

# Increasing instruction time in school does increase learning

Simon Calmar Andersen<sup>a,b,1</sup>, Maria Knoth Humlum<sup>c,1</sup>, and Anne Brink Nandrup<sup>c</sup>

<sup>a</sup>Department of Political Science, Aarhus University, 8000 Aarhus, Denmark; <sup>b</sup>TrygFonden's Centre for Child Research, Aarhus University, 8210 Aarhus, Denmark; and <sup>c</sup>Department of Economics and Business Economics, Aarhus University, 8210 Aarhus, Denmark

Edited by Jens Ludwig, University of Chicago, Chicago, IL, and accepted by Editorial Board Member Mary C. Waters May 11, 2016 (received for review August 21, 2015)

**Increasing instruction time in school is a central element in the attempts of many governments to improve student learning, but prior research—mainly based on observational data—disputes the effect of this approach and points out the potential negative effects on student behavior. Based on a large-scale, cluster-randomized trial, we find that increasing instruction time increases student learning and that a general increase in instruction time is at least as efficient as an expert-developed, detailed teaching program that increases instruction with the same amount of time. These findings support the value of increased instruction time.**

education | randomized controlled trial | school performance | school resources

All governments responsible for school systems must consider the amount of instruction time that they should provide. The time that students spend in the classroom varies by a factor of two across the OECD (Organization of Economic Co-operation and Development) countries both in total compulsory instruction time and within specific subjects, such as reading, writing, and literature (1). These international differences have generated sustained debates about whether students benefit from having more instruction or on the contrary, whether governments can cut spending on instruction time without negatively impacting student achievements (2). Increased instruction time has been an element in many educational reforms in the United States, Europe, and Japan (2–4).

However, the existing evidence of the effectiveness of increasing instruction time is deficient. A review of the research before 2009 concludes that there seems to be a neutral to small positive effect of extending school time on achievements; however, most studies are based on weak designs, and the effect remains disputed. Skeptics argue that longer school days generate behavioral problems caused by fatigue and boredom (2). More recently, studies based on observational data found positive effects on student achievements (4–6), and studies of the impact of increased instruction time in combination with other interventions (e.g., more effective teachers, data-driven instruction, ability tracking, or improved pedagogy) also found positive effects (7, 8). A randomized trial conducted in The Netherlands does not find significant effects of increased instruction time (3). Nonetheless, it should be noted that this trial had substantial noncompliance and was based on only seven schools, which seems low-powered (9).

Other than the methodological limitations of the existing evidence, there are two potential explanations for the lack of strong evidence for the effect of increasing instruction time. One explanation relates to the students, and the other relates to the teachers. First, to benefit from more instruction time, students may need to be motivated to sacrifice short-term pleasures to pay attention to the teaching and thereby, achieve long-term gains (10). This exercise, however, requires self-control. Self-control has been shown to be a scarce resource that is exhausted when used. When that happens, it is harder for students to control their thoughts, fix their attention, and manage their emotions, and they may become more aggressive (11). Thus, extending the

school day may be ineffective, because students' self-control is depleted, and they may have more trouble managing their emotions, become more aggressive, become hyperactive, and/or conflict with their classmates. Furthermore, previous research has shown that boys have less self-control capacity than girls (12), and immigrant children and children with low socioeconomic status also tend to have less self-control (13). [This finding does not imply that immigrants have less self-control or lower academic achievement, because they have a different cultural background. Their achievement may be strongly related to the lower socioeconomic status of immigrants on average. Also, there may be significant heterogeneity among immigrants with non-Western backgrounds.] More formally, children can be expected to maximize their learning in school relative to the effort that it requires. Because the cost of effort as well as the relationship between effort and learning may be different for boys and girls as well as immigrants and natives, it is worth examining the effect of increasing instruction time separately for each of these groups, even if the power of the study does not allow strong conclusions based on subgroup analyses.

Second, the effect of increasing school resources is likely to depend on how teachers spend the additional time, which relates to the instructional regime in the school (that is, the set of rules for how to regulate the interplay between assessment and instruction) (14). We compare two opposite instructional regimes. One type has formalized instruction in a teaching program. On the

## Significance

**Across the Organization for Economic Co-operation and Development countries, the instruction time at age 10 y old varies by a factor of two, reflecting, at least in part, differences in beliefs about the value of additional instruction time. Because educational resources are typically not randomly allocated, it has proven to be a major challenge to determine how different educational resources, such as instruction time, affect student learning. We present evidence from a large-scale, randomized trial that increasing instruction time in school increases student learning. Importantly, a regime with no formal requirements on how the extra time is spent is at least as efficient as an increase in instruction time with a detailed teaching program.**

Author contributions: S.C.A. and M.K.H. designed research; S.C.A., M.K.H., and A.B.N. performed research; M.K.H. and A.B.N. analyzed data; and S.C.A., M.K.H., and A.B.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.L. is a guest editor invited by the Editorial Board.

Data deposition: The data used in this study are survey and administrative data hosted by Statistics Denmark. The data cannot be made publicly available due to the sensitive nature of the data which contains individual-level information. Data security policy means that access can be obtained from Aarhus only and currently requires approval by Statistics Denmark and completion of a test on data security policy. M.K.H. will provide assistance for researchers who are interested in getting access to the data.

<sup>1</sup>To whom correspondence may be addressed. Email: sca@ps.au.dk or mhumlum@econ.au.dk.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516686113/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516686113/-DCSupplemental).

one hand, this format may have several advantages. If teachers do not know how to teach effectively or if they are satisfied with some level of student achievements and therefore, not motivated to use the additional time effectively, providing a detailed teaching program for the additional instruction time may help increase the effectiveness of this time. On the other hand, this instructional regime with a high level of formalized instruction leaves less room for assessment of the individual student's responsiveness to the intervention. Therefore, it may not tailor the instruction sufficiently to the needs of the different groups of students. We compare this regime with another type that has no formalized prescriptions for how the instruction should take place. This high discretion treatment leaves more room for individual assessments and thereby, more tailored instruction. This tradeoff between high discretion, allowing frontline bureaucrats to use their expertise, and low discretion, ensuring a more specific policy implementation, is a classic but topical dilemma (15–17). However, there is very little evidence on whether high or low discretion affects policy outcomes (18).

## Results

To (i) improve the methodological quality of the evidence on increasing instruction time, (ii) compare two different instructional regimes on how to regulate the use of additional time, and (iii) compare how they affect different groups of students in terms of reading skills and behavioral problems that may come with depleted self-control, we use a large-scale, cluster-randomized trial involving 90 schools and 1,931 fourth grade students in Denmark. Instruction time in reading, writing, and literature was increased by 3 h (four lessons) weekly over 16 wk, corresponding to a 15% increase in the weekly instruction time (correspondingly reducing the students' spare time). The cost was approximately US \$182 per student. In the first treatment condition, there were no requirements in terms of how teachers should spend the extra time. This instructional regime with high levels of teacher discretion allows teachers to accommodate their teaching to the specific needs of the students in the classroom across a broad range of outcomes. Conversely, a more detailed, expert-developed teaching program may better ensure high-quality teaching. In the second treatment condition, teachers had the same increase in instruction time but were required to follow a detailed program developed by national experts and aimed at improving general language comprehension. We compare the two treatments with a control group continuing with the same instruction time as usual. *SI Methods* has more details about the treatments.

To measure student achievement in reading, we use a national standardized, online, self-scoring adaptive reading test used by all of the schools in the country (19). The test is based on three subscales: language comprehension, decoding, and reading comprehension. The study was designed to test the effect on the combined measure, but we also examine the effects on the three subscales. Different versions of the test are developed for second and fourth grades. We use the results of the second grade test as a baseline test (along with a third grade math test) and the fourth grade test as an outcome measure. To measure behavioral problems, we use the student responses to the Strengths and Difficulties Questionnaire (SDQ), which is based on five subscales (emotional symptoms, conduct problems, peer relationship problems, hyperactivity/inattention, and prosocial behavior), of which the first four can be combined for a total difficulty score (20, 21). The SDQ responses are used as a second outcome (*SI Methods* has details about the assessments). Outcomes were measured at the end of the intervention period (see Fig. S1).

We find no significant differences in means of the baseline covariates between the control group and the treatment group without a teaching program. A formal  $F$  test of the null hypothesis that all baseline covariates are the same for the no teaching program treatment and the control group is not rejected ( $P = 0.71$ ). The teaching program treatment group differs significantly from

the control group on 2 baseline covariates (of 28), and the  $F$  test rejects the null hypothesis that all baseline covariates are the same ( $P < 0.01$ ). *SI Methods* and Table S1 has more details about the baseline balance. Table S2 presents a formal attrition analysis.

The estimated treatment effects on reading are presented in Fig. 1A. Increasing instruction time without a teaching program increases student achievement in reading by 0.15 SD ( $P = 0.02$ ) compared with the control group. The effect of increasing instruction time with a teaching program is small and insignificant. However, the estimated effects on the reading subscales suggest that the teaching program (aimed at improving general language comprehension) has a statistically significant effect on the subscale language comprehension of 0.14 SD ( $P = 0.03$ ), which would be expected, but not on the other two subscales (Table S3). The general increase in instruction time with no teaching program has significantly positive effects on both language comprehension and decoding. The differences between the two treatment groups are not statistically significant. It should be noted, however, that the statistical power of the trial does not allow us to detect minor differences between the treatment groups.

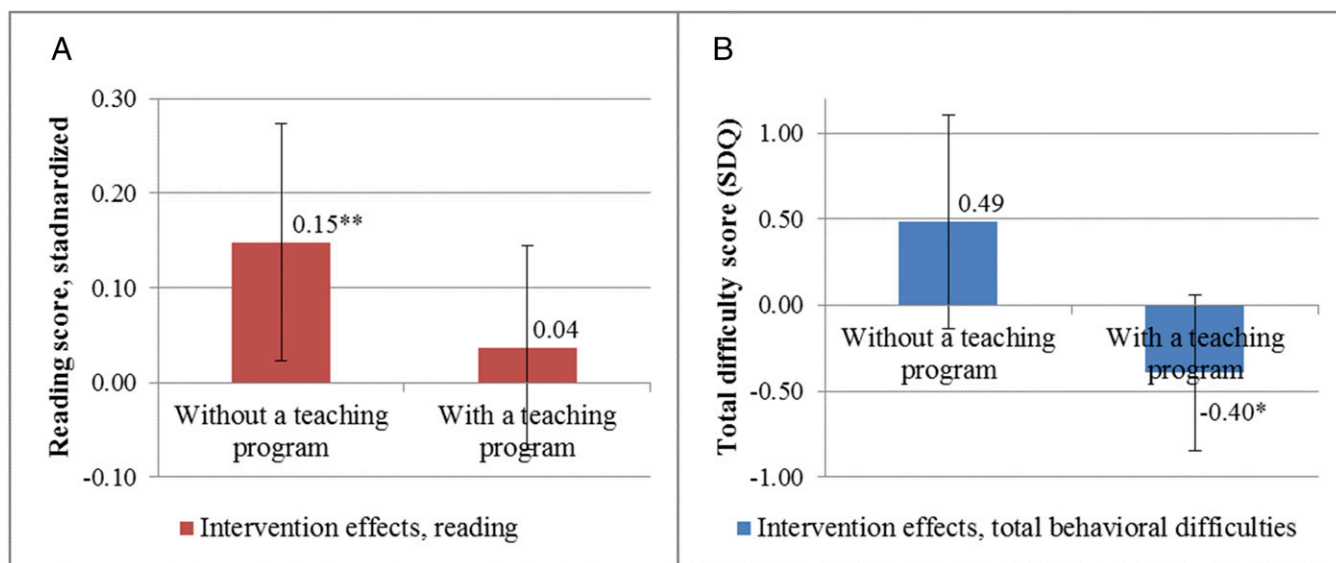
Fig. 1B shows that the teaching program significantly reduces student behavioral problems. Instruction time without a teaching program may increase behavioral problems, but the overall results are insignificant (see Table S4 for details).

The study is not powered for multiple tests of effects on groups, but because of the concerns that increasing instruction time will be less beneficial for students with less capacity for self-control, it is worth examining results for subgroups—although it should be emphasized that these results are exploratory. Subgroup analyses presented in Table S5 suggest that students of non-Western origin do not seem to benefit from any of the two interventions. Increased instruction time without a teaching program may cause increased behavioral difficulties for boys, whereas it has no effect on girls. Conversely, increased instruction time with a teaching program seems to reduce behavioral difficulties for girls.

## Discussion

These results suggest that increasing instruction time does increase average student achievement. An effect of 0.15 SD in reading means that a student at the median of the reading score distribution moves to the 44th percentile. This effect is substantial compared with the relatively cheap intervention. The effect size per US \$1,000 per student would be  $\sim 0.82$  SD. The results suggest that governments cannot reduce instruction time without the risk of adversely affecting student achievement, but this interpretation naturally depends on the generalizability of the results. First, for reasons explained above, the marginal effect of instruction time may be decreasing. OECD figures show that, in Denmark, where this study took place, instruction time in reading, writing, and literature at age 10 y old (which is the median age in this study) is above the average of OECD countries (Fig. S2). The total compulsory instruction time is very close to the OECD average (1). If marginal returns to instruction time are decreasing, the many countries that provide the same or less instruction time than Denmark would have at least as large of effects. The effects for countries providing more instruction time are more uncertain.

Second, the effects of increasing instruction time may depend on the instructional regime (14). Decision-makers face a tradeoff between a high-discretion program allowing teachers to use their assessment of the individual students to tailor instruction vs. a more detailed teaching program that—based on the best available evidence—regulates the instruction more. The latter primarily affected one subscale of reading. The benefits of a detailed, expert-developed program may be outweighed by a narrower focus on a specific learning domain, and more detailed instructional regulation curtails teachers' opportunities to differentiate their teaching to the needs of different students in the classroom. It is worth



**Fig. 1.** Effects of increasing instruction time on (A) student achievement in reading and (B) behavioral difficulties. Stratum indicators and baseline achievement in reading and math are included as controls. Error bars reflect 95% confidence intervals. SEs are corrected for clustering at the classroom level. \* $P < 0.1$ ; \*\* $P < 0.05$ .

emphasizing that implementation survey data suggest that teachers generally had a positive attitude toward the teaching program: 96% of teachers reported that they used the material to some or a large extent, 83% of the teachers found that the teaching program was useful, and 88% believed that it was beneficial for the whole class.

The high-discretion instructional regime with no teaching program had a significant average treatment effect. This finding does not prove that high-discretion programs will be better than a more detailed, evidence-based program. However, it does suggest that there are some benefits of giving teachers good opportunities to differentiate their instruction. Survey data show that 90% of the teachers report that they used (parts of) the increased instruction time for working more with existing materials, and at the same time, 90% used new materials, which supports the notion that they use the high-discretion regime to accommodate their teaching. The effect of increasing instruction time may also depend on other factors, such as the educational level of the teachers or other available school resources. No single study will, therefore, settle the debate. However, the very low regulated treatment tested in the no teaching program condition makes the results relatively applicable to other contexts.

However, the exploratory results of the high-discretion no teaching program raise two concerns. First, boys did benefit from the intervention in terms of their reading skills, but they may also have experienced increased behavioral problems. Boys have been found to have less self-control (12), and therefore, making them work longer during the day may exhaust their self-control and thereby, create behavioral problems. Second, non-Western students seemed to show no or very little benefit of the intervention. Nevertheless, 73% of the teachers report that they also believed that the intervention benefitted bilingual students. This issue points to the other aspect of the instructional regime than the instruction, namely the assessment of the students' progress. Teachers did not seem to notice if the non-Western students did not benefit from the instruction. Therefore, it might be that the effect of a high-discretion regime on the instruction side would be even more effective combined with more regulation on the assessment side, thereby making teachers more aware of how their students respond to their teaching.

It should be emphasized that these considerations should be seen as hypotheses for future research, because the power of the

trial does not allow strong inference about the differences between the student groups and treatment conditions. The results do confirm, however, that increasing instruction time in an instructional regime with little formalization has positive average treatment effects on the reading skills of the students.

## Methods

**Participants.** The randomized, controlled trial was approved and funded by the Danish Ministry of Education and Aarhus University. All schools have volunteered to participate in the trial. Parents of students were informed about the content of the trial beforehand and told how to withdraw their child from the trial if they wished.

All interventions were implemented for students in grade 4 in the fall of 2013. Danish public schools that expected to have at least 10% bilingual students in grade 4 in the school year 2013/2014 were eligible to participate in the trial.

The participating schools were fully reimbursed for the costs associated with participation in the trial.

**Procedure.** In March of 2013, the Minister of Education sent an email to all municipalities in Denmark informing them about the upcoming randomized trial, eligibility criteria, and enrollment procedures. The municipalities were invited to enroll all of their eligible schools in the trial; 126 schools enrolled in the trial. We estimate that this constitutes about 37% of the eligible schools.

The trial was a two-stage, cluster-randomized trial with three treatment arms. Fig. S3 shows a diagram of the flow of schools and students participating in the trial. The two levels of randomization were school and classroom. First, administrative records from the school year of 2011/2012 were used to divide schools into strata based on the share of students of non-Western origin in grade 2 and the average score on the national reading test in grade 2. Each stratum contained four schools, and allocation to one of four experimental conditions (the two treatment arms, the control group, and a third treatment not related to instruction time and not analyzed here) was random within the stratum. Second, one classroom in each school was selected to participate in the trial by simple randomization. Randomization assures an unbiased distribution of baseline characteristics between experimental conditions, although some imbalance will occur in any finite sample. We find no substantially large imbalances between control and treatment groups (among 28 baseline student characteristics reported in Table S1, none of the mean values were significantly different across the treatment group without a teaching program and the control group, and only two means were significantly different across the treatment group with a teaching program and the control group). However, some minor imbalance occurs in baseline reading achievement between the control group and the teaching program group. To be conservative and because of a strong expected relationship between

baseline achievement and outcomes, the effect estimates presented here are controlled for baseline achievement in reading and math. *SI Methods* and *SI Results* has more details about the balance of the experimental conditions and the robustness of the results to different model specifications.

Although the trial included three treatment arms, the focus of this study is the two interventions that involved an increase in instruction time. Thus, we exclude schools that received the third treatment from our analyses. In the interventions that involved an increase in instruction time, the classrooms received four extra (45-min) lessons per week for 16 wk. In the first treatment arm, classrooms received extra instruction time in Danish with high teacher discretion (i.e., without a teaching program). Thus, the teachers were not provided with any explicit teaching material. In the second treatment arm, classrooms received extra instruction time in Danish with low teacher discretion (i.e., teachers were provided with very detailed teaching material containing texts and classroom exercises for each week of the intervention).

**Analyses.** We estimate each of the two treatment effects separately (i.e., including only observations in the relevant treatment arm and the control group). The empirical analysis presented is based on a linear regression model that includes a treatment indicator, stratum fixed effects, and baseline test scores. The stratum fixed effects are included to take into account that treatment assignment was random within strata. We account for the hierarchical structure of the data (students within classrooms) by clustering SEs at

the classroom level. Similar results are found using hierarchical linear modeling (Table S6), which would be expected based on simulations comparing hierarchical linear or multilevel models with models using clustered SEs (22). Because of attrition, outcome data are missing for some students. Test score attrition does not correlate with the treatment assignment, but participants in the control group were less likely to respond to the postintervention survey containing the SDQ outcomes (Table S2). We adjust for baseline achievement to present the more conservative estimates. Effect sizes are generally larger when we do not include baseline achievement (Table S3). We estimate the intention to treat effect, which is an estimate of the effect of assigning students to increased instruction time that does not impose assumptions about noncompliance. The intention to treat effect is of immediate relevance to policymakers, because it reflects the average treatment effect taking into account that not all students will comply with a policy that increases instruction time. For instance, students may transfer to a private school or other schools with no increase in instruction time.

Additional details about the trial and the analyses are in *SI Methods* and *SI Results*.

**ACKNOWLEDGMENTS.** We thank Michael Rosholm for supporting the setup of the project, Rambøll for data collection assistance and support, and Jessica Goldberg for helpful comments and discussions. The implementation and evaluation of the randomized trial were funded by the Danish Ministry of Education.

1. OECD (2014) *Education at a Glance* (OECD Publishing, Paris).
2. Patall E, Cooper H, Allen A (2010) Extending the school day or school year. *Rev Educ Res* 80(3):401–436.
3. Meyer E, van Klaveren C (2013) The effectiveness of extended day programs: Evidence from a randomized field experiment in the Netherlands. *Econ Educ Rev* 36:1–11.
4. Kikuchi N (2014) The effect of instructional time reduction on educational attainment. *J Jpn Int Econ* 32:17–41.
5. Jensen VM (2013) Working longer makes students stronger? *Educ Res* 55(2):180–194.
6. Parinduri RA (2014) Do children spend too much time in schools? Evidence from a longer school year in Indonesia. *Econ Educ Rev* 41:89–104.
7. Fryer RG (2014) Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Q J Econ* 129(3):1355–1407.
8. Cortes KE, Goodman JS (2014) Ability-tracking, instructional time, and better pedagogy: The effect of double-dose algebra on student achievement. *Am Econ Rev* 104(5):400–405.
9. Schochet PZ (2008) Statistical power for random assignment evaluations of education programs. *J Educ Behav Stat* 33(1):62–87.
10. Duckworth AL, Seligman MEP (2005) Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychol Sci* 16(12):939–944.
11. Baumeister RF, Vohs KD, Tice DM (2007) The strength model of self-control. *Curr Dir Psychol Sci* 16(6):351–355.
12. Duckworth AL, Seligman MEP (2006) Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *J Educ Psychol* 98(1):198–208.
13. Raver CC (2004) Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Dev* 75(2):346–353.
14. Raudenbush SW (2008) Advancing educational policy by advancing research on instruction. *Am Educ Res J* 45(1):206–230.
15. Weber M (1922) *Wirtschaft und Gesellschaft* (Mohr, Tuebingen, Germany).
16. Bawn K (1995) Political control versus expertise: Congressional choices about administrative procedures. *Am Polit Sci Rev* 89(1):62–73.
17. Huber JD, Shipan C (2002) *Deliberate Discretion: The Institutional Foundations of Bureaucratic Autonomy* (Cambridge Univ Press, Cambridge, United Kingdom).
18. Carpenter D, Chattopadhyay J, Moffitt S, Nall C (2012) The complications of controlling agency time discretion: FDA review deadlines and postmarket drug safety. *Am J Pol Sci* 56(1):98–114.
19. Wandall J (2011) National tests in Denmark – CAT as a pedagogic tool. *J Appl Test Tech* 12(Special Issue):3.
20. Goodman R (1997) The Strengths and Difficulties Questionnaire: A research note. *J Child Psychol Psychiatry* 38(5):581–586.
21. Goodman A, Goodman R (2009) Strengths and difficulties questionnaire as a dimensional measure of child mental health. *J Am Acad Child Adolesc Psychiatry* 48(4):400–403.
22. Green DP, Vavreck L (2008) Analysis of cluster-randomized experiments: A comparison of alternative estimation approaches. *Polit Anal* 16(2):138–152.
23. Statistics Denmark (2014) *Statistical Yearbook 2014* (Statistics Denmark, Copenhagen).
24. The Danish Ministry of Education (2010) *Minimumstimetallet fra Skoleåret 2010/2011 og Frem* (The Danish Ministry of Education, Copenhagen).
25. Pøhler L, Sørensen S (2010) *Nationale Test og Anden Evaluering af Elevens Læsning* (Dafolo Forlag, Copenhagen, Denmark), 1st Ed.
26. Goodman R, Meltzer H, Bailey V (1998) The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *Eur Child Adolesc Psychiatry* 7(3):125–130.
27. Puma M, Olsen R, Bell S, Price C (2009) *What to Do When Data Are Missing in Group Randomized Controlled Trials* (National Center for Education Evaluation and Regional Assistance, US Department of Education, Washington, DC).
28. Altman D (1985) Comparability of randomized groups. *Statistician* 34(1):125–136.
29. Bruhn M, McKenzie D (2009) In pursuit of balance: Randomization in practice in development field experiments. *Am Econ J Appl Econ* 1(4):200–232.
30. Krueger AB, Zhu P (2004) Another look at the New York City school voucher experiment. *Am Behav Sci* 47(5):658–698.
31. Kézdi G (2004) Robust standard error estimation in fixed-effects panel data models. *Hung Stat Rev* 9:95–116.
32. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6(2):65–70.