students sorted into schools or classes focusing on a specific subject/s. Moreover, emphasis on a subject/s would have to be related to teaching practices. As we mentioned in section 3, such sorting is not common in the Czech Republic.

[17]Bear in mind that motivation and self-confidence are self-reported and measured on a 4 point scale.

Specifically, 8% of classes in our sample specialize in a technical subject (math, ICT, or a science subject). When we divide classes into specialized and non-specialized, the results are qualitatively similar, but the lower number of observations leads to fewer significant variables (Table 5, panel B1). The only important distinction is a positive effect of modern practices on test scores.[18] Similarly, the effects are not driven by grammar schools (selective schools for high achieving students) where there is a selection by general ability (Panel B2).

4.2.4 Linear and ordered-logit fixed-effects models

For our analysis, we assumed cardinality of answers in variables for socio-emotional skills. However, economists usually assume that self-reported answers are only ordinarily comparable. So far, evidence is not conclusive on what would constitute a better approach. Some studies, mainly those exploring happiness measures, argue that both approaches yield similar results (Dickerson et al., 2014, Ferrer-i Carbonell and Frijters, 2004, Frey and Stutzer, 2000). However, Baetschmann et al. (2015) showed in a recent study that the fixed-effects (FE) ordered model used in Ferrer-i Carbonell and Frijters (2004) is biased. Thus, a comparison with cardinal results is not reliable.

In TIMSS questionnaires, answers range from "agree a lot" to "disagree a lot" which is, strictly speaking, rather an ordinal measure. On the other hand, FE ordered models have one important disadvantage; that coefficients cannot be interpreted except for their sign and significance level. For its simplicity of interpretation, we chose the FE linear approach for our analysis. In this section we reanalyse all results with an ordered FE approach. The FE linear model is compared with four FE ordered-logit models - Blow-up and Cluster estimator and three estimators with individual-specific endogenous cut-offs (the mean, median, minimum Hessian (FF)). The theoretical background and description of the estimators used is presented in Appendix B.

[18]The results also hold when we look at all classes specializing in any course (languages, sports, music and arts). Results available upon request.

The results reported in Table 6 for intrinsic motivation show only minor differences across models.[19] Column 1 reports coefficients from an OLS estimator with fixed-effects using a four point scale. The coefficients of the linear model can be interpreted as marginal effects. Even though ordered models cannot be compared in terms of magnitude (Column 2-5), the coefficients have the same signs and resemble each other in the majority of significance levels.[20] Moreover, the relative magnitudes of coefficients across models are also roughly the same. This is especially important for coefficients in the variables of our interest, teaching practices.

5 Conclusion

Taken together, we show that modern teaching practices have a significant impact on socio-emotional skills in technical subjects, and no adverse effects on test scores. Standard practices tend to demotivate boys and lead to lower self-confidence. On the other hand, high-achieving girls benefit from standard practices. Importantly, the positive effects of modern practices are higher than those of standard practices for both genders.

We believe that our findings can shed light on the current public debate about teaching methods and their effectiveness for children's development. In spite of a higher focus on socio-emotional skills among researchers and policy makers in recent years, we still know very little about how teaching methods used every day in classrooms affect students. This is quite surprising since changes in composition of lessons could be very cheap and scalable. Moreover, recent studies (Algan et al., 2013, Bietenbeck, 2014) suggest that modern practices such as working in small groups account for increases in skills not captured in test scores such as social capital and application of gained knowledge.

[19] Estimates for other outcomes variables are not included but show similar patterns. Dummy variables for teacher motivation, indicator variables for imputed values and dummy variables for each subjects are not reported in Table 6. However, results are consistent with the results presented in this section.

[20] Complicated procedures for executing FE ordered models prevent us from using exactly the same clustering procedure appropriate for TIMSS dataset. This could be the reason for a few of the dissimilarities in significance levels across models.

Our results advocate for wider inclusion of modern teaching practices in lessons. However, we have to emphasize that our results should be interpreted with caution. First, we analyzed the impact of teaching practices on socio-emotional skills in a country with both very low levels of socio-emotional skills and modern teaching practices. Therefore, marginal effects in this study could be higher than in other countries. For example, modern practices may be perceived by Czech students as new and inspiring, while they are considered more standard and dull for students in the US, where usage of modern practices has a much longer tradition. Second, our results are based on the analysis of science subjects and math, and the relationship may differ for other subjects. Therefore, more research is needed to confirm our findings for other countries and to specifically disentangle elements of modern practices which are useful for supporting the development of socio-emotional skills.

References

Algan, Y., Cahuc, P., and Shleifer, A. (2013). Teaching Practices and Social Capital. *American Economic Journal: Applied Economics*, 5(3):189–210.

Altonji, J., Elder, T., and Taber, C. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113(1):151–184.

Aslam, M. and Kingdon, G. (2011). What can teachers do to raise pupil achievement? *Economics of Education Review*, 30(3):559–574.

Baetschmann, G., Staub, K. E., and Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(3):685–703.

Bertrand, M. and Pan, J. (2013). The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. *American Economic Journal: Applied Economics*, 5(1):32–64.

Bietenbeck, J. (2014). Teaching practices and cognitive skills. *Labour Economics*, 30:143–153.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47(1):1980.

Cornwell, C., Mustard, D. B., Parys, J. V., Cornwell, C., Mustard, D. B., and Parys, J. V. (2013). Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments : Evidence from Primary School Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments Evidence from Primary School. *Journal of Human Resources*, 48(1):236–264.

Cunha, F. and Heckman, J. J. (2006). Interpreting the Evidence on Life Cycle Skill Formation. *Handbook of the Economics of Education*, 1:697–812.

Dickerson, A., Hole, A. R., and Munford, L. A. (2014). The relationship between well-being and commuting revisited: Does the choice of methodology matter? *Regional Science and Urban Economics*, 49:321–329.

Dubey, P. and Geanakoplos, J. (2010). Grading exams: 100,99,98,... or A,B,C? *Games and Economic Behavior*, 69(1):72–94.

Ferrer-i Carbonell, A. and Frijters, P. (2004). How Important is Methodology for the Estimates of the Determinants of Happiness? *The Economic Journal*, 114(497):641–659.

Frey, B. S. and Stutzer, A. (2000). Happiness , Economy and Institutions. *The Economic Journal*, 110(466):918–938.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, 25(4):191–210.

Golsteyn, B. H. H., Grönqvist, H., and Lindahl, L. (2014). Adolescent time preferences predict lifetime outcomes. *Economic Journal*, 124(580):F739–F761.

Gorard, S., Rees, G., and Salisbury, J. (1999). Reappraising the Apparent Underachievement of Boys at School. *Gender and Education*, 11(4):441–454.

Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., and Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics*, 1(1):1–46.

Heckman, J., Stixrud, J., and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3):411–482.

Houtte, M. V. (2004). Why boys achieve less at school than girls: the difference between boys' and girls' academic culture. *Educational Studies*, 30(2):159–173.

Hoxby, C. M. H. U. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. *National Bureau of Economic Research*, No. w7867.

Jacob, B. A. (2002). Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education review*, 21(6):589–598.

Joyce, T., Crockett, S., Jaeger, D. A., Altindag, O., and O'Connell, S. D. (2015). Does classroom time matter? *Economics of Education Review*, 46:64–77.

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., and Borghans, L. (2014). Fostering and Measuring Skills. *National Bureau of Economic Research,*, No. w19656.

Klaveren, C. V. (2011). Lecturing style teaching and student performance. *Economics of Education Review*, 30(4):729–739.

Koch, A., Nafziger, J., and Nielsen, H. S. (2015). Behavioral economics of education. *Journal of Economic Behavior & Organization*, 115:3–17.

Lavy, V. (2015). What makes an effective teacher? Quasi-experimental evidence. *CESifo Economic Studies*.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., and Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7):2693–8.

Mullis, I., Martin, M., Foy, P., and Arora, A. (2012a). TIMSS International Results in Science. Technical report.

Mullis, I. V. S., Martin, M. O., Foy, P., and Arora, A. (2012b). TIMSS 2011 International Results in Mathematics. Technical report, Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Neyman, J. and Scott, E. L. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, 16(1):1–32.

OECD (2015). *The ABC of Gender Equality in Education*, volume PISA.

Oster, E. (2013). Unobservable Selection and Coefficient Stability: Theory and Validation.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2):247–252.

Schwerdt, G. and Wuppermann, A. C. (2011). Is traditional teaching really all that bad? A within- student between-subject approach. *Economics of Education Review*, 30(2):365–379.

Staiger, D. O. and Rockoff, J. E. (2010). Searching for Effective Teachers with Imperfect Information. *Journal of Economic Perspectives*, 24(3):97–118.

Wigfield, A., Wentzel, K. R., and Miele, D. (2009). *Handbook of Motivation at School*. Routledge.

Table 1: Descriptive statistics
Teaching practices

	Mean	SD
Standard teaching practices		
Total average	0.64	(0.07)
Mathematics	0.67	(0.07)
Physics	0.65	(0.08)
Biology	0.61	(0.09)
Chemistry	0.68	(0.07)
Earth science	0.60	(0.08)
Difference (p-value)	0.00	
Modern teaching practices		
Total average	0.44	(0.06)
Mathematics	0.39	(0.08)
Physics	0.48	(0.08)
Biology	0.45	(0.08)
Chemistry	0.49	(0.08)
Earth science	0.39	(0.07)
Difference (p-value)	0.00	
Other activities	0.28	(0.08)

Notes: Class-aggregated indices leave out each student's own observation. The index expresses the share of a lesson devoted to a particular teaching practice. Student weights are used. Chi-square multivariate test is used to compare means. Standard deviation in parentheses.

Table 2: Descriptive statistics
Teacher and class characteristics

Variable	Mathematics		Physics		Biology		Chemistry		Earth science		Difference
Teacher characteristics	207 teachers		206 teachers		208 teachers		203 teachers		202 teachers		p-value
Age:											
Under 25	0.01	(0.09)	0.00	(0.09)	0.00	(0.06)	0.02	(0.16)	0.01	(0.13)	0.30
25 to 29	0.12	(0.32)	0.09	(0.29)	0.12	(0.33)	0.07	(0.26)	0.21	(0.41)	0.00
30 to 39	0.21	(0.40)	0.21	(0.41)	0.12	(0.32)	0.24	(0.42)	0.19	(0.39)	0.00
40 to 49	0.33	(0.47)	0.32	(0.46)	0.23	(0.42)	0.23	(0.42)	0.23	(0.42)	0.02
50 to 59	0.26	(0.43)	0.20	(0.40)	0.38	(0.48)	0.31	(0.46)	0.23	(0.42)	0.00
60 or older	0.07	(0.26)	0.15	(0.35)	0.12	(0.33)	0.11	(0.31)	0.09	(0.29)	0.15
Female	0.77	(0.41)	0.56	(0.49)	0.83	(0.37)	0.85	(0.35)	0.61	(0.48)	0.00
Experience:											
0-2 years	0.03	(0.19)	0.06	(0.23)	0.07	(0.26)	0.08	(0.28)	0.12	(0.32)	0.02
3-5 years	0.10	(0.31)	0.09	(0.29)	0.06	(0.24)	0.06	(0.25)	0.11	(0.32)	0.25
6 or more years	0.85	(0.35)	0.84	(0.36)	0.86	(0.34)	0.84	(0.36)	0.76	(0.42)	0.11
University degree	0.96	(0.18)	0.93	(0.24)	0.96	(0.19)	0.96	(0.19)	0.92	(0.26)	0.24
Motivation (1-4 scale)											
Discuss concepts with others	1.81	(0.63)	1.89	(0.74)	1.77	(0.73)	1.83	(0.73)	1.75	(0.75)	0.39
Preparation of materials	1.99	(0.82)	2.10	(0.99)	1.95	(0.85)	1.96	(0.83)	1.95	(0.85)	0.45
Visits in other classes	1.13	(0.38)	1.15	(0.35)	1.12	(0.32)	1.11	(0.32)	1.14	(0.39)	0.80
Informal visits	1.08	(0.29)	1.16	(0.38)	1.12	(0.34)	1.13	(0.37)	1.10	(0.33)	0.15
Further development											
Subject content course	0.52	(0.50)	0.59	(0.49)	0.63	(0.48)	0.72	(0.44)	0.51	(0.50)	0.00
Pedagogy course	0.51	(0.50)	0.54	(0.49)	0.39	(0.49)	0.54	(0.49)	0.32	(0.46)	0.00
Curriculum improvement course	0.39	(0.49)	0.31	(0.46)	0.30	(0.45)	0.40	(0.49)	0.22	(0.41)	0.00
Subject related to IT	0.54	(0.49)	0.61	(0.48)	0.58	(0.49)	0.57	(0.49)	0.52	(0.50)	0.40
Critical thinking course	0.29	(0.45)	0.37	(0.48)	0.30	(0.45)	0.30	(0.46)	0.30	(0.46)	0.43
Student evaluation course	0.22	(0.41)	0.23	(0.42)	0.23	(0.42)	0.22	(0.41)	0.20	(0.40)	0.94
Class characteristics											
Class size	24.29	(4.20)	24.36	(4.03)	24.53	(4.13)	24.45	(4.12)	24.40	(4.00)	0.98
Minutes in class	203.12	(30.03)	116.60	(51.60)	91.18	(24.88)	100.41	(30.45)	92.98	(40.95)	0.00

Notes: Teacher probability weights are used to calculate all descriptive statistics. Variables on motivation range from 1-4 (1 - never or almost never, 2 - 2 or 3 times per month, 3 - 1-3 times a week, 4 - daily or almost daily), other variables from 0 to 1. Chi-square multivariate test is used to compare means. Standard deviation in parentheses.

Table 3: Teaching practices, socio-emotional skills and test scores

	Test score	Intrinsic motivation	Extrinsic motivation		Self-confidence
			University	Job	
	(1)	(2)	(3)	(4)	(5)
Standard practices	0.108	0.202	0.400***	0.200	-0.393***
	(0.161)	(0.223)	(0.120)	(0.128)	(0.146)
Modern practices	0.175	2.460***	0.396**	0.690***	1.140***
	(0.140)	(0.214)	(0.156)	(0.132)	(0.170)
Teacher female	-0.014	-0.034	-0.033	-0.046**	0.018
	(0.018)	(0.044)	(0.020)	(0.021)	(0.031)
Age					
25-29	0.017	0.008	-0.006	0.075	0.066
	(0.078)	-0.117	(0.054)	(0.049)	(0.137)
30-39	-0.023	-0.191	-0.048	0.035	-0.051
	(0.087)	(0.121)	(0.066)	(0.054)	(0.139)
40-49	-0.039	-0.186	-0.071	0.027	-0.059
	(0.093)	(0.129)	(0.069)	(0.055)	(0.142)
50-59	-0.033	-0.281**	-0.085	0.028	-0.108
	(0.094)	(0.121)	(0.065)	(0.059)	(0.136)
>60	-0.008	-0.231*	-0.058	0.026	-0.055
	(0.101)	(0.132)	(0.069)	(0.058)	(0.147)
Experience					
3-5 years	-0.02	0.151**	-0.049	-0.088**	0.035
	(0.030)	(0.077)	(0.042)	(0.041)	(0.070)
>5 years	0.027	0.227***	0.038	-0.029	0.051
	(0.047)	(0.064)	(0.051)	(0.041)	(0.058)
University diploma	0.043	-0.177***	-0.052	-0.027	-0.028
	(0.028)	(0.065)	(0.049)	(0.034)	(0.059)
Number of min./week	0.0003	0.0004	-0.0001	-0.0003	0.0004
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0002)
Class size	-0.0013	-0.0008	-0.0002	0.001	0.001
	(0.0014)	(0.003)	(0.002)	(0.001)	(0.002)
Further development					
Subject content course	-0.030*	-0.026	0.038**	0.008	-0.03
	(0.016)	(0.033)	(0.019)	(0.019)	(0.026)
Pedagogy course	0.017	0.011	-0.019	0.006	0.01
	(0.017)	(0.032)	(0.020)	(0.017)	(0.020)
Curriculum improvement course	-0.022	-0.04	0.002	0.012	-0.036
	(0.017)	(0.033)	(0.020)	(0.020)	(0.031)
Subject related to IT	-0.011	0.016	0.003	-0.016	-0.005
	(0.016)	(0.033)	(0.019)	(0.017)	(0.026)
Critical thinking course	-0.002	-0.013	-0.006	-0.004	-0.012
	(0.023)	(0.039)	(0.023)	(0.018)	(0.028)
Student evaluation course	0.014	0.064	0.004	-0.019	0.059*
	(0.028)	(0.040)	(0.019)	(0.020)	(0.030)
STP = MTP (p-value)	0.794	0	0.988	0.030	0
Observations	22,633	22,633	22,633	22,633	22,633
R-squared	0.808	0.428	0.521	0.502	0.435
δ ratio (R^2 max $=1$)					
Standard practices			0.273		0.748
Modern practices		0.598	0.389	0.477	2.049
δ ratio (R^2 max $=0.7$)					
Standard practices			0.731		1.595
Modern practices		1.220	1.041	1.196	4.334

Notes: All regressions control for teacher and class characteristics from Table 2. Furthermore, regressions control for subject dummies and imputation indicators. Test score coefficients are estimated from five plausible values. Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Table reports p-values from hypothesis testing if the standard teaching practices coefficient equals the modern teaching practices coefficients.

Table 4: Teaching practices and gender effects

	Test score	Intrinsic motivation	Extrinsic motivation		Self-confidence
			University	Job	
	(1)	(2)	(3)	(4)	(5)
Panel A: Girls					
Standard practices	0.203	0.826***	0.526***	0.373**	-0.081
	(0.205)	(0.302)	(0.148)	(0.188)	(0.183)
Modern practices	0.155	2.196***	0.393**	0.507***	0.910***
	(0.132)	(0.270)	(0.188)	(0.171)	(0.197)
Observations	11,053	11,053	11,053	11,053	11,053
R-squared	0.819	0.424	0.527	0.514	0.442
Panel A1: Low-achieving					
Standard practices		0.432	0.291	0.0735	-0.472*
		(0.364)	(0.270)	(0.316)	(0.284)
Modern practices		2.253***	0.223	0.512**	0.816**
		(0.356)	(0.290)	(0.225)	(0.325)
Panel A2: High-achieving					
Standard practices		1.067***	0.799***	0.829***	0.261
		(0.365)	(0.242)	(0.266)	(0.259)
Modern practices		2.070***	0.637**	0.567**	0.938***
		(0.345)	(0.250)	(0.255)	(0.272)
Panel B: Boys					
Standard practices	0.013	-0.413*	0.257	0.007	-0.726***
	(0.160)	(0.217)	(0.178)	(0.154)	(0.180)
Modern practices	0.178	2.754***	0.394*	0.892***	1.376***
	(0.197)	(0.244)	(0.202)	(0.197)	(0.202)
Observations	11,580	11,580	11,580	11,580	11,580
R-squared	0.803	0.458	0.536	0.520	0.439

Notes: All regressions control for teacher and class characteristics from Table 2. Furthermore, regressions control for subject dummies and imputation indicators. Test score coefficients are estimated from five plausible values. Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Robustness checks

	Test score	Intrinsic motivation	Extrinsic motivation		Self-confidence
			University	Job	
	(1)	(2)	(3)	(4)	(5)
Alternative definitions of teaching practices					
Panel A1: Definition Alt. 1					
Standard practices	0.131	0.128	0.382***	0.117	-0.457***
Modern practices	0.218	2.639***	0.466***	0.740***	1.272***
Observations	22,392	22,392	22,392	22,392	22,392
R-squared	0.809	0.429	0.523	0.504	0.438
Panel A2: Definition Alt. 2					
Standard practices	-0.0177	0.231	0.346**	0.129	-0.451***
Modern practices	0.201	2.469***	0.444***	0.727***	1.125***
Observations	22624	22624	22624	22624	22624
R-squared	0.807	0.428	0.522	0.503	0.435
Course specific teaching practices					
Panel B1: Classes with a focus on technical subjects					
Standard practices	-0.546	-1.261	-0.859	-0.465	-0.664
Modern practices	0.784*	2.617**	0.862	0.307	1.454
Observations	1,803	1,803	1,803	1,803	1,803
R-squared	0.779	0.438	0.527	0.515	0.424
Restricted sample for special classes					
Standard practices	0.0732	0.224	0.386***	0.193	-0.438***
Modern practices	0.149	2.429***	0.365**	0.693***	1.141***
Observations	20,830	20,830	20,830	20,830	20,830
R-squared	0.808	0.429	0.522	0.502	0.437
Panel B2: Grammar schools					
Standard practices	-0.293	-1.504	-0.147	0.249	-2.124**
Modern practices	-0.147	0.819	0.343	0.686	-0.445
Observations	1,967	1,967	1,967	1,967	1,967
R-squared	0.713	0.388	0.420	0.427	0.419
Restricted sample for grammar schools					
Standard practices	0.0963	0.240	0.397***	0.148	-0.362**
Modern practices	0.150	2.512***	0.417**	0.719***	1.186***
Observations	20,666	20,666	20,666	20,666	20,666
R-squared	0.783	0.435	0.531	0.510	0.440

Notes: Panel A uses alternative definitions from Table A1. Panel B1 splits the sample classes with and without technical specialization (math, ICT, science). All regressions control for teacher and class characteristics from Table 2. Furthermore, regressions control for subject dummies and imputation indicators. Test score coefficients are estimated from five plausible values. Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Comparison of linear and order models

Dependent var.: Intrinsic motivation	OLS	BUC	Mean	Median	FF
	(1)	(2)	(3)	(4)	(5)
Standard practices	0.202	0.385	0.202	0.0132	-0.135
	(0.223)	(0.297)	(0.278)	(0.277)	(0.365)
Modern practices	2.460***	5.651***	5.525***	5.387***	6.287***
	(0.214)	(0.299)	(0.279)	(0.277)	(0.382)
Teacher female	-0.034	-0.084*	-0.129***	-0.107**	-0.129**
	(0.044)	(0.047)	(0.043)	(0.044)	(0.058)
Age					
25-29	0.008	0.075	0.213	0.224	0.398*
	-0.117	(0.169)	(0.154)	(0.155)	(0.217)
30-39	-0.191	-0.426**	-0.220	-0.186	-0.028
	(0.121)	(0.184)	(0.167)	(0.168)	(0.233)
40-49	-0.186	-0.387**	-0.190	-0.140	0.002
	(0.129)	(0.187)	(0.169)	(0.170)	(0.237)
50-59	-0.281**	-0.621***	-0.365**	-0.325*	-0.221
	(0.121)	(0.187)	(0.170)	(0.171)	(0.237)
> 60	-0.231*	-0.480**	-0.263	-0.242	-0.182
	(0.132)	(0.199)	(0.179)	(0.179)	(0.247)
Experience					
3-5 years	0.151**	0.370***	0.218**	0.211**	0.195
	(0.077)	(0.0980)	(0.0897)	(0.0904)	(0.122)
> 5 years	0.227***	0.552***	0.428***	0.364***	0.472***
	(0.064)	(0.107)	(0.099)	(0.010)	(0.138)
University diploma	-0.177***	-0.427***	-0.347***	-0.334***	-0.546***
	(0.065)	(0.096)	(0.091)	(0.092)	(0.123)
Number of min./week	0.0004	0.0008*	0.0008*	0.0011***	0.0026***
	(0.0003)	(0.0005)	(0.0004)	(0.0004)	(0.0006)
Class size	-0.0008	-0.0013	0.0007	0.0008	-0.0055
	(0.003)	(0.0040)	(0.0036)	(0.0036)	(0.0049)
Further development					
Subject content course	-0.026	-0.041	-0.022	-0.061	-0.001
	(0.033)	(0.045)	(0.043)	(0.043)	(0.059)
Pedagogy course	0.011	0.018	-0.008	0.011	0.066
	(0.032)	(0.046)	(0.043)	(0.042)	(0.057)
Curriculum improvement course	-0.04	-0.093**	-0.095**	-0.075*	-0.111*
	(0.033)	(0.046)	(0.044)	(0.043)	(0.059)
Subject related to IT	0.016	0.044	0.072*	0.055	0.004
	(0.033)	(0.041)	(0.038)	(0.039)	(0.053)
Critical thinking course	-0.013	-0.038	-0.062	-0.041	-0.049
	(0.039)	(0.049)	(0.046)	(0.045)	(0.063)
Student evaluation course	0.064	0.139***	0.121**	0.122**	0.067
	(0.040)	(0.052)	(0.048)	(0.048)	(0.066)
Observations	22,633	40,073	21,092	21,087	12089
R-squared	0.428				

Notes: The dependent variable is intrinsic motivation in all columns. The dependent variable is on a 4 point scale (Agree a lot, agree a little, disagree a little, disagree a lot). Apart from controls in the table, there are controls for motivation of teachers, subjects dummy variables and imputed indicators. Column 1 is estimated with OLS, others with ordered-logit fixed-effects models (Blow-up and cluster estimator (column 2) and 3 estimators with endogenous cut-off (Mean, median and Hessian)). Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A1: Definitions of teaching practices

Original			
Standard practices	1		We listen to the teacher give a lecture-style presentation
	2		We memorize formulas and procedures
	3		We work problems on our own
Modern practices	1		We explain our answers
	2		We relate what we are learning in *subject* to our daily lives
	3	Math	We work together in small groups
		Science	We work in small groups on an experiment or investigation
Alternative 1			
The fourth practice is added			
Standard practices	4	Math	We write equations and functions to represent relationships
		Science	We use scientific formulas and laws to solve problems
Modern practices	4	Math	We decide on our own procedures for solving complex problems -
		Science	We design or plan an experiment or investigation
Alternative 2			
Standard teaching practice "Work alone" replaced			
Standard practices	3	Math	We write equations and functions to represent relationships
		Science	We read our science textbooks and other resource materials

Appendix

Appendix A - Additional results

Table A2: Interaction of test scores and socio-emotional skills

	Test score	Intrinsic motivation	Extrinsic motivation		Self-confidence
			University	Job	
	(1)	(2)	(3)	(4)	(5)
Standard practices	0.096	0.184	0.391***	0.189	-0.407***
	(0.161)	(0.222)	(0.119)	(0.128)	(0.146)
Modern practices	0.030	2.428***	0.383**	0.672***	1.114***
	(0.137)	(0.214)	(0.156)	(0.132)	(0.171)
Test score		0.179***	0.077**	0.099***	0.145***
		(0.026)	(0.025)	(0.028)	(0.025)
Intrinsic motivation	0.059***				
	(0.008)				
Extrinsic - university	0.017				
	(0.010)				
Extrinsic - job	0.024**				
	(0.010)				
Self-confidence	0.031***				
	(0.008)				
Observations	22,633	22,633	22,633	22,633	22,633
R-squared	0.806	0.434	0.523	0.504	0.439

Notes: Column 1 reports coefficients from five regressions, test scores and one of socio-emotional outcome as a control. The coefficients for teaching practices in column 1 are from regressionwith intrinsic motivation. The coefficients in other regressions are also insignificant and range from 0.103 to 0.127 for SP and from 0.117 to 0.162 for MP. Similarly, R^2 ranges from 0.807 to 0.809. All regressions control for teacher and class characteristics from Table 2. Furthermore, regressions control for subject dummies and imputation indicators. Test score coefficients are estimated from five plausible values. Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Interaction of subjects and socio-emotional skills

	Test score	Intrinsic motivation	Extrinsic motivation		Self-confidence
			University	Job	
	(1)	(2)	(3)	(4)	(5)
Standard practices	0.160	-0.376	0.063	0.062	-0.474**
	(0.212)	(0.261)	(0.194)	(0.204)	(0.213)
Modern practices	0.056	3.016***	0.749***	0.928***	1.112***
	(0.197)	(0.354)	(0.226)	(0.181)	(0.315)
Math x standard	-0.031	1.152***	0.672**	0.299	-0.367
	(0.354)	(0.383)	(0.335)	(0.351)	(0.347)
Physics x standard	-0.023	0.686*	0.502*	0.271	-0.355
	(0.289)	(0.382)	(0.264)	(0.233)	(0.309)
Biology x standard	0.078	0.607	0.110	-0.148	0.08
	(0.290)	(0.413)	(0.266)	(0.265)	(0.349)
Chemistry x standard	-0.310	0.463	0.565**	0.379	0.207
	(0.280)	(0.453)	(0.285)	(0.241)	(0.381)
Math x modern	0.036	-1.423***	-0.655**	-0.289	0.477
	(0.318)	(0.530)	(0.325)	(0.299)	(0.395)
Physics x modern	-0.159	-0.581	-0.520*	-0.305	0.059
	(0.329)	(0.591)	(0.280)	(0.276)	(0.457)
Biology x modern	0.309	-0.208	-0.204	-0.228	-0.199
	(0.330)	(0.543)	(0.349)	(0.354)	(0.436)
Chemistry x modern	0.487*	-0.496	-0.375	-0.393	-0.408
	(0.291)	(0.516)	(0.338)	(0.304)	(0.418)
Observations	22,633	22,633	22,633	22,633	22,633
R-squared	0.808	0.429	0.522	0.502	0.436
Controls	YES	YES	YES	YES	YES

Notes: Earth science is the omitted subject. All regressions control for teacher and class characteristics from Table 2. Furthermore, regressions control for subject dummies and imputation indicators. Test score coefficients are estimated from five plausible values. Clustered standard errors. in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Teaching practices separately and functional forms

	Test score (1)	Intrinsic motivation (2)	Extrinsic motivation: University (3)	Extrinsic motivation: Job (4)	Self-confidence (5)
Panel A					
Standard practices	0.170	1.096***	0.560***	0.456***	-0.0336
	(0.147)	(0.243)	(0.106)	(0.114)	(0.146)
Observations	23,570	23,327	23,464	23,489	23,323
R-squared	0.804	0.405	0.515	0.493	0.425
Panel B					
Modern practices	0.192	2.509***	0.545***	0.748***	-0.983***
	(0.126)	(0.190)	(0.129)	(0.107)	(0.161)
Observations	23,484	23,245	23,387	23,407	23,245
R-squared	0.805	0.423	0.515	0.496	0.429
Panel C					
Standard practices	0.646	-0.701	-0.489	-0.051	-1.357*
	(0.749)	(1.098)	(0.763)	(1.042)	(0.802)
Modern practices	-0.270	2.923***	0.021	0.741	0.225
	(0.527)	(0.989)	(0.533)	(0.601)	(0.812)
Standard practices^2	-0.423	0.719	0.732	0.202	0.810
	(0.606)	(0.908)	(0.606)	(0.772)	(0.654)
Modern practices^2	0.490	-0.505	0.434	-0.053	1.041
	(0.592)	(1.166)	(0.597)	(0.669)	(0.921)
Observations	22,633	22,633	22,633	22,633	22,633
R-squared	0.808	0.428	0.521	0.502	0.435

Notes: Panel A estimated for standard practices only and Panel B for modern practices only. Panel C added to the original specification practices squared. All regressions control for teacher and class characteristics from Table 2. Furthermore, regressions control for subject dummies and imputation indicators. Test score coefficients are estimated from five plausible values. Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Consensus of students on teaching practices

	Test score	Intrinsic motivation	Extrinsic motivation		Self-confidence
			University	Job	
	(1)	(2)	(3)	(4)	(5)
Standard practices	0.108	0.187	0.374***	0.181	-0.387***
	(0.159)	(0.231)	(0.127)	(0.133)	(0.150)
Modern practices	0.175	2.466***	0.407***	0.697***	1.138***
	(0.140)	(0.215)	(0.157)	(0.134)	(0.171)
Consensus	-0.003	-0.415	-0.725*	-0.518	0.147
	(0.273)	(0.636)	(0.421)	(0.315)	(0.598)
Constant	-0.188	-1.083*	-0.323*	-0.417**	-0.192
	(0.261)	(0.626)	(0.180)	(0.165)	(0.571)
Observations	22,633	22,633	22,633	22,633	22,633
R-squared	0.808	0.428	0.522	0.502	0.435
Controls	YES	YES	YES	YES	YES

Notes: All regressions control for teacher and class characteristics from Table 2. Furthermore, regressions control for subject dummies and imputation indicators. Variable consensus is created as a class average of difference individual and mean class answers on teaching practices between which are then averaged across both teaching practices and subjects. Test score coefficients are estimated from five plausible values. Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix B - Fixed-effects ordered logit models

This section defines the fixed-effects ordered logit model and then reviews estimators that have been suggested in the literature based on Baetschmann et al. (2015), Dickerson et al. (2014). The model considers latent variable y_{it}^* for individual i in time t (in our case in 5 subjects) to a vector of observable characteristics x_{it} and unobserved characteristics α_i, ϵ_{it}:

$$y_{it}^* = \beta x_{it}' + \alpha_i + \epsilon_{it}, \qquad i = 1, \ldots, N \quad t = 1, \ldots, T. \tag{B1}$$

α_i is invariant unobserved component which may or may not be correlated with x_{it}. Observed ordered variable y_{it} is related to the latent variable y_{it}^* in the following way:

$$y_{it} = k \quad \text{if} \quad \upsilon_{ik} < y_{it}^* \leq \upsilon_{ik+1}, \quad k = 1, \ldots, K \tag{B2}$$

and the thresholds υ_k are increasing ($\upsilon_{ik} \leq \upsilon_{ik+1} \quad \forall k$) with $\upsilon_{i1} = -\infty$ and $\upsilon_{iK+1} = \infty$. The fixed-effects ordered logit model assumes that ϵ_{it} are IID with logistic cumulative distribution function $\Lambda(\cdot)$ and the probability of observing outcome k for individual i in time (subject) t is

$$Pr(y_{it} = k | x_{it}, \alpha_i) = \Lambda(\upsilon_{ik+1} - \beta x_{it}' - \alpha_i) - \Lambda(\upsilon_{ik} - \beta x_{it}' - \alpha_i) \tag{B3}$$

Baetschmann et al. (2015) discuss two problems with estimation of maximum likelihood from the equation B3. The first is that only $\upsilon_{ik} - \alpha_i = \alpha_{ik}$ can be identified. The second is that under T asymptotics, estimation of α_{ik} is not consistent due to incidental parameters problem (Neyman and Scott, 1948). The bias of $\hat{\beta}$ can be substantial, especially in short panels.

The Blow-up and Cluster (BUC) estimator

Baetschmann et al. (2015) proposed an estimator which combines information from different cut-offs into a single likelihood function, yielding a one-step estimator of β. The model uses all $K - 1$ cutoffs simultaneously and imposes a restriction that $\beta^2 = \ldots = \beta^K$. It is implemented in the following way - every observation is replaced in the sample by $K - 1$ copy of itself and each of the $K - 1$ copies of the individual is dichotomized at a different cut-off point. The expanded sample can be then estimated using approach developed by Chamberlain (1980). Standard errors must be clustered at the individual level since some observations are used multiple times. Baetschmann et al. (2015) show that the BUC estimator is consistent and performs well even in small panels.

The Ferrer-i-Carbonell and Frijters (FF) estimator

Ferrer-i Carbonell and Frijters (2004) suggested an estimator that identifies a single and "optimal" cut-off point for each individual. The optimal cut-off is found by minimizing the Hessian matrix. It is done in practice at a preliminary estimate of $\hat{\beta}$. For comparison, we also include in our analysis a computationally simpler approach which chooses the cut-off at individual mean or median of y_{it}. However, Baetschmann et al. (2015) show that FF estimators are inconsistent, since the procedure chooses the cut-off point endogenously.

Abstrakt

Nedávné studie zdůrazňují význam socioemočních dovedností pro úspěch ve škole, stejně jako pro další životní výsledky. Nicméně zatím není jasné, jak různé metody výuky ovlivňují tyto dovednosti. Na datech z České republiky ukazujeme, že moderní způsob výuky jako práce ve skupinkách zlepšují socioemoční dovednosti. Především vnitřní motivace a sebevědomí jsou pozitivně ovlivněny. Navíc moderní metody nemají negativní dopad na výsledky žáků v testech. Na druhou stranu standardní metody jako frontální výuka nebo důraz na memorování nemají žádný dopad na socioemoční dovednosti a výsledky v testech. Naše výsledky zdůrazňují, že klasické testování žáků nezachycuje všechny dovednosti rozvíjené ve škole, a naznačují, že malá změna ve složení výuky může mít podstatný dopad na socioemoční dovednosti.

Working Paper Series
ISSN 1211-3298
Registration No. (Ministry of Culture): E 19443

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from institutional support RVO 67985998 from Economics Institute of the CAS, v. v. i.

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

Published by
Charles University, Center for Economic Research and Graduate Education (CERGE)
and
Economics Institute of the CAS, v. v. i. (EI)
CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.
Printed by CERGE-EI, Prague
Subscription: CERGE-EI homepage: http://www.cerge-ei.cz

Phone: + 420 224 005 153
Email: office@cerge-ei.cz
Web: http://www.cerge-ei.cz

Editor: Jan Zápal

The paper is available online at http://www.cerge-ei.cz/publications/working_papers/.

ISBN 978-80-7343-398-7 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-427-3 (Národohospodářský ústav AV ČR, v. v. i.)