CESifo WORKING PAPERS

Central Exams and Adult Skills: Evidence from PIAAC

Lisa Leschnig, Guido Schwerdt, Katarina Zigova

CES**ifo**

Central Exams and Adult Skills:
Evidence from PIAAC

Abstract

Central exams are often hypothesized to favorably affect incentive structures in schools. Indeed, previous research provides vast evidence on the positive effects of central exams on student test scores. But critics warn that these effects may arise through the strategic behavior of students and teachers, which may not affect human capital accumulation in the long run. Exploiting variation in examination types across school systems and over time, we provide the first evidence that central exams positively affect adult skills. However, our estimates are small compared to the existing estimates for students, which may indicate some fade-out in the effect on skills over time.

Lisa Leschnig
Institute for Employment Research (IAB)
Nürnberg / Germany
Lisa.Leschnig@iab.de

Guido Schwerdt*
University of Konstanz / Germany
guido.schwerdt@uni-konstanz.de

Katarina Zigova
University of Konstanz / Germany
katarina.zigova@uni-konstanz.de

*corresponding author

1   Introduction

Central exams are associated with substantially higher test outcomes of students (see Bishop, 1997; Woessmann, 2003, 2005; Jürges et al., 2005; Fuchs and Wößmann, 2008). While this reduced-form pattern is well documented, critics warn that this result may simply reflect differences in students' test-taking ability, rather than actual differences in knowledge and skills. If that were so, central exams might not genuinely improve the productive skills of adults. Indeed, the existing evidence on the relationship between central exams and labor market outcomes is at best mixed (Bishop et al., 2000; Backes-Gellner and Veen, 2008; Piopiunik et al., 2013). However, there is no evidence on the direct effect on cognitive skills of adults, and the available evidence on the long-term impacts of central exams on labor market outcomes rests entirely on observational studies that estimate cross-sectional regression models.

This is the first paper to study the relationship between the type of exit examination at the end of secondary school and the cognitive skills of adults, as well as labor market outcomes in a panel framework. For this purpose, we draw upon data from OECD's survey of adult skills (OECD, 2013b). Internationally, the Programme for the International Assessment of Adult Competencies (PIAAC) is the most comprehensive survey of adult skills ever undertaken. It assesses literacy and numeracy skills and the ability to solve problems in technology-rich environments. It also collects information on labor market outcomes and background characteristics of representative samples of respondents aged 16-65. We supplement the PIAAC data with data on reforms of examination regimes since 1960 for 30 countries participating in PIAAC. We collected these data specifically for this analysis from a wide variety of sources (cf. Table A1). The data allow us to assign to each participant of PIAAC the type of exit exam that was in place in his or her country at the time of their graduation.

We leverage the panel structure (by graduation cohort across countries) of our data to identify the long-term effects of central exams by exploiting within-country variation in exam types over time and across countries.[1] In particular, we estimate regression models that include country fixed effects as well as country-specific linear trends, which ensure that unobserved time-invariant differences in relevant characteristics or differences in linear trends across countries do not confound our estimates. To benchmark our panel estimates, we also present results based on commonly estimated cross-sectional

---

[1]Note that exploiting this variation is different from comparing U.S. states with and without minimum competency exams or differing graduation requirements (e.g., Baker and Lang, 2013; Bishop and Mane, 2001; Dee and Jacob, 2007), as here graduation requirements in many cases do not change.

models that rely entirely on strong selection-on-observable assumptions.

Our findings reveal a positive relation between central exit exams at the end of secondary school and adult outcomes. We find that central exams increase the cognitive skills of adults by about 7 percent of a standard deviation. Effect sizes are very similar for numeracy and literacy, and only slightly higher for problem-solving skills. We find no significant effects on earnings, employment, or the probability of holding a tertiary degree. In terms of earnings, we can rule out effect sizes larger than 4,6 percent.

Our estimated effects on skills are small compared to the results of studies that examine student skills, which typically find central exam effects of more than 10 percent of a standard deviation. This difference might be the consequence of some fade-out in central exam effects. Given recent estimates of the returns to adult skills based on the same data (Hanushek et al., 2017), our small estimate of the skill effect is consistent with not finding any earnings effects above 4,6 percent.

We also conduct a series of robustness checks that reveal some interesting aspects. First, our main results are robust to a large set of sensitivity checks. Second, we supplement our main analysis with comparable estimates of the central exam effect on student skills. While these results are purely descriptive, they support the conjecture that central exam effects on skills fade out to some extent over time. Finally, we also provide some evidence for the existence of a positive relation between central exams and adult skills by exploiting the variation in exam types across German federal states. This finding within a single country concurs with our international results.

The remainder of the paper is structured as follows. Section 2 reviews existing evidence on the effects of central exams. In Section 3, we describe our self-collected longitudinal data on exit examinations, the PIAAC data as well as the construction of our estimation sample. In Section 4, we present our empirical strategy. Section 5 presents our main results, Section 6 discusses their robustness, and Section 7 presents further supporting evidence. Our conclusion is in Section 8.

2   Conceptual background

The theoretical foundation for our empirical analysis of the effects of central exams is laid out in models such as Bishop and Wößmann (2004) and Bishop (2006). Central exams may affect student skills and later outcomes through a variety of channels. First, by providing information on the outcomes of the educational process, central exams can improve the monitoring of the behavior of teachers and schools. Thus, central exams may raise educational outcomes as they improve accountability in school systems. Sec-

ond, if central exams improve the signal of educational achievement for employers and institutions of higher education (Becker and Rosen, 1992; Schwerdt and Woessmann, 2017), they can have true productivity effects by creating incentives for students to exert more effort.[2] Hvidman and Sievertsen (2019) provide evidence that grading incentives in high stakes tests significantly impact students' learning effort. Exploiting a school reform that exogenously re-coded the GPA of high school students, the authors show that students whose prior achievement was downgraded exerted a greater effort in the following school years to compensate for the reform-induced demotion. Finally, central exams may also decrease collective peer pressure against learning, because they render futile any collective strategy to lower standards in a classroom—which again increases learning outcomes.

An extensive empirical literature has investigated the effects of external exit exams on students' skills. Evidence from international student achievement tests indeed shows that students perform substantially better in countries with external exit exams than in countries without them (see Hanushek and Woessmann, 2011, for an overview). This pattern has been found on the 1991 International Assessment of Educational Progress (IAEP) in math, science, and geography (Bishop, 1997), the 1991 International Association for the Evaluation of Educational Achievement (IEA) Reading Literacy study (Bishop, 1999), the 1995 Third International Mathematics and Science Study (TIMSS; cf. Bishop, 1997; Woessmann, 2003), the 1999 TIMSS-Repeat study (Woessmann, 2005), the 2000 Programme for International Student Assessment (PISA) in reading, math, and science tests (Bishop, 2006; Fuchs and Wößmann, 2008), and the 2003 PISA reading, math, and science tests (Woessmann et al., 2009). These studies suggest that the effects of external exit exams on student achievement may well be larger than a whole grade-level equivalent, or between 20 and 40 percent of a standard deviation of the respective international tests. Studies in countries with regional variations in exam systems find similar results. Positive effects of external exit exams on test-score outcomes have been shown for Canadian provinces (Bishop, 1997, 1999), for U.S. states (e.g., Graham and Husted, 1993; Bishop et al., 2000), and for German states (Jürges et al., 2005; Jürges and Schneider, 2010; Jürges et al., 2012; Lüdemann, 2011). To probe causality, Jürges et al. (2005) apply a differences-in-differences approach to the German TIMSS 1995 data that exploits the fact that in some secondary-school tracks, the states with central exit exams have exams in math but not in science, finding a smaller, but still substantial, effect of about 0.13 of a national standard deviation.

---

[2]In a similar spirit, principal-agent models of educational standards model how educational credentials affect the level of learning effort that students choose (e.g., Costrell, 1997; Betts, 1998).

More recently, Bergbauer et al. (2018) show that an expansion in standardized external comparisons is associated with improvements in student achievement. Their panel analysis is based on a dataset of more than 2 million students in 59 countries observed over six waves in the international PISA student achievement test 2000–2015, and exploits reforms in assessment systems over time for identification. Most importantly, they show that both school-based and student-based external comparisons have positive effects on student achievement, where their definition of a student-based external comparison is closely aligned with our definition of a central exam.[3]

But students' higher test scores, as induced by central exams, may not automatically translate into better outcomes for adults. External testing could also have negative effects (see Figlio and Loeb, 2011, for a review). Critics often argue that test-based accountability systems may raise only test-taking skills, not genuine educational achievement (see Popham, 2001; Koretz, 2002; Volante, 2004). Increased test scores can also just be due to fraudulent behavior like outright cheating, as educators have an incentive to improve observed test outcomes (for example, Jacob and Levitt, 2003). In addition, Jürges and Schneider (2010) find that while central exit exams do improve students' academic skills, they negatively affect students' attitudes toward learning, as indicated in self-reported enjoyment of mathematics. And Jürges et al. (2012) indicate that the positive effect of central exams on curriculum-based knowledge does not extend to a positive effect on mathematical literacy.[4]

Only a few studies have investigated the relationship between central exams and adult outcomes. Backes-Gellner and Veen (2008) and Piopiunik et al. (2013) study labor market outcomes exploiting cross-sectional variations in exit exams across German federal states. Backes-Gellner and Veen (2008) fail to find a positive effect of central school exit exams on labor-market earnings for graduates of the highest school track of the highly tracked German school system. But Piopiunik et al. (2013) show that graduates from the lowest track have higher earnings if they received their high school leaving certificate in a state with central exams. Piopiunik et al. (2013) also show that graduates from both low-track and high-track schools have lower rates of unemployment when their school exit exam was centrally administered. However, these findings may be specific to Germany's rigid labor market, where earnings structures are mostly determined by central bargaining.

---

[3]In fact, Bergbauer et al. (2018) use our compiled dataset of the existence of central exit exams at the end of secondary school to create their measure of student-based external comparisons.

[4]Jürges et al. (2012) distinguish between test achievements reflecting the German mathematics curriculum and achievements in applying the knowledge to real world issues (i.e., mathematical literacy).

For the United States, Bishop et al. (2000) provide evidence on the effects of external exit exams, but their analysis ultimately boils down to a comparison of New York State to the remaining states. Wider variation across U.S. states is restricted to course graduation requirements and minimum competency exams, which assess only low-level skills in public schools and have no consequences for university entrance. Based on longitudinal data that allows linking the exam type of individual students with later labor-market outcomes, Bishop and Mane (2001) find minimum competency exams, but not mere course graduation requirements, to be positively associated with earnings.

Our paper directly contributes to this literature on central exams by providing a first analysis of the direct effect of central exams on adult skills, and a first analysis of the effect on labor market outcomes, in a panel framework that exploits several reforms to examination systems across a wide range of countries. We believe that this extension is important for several reasons. First, the existing evidence on the effect on labor market outcomes is based on regression models that exploit variation in exam regimes across regions. The evidence is therefore prone to biases arising because of unobserved institutional or cultural differences across regions. Second, the existing evidence on labor market effects is also limited to specific countries and might not be generalizable. Third, studying the direct effect of central exams on adult skills is important because positive effects of centralized examination systems may also arise through other channels besides increased human capital accumulation. For example, central exams may facilitate the matching process between vacancies and workers, as it allows for an improved sorting of students by productive skills. This may increase overall productivity in the long run.

3 Data

In this section we describe our data collection and the construction of our estimation sample. We also elaborate on how we match our data on exit examinations in 30 countries to individual-level data of the Programme for the International Assessment of Adult Competencies (PIAAC).

3.1 Longitudinal data on central exit exams

We begin our discussion with a definition of central exams, based on Bishop (1997, 1999). A central school leaving examination (in short, a central exam) is a mandatory written test, administered by a central authority (such as a ministry of education), that provides centrally developed, curriculum-based test questions covering such core

subjects as math and the mother tongue. The standardized questions should test content that is taught during upper-secondary schooling rather than general skills. Central exams impose consistent test implementation, grading, and passing threshold. Thus, central exams set external and consistent quality standards of secondary education, and ensure comparability of students' performance. Additionally, achievement in a central exam—and, especially, failing a central exam—is required to have direct consequences for students who take the exam, for example by representing a significant part of the final grade and by a limited number of valid retakes. Finally, central exams can be organized either on a national or regional level and must be mandatory for a cohort in a certain school track to ensure that a sufficient number of students take the exam.[5]

Following Bishop (1998 and 1997), we do not consider commercially prepared tests such as the Scholastic Assessment Test (SAT), American College Testing (ACT), or the California Achievement Test (CAT) to be central exams. These tests do not have direct consequences for students, and students can retake them multiple times. What is more, their content is not necessarily curriculum-specific, nor are these tests mandatory.

We also distinguish between university entrance exams and high school leaving exams, as only students who are aiming to attend a university typically participate in university entrance exams. But participating in a centrally organized university entrance exam is very common in Japan and Korea, even among students who do not actually intend to start tertiary education. Thus, we include these two countries in our definition of central exams in our preferred specification, but also show that our main results are robust to excluding these two countries from our analysis.

While our treatment is well defined, the counterfactual is not. School systems may not satisfy our definition of having a central exam for a variety of reasons. On the one hand, there might be a general absence of any school leaving exam (as in Spain and Belgium). On the other hand, some school leaving exams violate our definition across one or more dimensions. For example, exit exams could be administered at the local level (as in Greece and some federal states in Germany), may not be mandated for graduation (Sweden after 1968), may fail to have direct consequences for graduates (perhaps because of the option of multiple retakes, as with CAT and SAT tests), or may not be curriculum-based (as in many U.S. states).

For this analysis, we make use of a large variety of sources to collect data on exit exams at the end of secondary school. Our sources include international and national publications, both offline and online sources, and personal communications with coun-

---

[5]See Bishop (1997, pg. 260) for a more detailed list of characteristics of a central exam.

try experts. We provide a detailed documentation of our country-specific sources in Appendix—Table A1 and visualize the variation in central exam regimes by country and graduation cohort in Figure 1.

The history of national central exams started in France, where nationwide coordinated testing of high school graduates dates back to the early 19th century. Denmark followed 100 years later in 1908. In 1919 and 1923, Finland, Italy, and Israel introduced central exam systems, followed by Ireland, UK, Norway, and the Netherlands from the 1950s through the 1960s.[6] New Zealand implemented central exams, according to our definition, in 1968, and Singapore did so in 1975, shortly after gaining independence. Japan and Korea introduced centrally administered university entrance exams in 1979 and 1994. The more recent, larger wave of countries implementing central exams took place in Eastern Europe at the turn of the 21th century, with Estonia in 1997, Lithuania in 2002, Poland and the Slovak Republic in 2005, and the Czech Republic in 2011. Cyprus also introduced central exams in 2006.

The only countries that never introduced central exams until 2012, or 2015 for countries sampled in the second wave of PIAAC, are Belgium, Austria, Chile, Spain, and Turkey.[7] Sweden and Greece currently have no central exams, but these two are interesting cases. Sweden abolished mandatory central exams in 1968.[8] Greece introduced central exams in 1983 but abolished them in 2005, so we only consider graduation cohorts in Greece who graduated between 1983 and 2005 as being exposed to central exams.

The education systems in Canada, Germany, and the United States are organized at the province or state level rather than nationally. Examination systems also vary within these countries. Table A2 provides detailed information on these variations over time.

Currently, six of Canada's 13 provinces have centralized school leaving exams.[9] Three provinces—Manitoba, Nova Scotia, and Ontario—abolished central exams in the 1960s and 1970s, but Manitoba reintroduced them in 1991. In the U.S., according

---

[6]Israel introduced the "Bagrut" central exams before its independence, during the British Mandate of Palestine (1920–1948).

[7]Austria only recently introduced central exams in the 2015/16 school year. Therefore, Austria is treated as a non-central exam country for the purpose of our analysis.

[8]Swedish students still have the option of taking a centralized exam, but since this is entirely voluntary it does not meet our definition of a central exam. This coding is in line with Bishop (1998) and the assessment of country experts. Fuchs and Wößmann (2008), however, assumed for Sweden a share of 50 percent of students participating in central exams in the year 2000. Our results do not change if we assume a factor of 0.5 for Sweden after 1968.

[9]In the province of New Brunswick, the francophone sector introduced central exams in 1991; the anglophone sector has yet to do so.

to our definition, only New York State introduced central exams in 2003. During the 2000s, other U.S. states introduced curriculum-based school leaving exams, the so-called end of course exams (EoC exams), or made participating in the commercial ACT or SAT a mandatory part of graduation. A further limitation is that students can retake those tests multiple times, so they fail to provide a universal signal of graduate skills.[10]

By 2012, all of Germany's federal states but one (Rhineland-Palatine) had introduced central exams at the end of secondary education. Three states—Baden-Wuerttemberg, Bavaria, and Saarland—were the first to introduce central exams, between the late 1940s and the early 1950s. Before the German reunification, the former GDR states (or new federal states) also had some form of central exams in place (Kühn, 2010) but except for the state of Thuringia, they initially abolished central exams after reunification. However, all former GDR states quickly reintroduced central exams in the 1990s. In the 2000s, a number of other states, including Brandenburg, Hamburg, Berlin, Lower Saxony, North-Rhine Westphalia, Bremen, Hesse, and Schleswig-Holstein, switched to central exam regimes.[11]

Since PIAAC data does not provide information on the province or state of residence for surveyed individuals, we follow the common practice in empirical studies on central exams and assign a central exam factor to graduation cohorts in countries with education systems organized at a sub-national level (for example, Fuchs and Wößmann, 2008).[12] For this purpose, we weight regional central exam dummies for each state and each graduation year with the state's respective population share in 2012. Aggregating the weighted dummies, we obtain a single factor at the country level. For a given graduation cohort and on the federal level, this factor should roughly reflect the probability that an individual was exposed to a central exam regime. Figure 1 displays the evolution of these factors for the three countries in our sample with federal education systems. These factors range in the U.S. from 0 to 0.063, in Canada from 0.72 to 0.53, and in Germany from 0.33 to 0.95.

3.2 PIAAC data

The Programme for the International Assessment of Adult Competencies, was developed by the OECD to provide internationally comparable data on adult skills (OECD, 2013a; OECD, 2016). The first round of PIAAC was administered to 22 countries be-

---

[10]For a detailed survey of high school graduation testing across U.S. states, see Caves and Balestra (2018).
[11]For a detailed summary of the German school system, see Jürges et al. (2005), Lüdemann (2011), or Schwerdt and Woessmann (2017).
[12]Germany is an exception, as we can identify the federal state of residence in the German data. We exploit this information in Section 6.3.

tween August 2011 and March 2012; the second was administered to an additional eight countries between April 2014 and March 2015. The two waves combined thus provide comparable skill data for 30 countries.[13] In each participating country, a representative sample of at least 5,000 adults between 16 and 65 years of age was interviewed at home in the language of their country of residence. The standard survey mode was to answer questions on a computer, but respondents without sufficient computer knowledge could also do a pencil-and-paper survey.

PIAAC was designed to measure key cognitive and workplace skills needed for individuals to advance at work and participate in society. The survey assessed cognitive skills in three domains: numeracy, literacy, and problem-solving in technology-rich environments.[14] The tasks that respondents had to solve were often framed as real-world problems, such as maintaining a driver's logbook (numeracy domain) or reserving a meeting room on a particular date using a reservation system (problem-solving domain). The domains are described in more detail in OECD (2013a). For analytical purposes, we standardize skills in the full sample to have mean zero and standard deviation (SD) one, and employ the sample weights provided in PIAAC. In the empirical analysis, we focus on numeracy skills, which we deem most comparable across countries, but we report the results for the other two skill domains as well.

We also investigate whether central exams affect labor market and education outcomes. For this analysis, we make use of the data on labor-market status, earnings, education, and demographic characteristics, as provided by PIAAC. Specifically, we estimate central exam effect on gross hourly earnings of wage and salary workers, employment status, and educational attainment. These outcome measures have been used previously in the literature analyzing PIAAC data (for example Hanushek et al., 2015).[15]

We match our self-collected data on exam regimes to the PIAAC data by country of residence and year of completion of the highest level of schooling. That is, we assign to each PIAAC participant the type of exit examination at the end of high school that was in place in his or her country at the time of their graduation. High school

---

[13]The actual number of countries surveyed in the first and the second PIAAC rounds was 24 and 9. Following the OECD recommendation, we excluded Russia because the Russian sample doesn't cover the population of the Moscow region. We also had to exclude Australia because of restrictive access to the Australian data. From the second-round countries, we excluded Indonesia as it withdrew its participation in the PIAAC study and submitted only data from the Jakarta region.

[14]Participation in the problem-solving domain was optional; Cyprus, France, Italy, and Spain did not participate in this domain.

[15]In particular, see Hanushek et al. (2015) for detailed information on the construction of our earnings measure.

graduation in the PIAAC data is defined as having completed at least the Level 3a of the International Standard of Classification of Education (ISCED), according to the 1997 classification. As PIAAC provides information on the highest educational degree obtained and the age at which individuals left full-time education, we can derive the year of graduation directly for individuals whose highest educational degree is the high school diploma. Individuals who report having not graduated from high school are kept in our baseline sample and are assigned to the graduation cohort corresponding to the year they report having left full-time education. For individuals with higher educational degrees, we assume their graduation age to be the country-specific median age at high school graduation among those with high school as their highest degree. To ensure that all individuals in our sample have been exposed to at least two years of secondary schooling, we restrict our baseline sample to individuals with a graduation age of 15 or above.[16]

Note that our matching procedure implies a specific treatment definition, as some individuals in our sample will be considered to have been exposed to a central exam regime even though central exams were only introduced in their final high school year. Thus, some students whom we consider to be treated, might not themselves have participated in a central exam. However, they may be indirectly affected by the change in exam regimes (such as by spillover effects of increased school accountability). As a consequence, the treatment we are studying here could be described as "*at least some (direct or indirect) exposure to a central exam regime.*"

Table 1 provides summary statistics for our baseline sample in all 30 countries by exam regime. The table shows that, on average, individuals exposed to central exams are quite different from those in non-central exam regimes. In particular, they are more skilled, earn more, are more likely to be employed, and are more likely to hold a tertiary degree. These differences are not surprising, as the younger population is over-represented in the central exam category.

4   Empirical strategy

We estimate the effect of central exams on adult outcomes based on regression models of the following kind:

$$y_{icg} = \beta\, CE_{cg} + X'_{icg}\gamma + \delta_c + \theta_g + \lambda_c g + \varepsilon_{icg}, \tag{1}$$

---

[16]Results are, however, almost identical when early school leavers are not excluded from the sample. See Table 4, "Incl. early leavers" row .

where y_{icg} is the outcome of individual i in country c belonging to graduation cohort g, CE_{cg} is a dummy that indicates whether an individual belonging to graduation cohort g in country c was exposed to a central exit exam, δ_c is a country fixed effect, θ_g is a graduation cohort fixed effect, $\lambda_c g$ is a country-specific linear trend in graduation years, X_{icg} is a vector of student background variables, and ε_{icg} is an idiosyncratic error.

The key parameter of interest is β. It captures the effect of graduating in a central as opposed to non-central exam regime. By leveraging the panel nature of our data at the cohort-country level, we identify β in our preferred model solely based on within-country variation in examination regimes across graduation cohorts.[17] In particular, our identification strategy makes use of the fact that some countries changed their exam regimes in different years between 1960 and 2012, while others did not. Thus, unobserved institutional or cultural differences across countries that have a constant impact on adult outcomes do not bias our estimates.

Note that our preferred specification also includes country-specific linear trends in graduation years, $\lambda_c g$. This allows us to control, to some extent, for country-specific confounders that have a different impact on outcomes across graduation cohorts. This might be important, as Stephens and Yang (2014) have shown, in the context of studies exploiting reforms to compulsory schooling laws for identification, that identification based on the timing of policy changes across states over time can be very sensitive to the inclusion of state-specific trends.

To benchmark our preferred estimates, we also estimate commonly estimated cross-sectional regression models that do not control for country-specific effects. And in Section 6 we conduct a wide range of sensitivity checks and a permutation test to assess the robustness of our main results.

Throughout the empirical analysis, we estimate and report standard errors clustered at the country by graduation cohort level. We also follow other cross-country studies based on PIAAC data (such as Hanushek et al., 2015) and estimate weighted regressions using sampling weights, which we adjusted so that each country is given the same weight to account for different sample sizes across PIAAC countries.

5   Results

All else being equal, do central exams at the end of secondary education affect adult outcomes? In addressing this question, we begin with a set of estimates of the effects on

---

[17]Other studies that exploit variation in the exposure to education policies across cohorts in cross-sectional data for identification include Pischke (2007) and Pischke and von Wachter (2008).

adult skills, before presenting estimates of the effects on labor market and educational outcomes.

5.1 The effect of central exams on adult skills

We report our main estimates of the effects of central exams on adult skills in Table 2. Before presenting estimates of our preferred model given by Equation 1 in columns 3 to 8, we start by presenting a set of benchmark estimates that do not control for unobserved heterogeneity across countries in columns 1 to 3. As skill measures are standardized to have a mean of zero and standard deviation of one, the coefficient of interest can be interpreted as the effect of central exams on adult skills measured in percent of a standard deviation of the respective skill distribution.

Across the different specifications, there is clear evidence that central exams are positively related to adult skills. Without taking into account any control variables and unobserved heterogeneity across countries, graduates in central exam regimes out-perform graduates in non-central exam regimes by about 20 percent of a standard deviation in the numeracy test. The estimate decreases slightly to about 17 percent when controlling for individual background characteristics and the age at graduation. Interestingly, these effect sizes are comparable to estimates of central exam effects on student skills based on cross-sectional data (for example Bishop, 1997; Woessmann, 2002; Fuchs and Wößmann, 2008). The estimates on the background variables reveal several typical features of estimated education production functions that are in line with previous evidence.

Turning to our panel estimates, we observe that estimates become much smaller in absolute terms, but remain positive and significant. Including country fixed effects suggests that central exams increase numeracy skills on average by about 7 percent of a standard deviation. Introducing a country-specific linear time trend in column 4, does not change this estimate. For literacy skills, estimated effects are similar to those for numeracy, while the estimated effect on problem-solving skills is somewhat larger, at 9 percent of a standard deviation.

Our estimates on adult skills are about half the size of well-identified estimates of central exam effects on student skills (such as Jürges et al., 2005). This may indicate that the effect of central exams fades out as individuals grow older, which may reflect that part of the initial effect on student test scores is indeed driven by an increased test-taking ability gained during secondary education by students in central exam regimes.

5.2 The effect of central exams on labor market outcomes

To investigate whether the increase in adult skills induced by central exams also translates into better labor market outcomes, in Table 3 we report estimates of the central exam effects on earnings, employment, and college degree. The first column in Table 3 reports coefficient estimates of variants of Equation 1, with individual earnings as dependent variable. In column 2 we report estimates of the effect on employment.[18] Finally, we test whether central exams affect the probability of completing tertiary education and receiving a high school diploma in columns 3 and 4, respectively.[19]

Once we control for unobserved country-specific trends, we do not find any significant effects on these outcomes. Most importantly, our estimates let us rule out the existence of earnings estimates larger than 4,6 percent.[20] This finding is, however, consistent with the finding of effects on adult skills of about 7 percent of a standard deviation if central exams affect earnings primarily through a skill production channel. Hanushek et al. (2017) report an average wage return to a one-standard-deviation increase in numeracy skills of 20 percent. Based on a back-of-the-envelope calculation, we should in this case expect that a shift in skills of 7 percent of a standard deviation would produce a shift in earnings of only about 1 percent. Interestingly, our point estimate actually suggests an effect on earnings of about 1 percent, but we lack the statistical power to detect such small effect sizes.[21]

6 Robustness

In this section we conduct a set of sensitivity checks to assess the robustness of our main results.

6.1 Sensitivity checks

We begin by reporting estimates of a series of sensitivity checks in Table 4. To consolidate our presentation, we report estimates of the central exam effect based on all six outcomes in the first row. As shown in the second row of Table 4, our results prove to be mainly robust to estimating regression models that do not use any survey weights.

---

[18]Any participant in PIAAC who reports positive labor earnings is coded as employed.

[19]Note that in columns 3 and 4, we follow Hanushek et al. (2015) and exclude workers below age 25, as they are not likely to have completed their first phase of full-time education.

[20]The upper bound of the 95 percent confidence interval of the earnings estimate in the first column of Table 3 is 0.0461.

[21]The calculation is as follows: Central exams positively affect skills by about 7 percent of a standard deviation, while a standard deviation increase in skills increases earnings by about 20 percent (see Hanushek et al., 2017)). Thus, the expected effect of central exams on earnings is 0.07×20, which comes to 1.4 percent.

Another robustness concern may be that the graduation year is partly endogenous to the type of exit exam. We test the relevance of this concern by making an alternative assumption about the graduation age of individuals. Estimations in the third row are based on assigning the median graduation age within a country to all graduates in the country. The results show the robustness of our main results to this alternative assumption about individuals' graduation age. Finally, we show that our main results are generally robust to the inclusion of individuals who report having left school before age 15.

While some degree of country heterogeneity can be expected, another key concern is that our overall finding is entirely driven by a specific country. To investigate this possibility, we estimate our preferred specification several times, each time excluding a different country. Table 5 presents the results from this piecewise deletion exercise. The results are very similar, which shows that our main result is not driven by any specific country.

6.2 Permutation test

To check whether our estimates are driven by an accidental correlation between changes in exam regimes and other unobserved discrete changes that impact skill development, we conduct a placebo analysis in the spirit of the placebo and permutation tests conducted in Chetty et al. (2009).[22] Our placebo check replaces the actual year of the reform with a randomly chosen year between 1960 and 2012 for each of the 13 de facto regime-changing countries. We repeat this randomization procedure 2,000 times, estimating the reform effect on adult skills based on our baseline specification for each of those 2,000 iterations.[23]

Figure 2 displays the empirical cumulative distribution of central exam effect estimates from the 2,000 placebo randomizations. The results reveal that more than 95 percent of the placebo estimates are well below our baseline estimate. In most cases, estimates are close to zero and insignificant. Actually, the probability that the baseline effect of 0.067 would appear by an accidental correlation is 0.03. Thus, this exercise re-

---

[22]Chetty et al. (2009) use the test to check for robustness in their estimates of posting-tax on demand effect. They estimate a placebo effect immediately before and immediately after the posting-tax experiment in order to check for unusual patterns in demand around the timing of the experiment.

[23]Note that the direction of the placebo policy change is assumed to be identical to the actual direction of the reform, i.e., abolishment or introduction of central exams. For Greece, which experienced two changes in exam regimes between 1960 and 2012, the randomization assumes that the first randomized year of change must be smaller than the second. For simplicity, in this analysis we dropped the three countries with federal education systems. The data for the 14 countries that didn't change their central exam regime are not manipulated in this placebo analysis.

veals that the potential for discrete changes in unobserved impacts on skills that could generate effect sizes comparable to our baseline estimate is very low in our data.

6.3 Exploiting within-country variation

For the German subsample, PIAAC provides identifiers for the respondents' federal state of residence, which allows us to exploit the variation in exam regimes across federal states within a single country. The results should be directly comparable to the findings of Backes-Gellner and Veen (2008) and Piopiunik et al. (2013).

Table 6 presents the results of this analysis. The set of controls used for the skills outcomes (columns 1 to 3) is identical to Table 2, and the set of controls used for labor market outcomes (columns 4 to 6) is identical to Table 3's. Given that we cannot use specification with state (*Länder*) fixed effects, the estimate for numeracy in Table 6 corresponds to the international estimate in column 2 of our baseline table (Table 2).

For numeracy and literacy skills, we find a significant positive association with central exams. The estimates for problem-solving skills are also positive and economically important, but only weakly significant. Our estimated effect of 0.11 for numeracy is fairly close to the preferred effect of central exams to the TIMSS math score, namely 0.13 of national standard deviation, as evidenced by Jürges et al. (2005).

As for labor market outcomes, we do not find a significant link between central exams and earnings, but we do find a positive and highly significant relation with employment probabilities. This pattern is broadly in line with the estimation results of Backes-Gellner and Veen (2008) and Piopiunik et al. (2013), which exploit the same variation but use data from the German Socio-Economic Panel (GSOEP) for their analysis.

While these associations do not necessarily warrant a causal interpretation, overall they support the interpretation of our baseline estimates. In particular, they increase confidence in the validity of our analysis of the labor market effects of central exams in PIAAC, as we find a pattern of results that is similar to the existing evidence on the same association based on an independent data source.

6.4 Other education reforms

Another concern is that education reforms often come in packages that not only target a specific element (such as introducing central exams) but also impact multiple levels of an education system. Thus, the effect that we're capturing in our analysis might not be attributable solely to changes in the central versus local nature of exit exams if

other elements of the education system systematically covary with exit exam systems.

Our database on exit exams is the result of a substantial data collection effort that specifically focused on collecting information on exit exams. While we are confident that we can adequately capture changes in the central versus local nature of exit exams in the 30 PIAAC countries between 1960 and 2012, we have no time-varying information on other elements of the education systems. Thus, we have no formal way of testing whether other features of education systems systematically covary with exit exams.

We did, however, further investigate the reforms to the exam systems we have identified, and checked what other features of the education systems might have changed simultaneously. Specifically, for the countries that either introduced or abolished central exams in 1960–2012, we also collected information on other education reforms that had been passed during the time frame of five years before and after the respective change to the exam system. While we have not been successful in all cases, we see no clear pattern suggesting that there is indeed a common set of other institutional changes that typically accompanies changes in exit exams in the countries where we can study this.

To provide just two examples, we elaborate further on the cases of Sweden and Greece.[24] Sweden abolished central exams in a comprehensive upper-secondary education reform in 1968. With this reform, the former division of academic and vocational post-compulsory education was rescinded and a common curriculum was implemented for all school tracks. Prior to the reform, academic upper-secondary schools, upper-secondary technical schools, junior colleges of commerce, and the two-year continuation schools had already been brought under the control of a single authority in 1964. After the of the 1968 reform, substantial effort was devoted to modernizing vocational education at the upper-secondary level and preparing the work force for the economic challenges of the time (Lundahl et al., 2010; Garrouste, 2010). The 1983 introduction of central exams in Greece was also surrounded by other education reforms. Law 186, passed in 1975, brought changes in teachers' recruitment, training, and salaries. In 1976, Law 309 laid the foundation for changes at the organizational and administrative level of general education, and established several councils to address teachers' working conditions.[25] The Presidential Decree of 1978, and Law 1286 and Law 1304 of 1982, further reformed teacher recruitment and training. Law 1566, passed in 1985, targeted

---

[24]A more comprehensive documentation of our investigation of educational reforms surrounding changes to the examination system is available upon request.

[25]Law 309 also replaced classical Greek by modern Greek as official language and extended compulsory schooling from 6 to 9 years (Garrouste, 2010).

the structure and operation of primary and secondary education, and again set higher training standards for teachers (Garrouste, 2010).

These two examples show that changes to exam systems are indeed, in some cases, embedded in a larger set of reforms that typically share a specific educational philosophy or are aimed at achieving common objectives. Thus, a more careful interpretation of our results shows that educational reform packages that typically include strengthening the central nature of exit exams as one element do lead to higher adult skills, while a change in the central versus local nature of exit exams alone might have no effect. However, our investigation of other reforms does not suggest that there actually was a set of other institutional changes that systematically accompanied the changes in the exam regimes we are studying. In fact, while we admit that we cannot fully exclude the possibility that we missed some reforms or specific elements of reforms, we believe that it is generally a strength of our empirical strategy that we exploit not just a specific reform for identification, but several reforms implemented in different countries in different decades. The one shared element of all these reforms is that they changed the central versus local nature of exit exams at the end of high school.

7 Heterogeneous effects

In this section we explore potential effect heterogeneities. We are particularly interested in understanding whether central exam effects vary across different subgroups of the population, whether reforms introducing or abolishing central exams have immediate effects on student performance, and whether the effect of central exams on individuals' skills tends to decay over time.

7.1 Subgroup analysis

Overall, we provide evidence for the average effects of central exams on adult skills and labor market outcomes. But this may conceal a substantial detrimental effect of central exams for some students. In Table 7 we address this concern by replicating our main analyses for specific subgroups of the population. Overall, the results of these analyses provide no evidence of detrimental effects of central exams for any subgroup. In the broader literature on school accountability, one issue that is often discussed is whether school accountability policies can effectively decrease gaps in student outcomes for disadvantaged groups.[26] Such effect heterogeneities could arise if central exams act

---

[26]For example, Hanushek and Raymond (2005) find that the U.S.-wide National Assessment of Educational Progress narrowed the Hispanic-White gap, but actually increased the Black-White gap.

as a strong monitoring instrument in schools, which may especially benefit students whose socioeconomic backgrounds are otherwise less likely to detect or to react to educational failures.

Table 7 reveals, however, that the effect of central exams on adult skills is largely independent of social background. We find positive effects on skills for all groups. With respect to the educational background of their parents, we find that effects, if anything, may be slightly larger for students from more educated households. Interestingly, the estimates of the effects on earnings point in a different direction. While the effects are not significantly different from zero for individuals from more educated households, individuals whose parents lack a secondary degree actually seem to benefit from central exams in terms of their later earnings. With respect to migration background, we find that second-generation migrants benefit slightly more from central exams in terms of their skill development, but differential effects on labor market outcomes do not exist.

Finally, gender may be another important aggravating or mitigating factor. For example, central exams may reduce the relative competition for good grades within a class, as all the students are evaluated with respect to an external standard. This may trigger behavioral responses that differ by gender. Our subgroup analysis indeed reveals some positive effects of central exams on labor market outcomes for women, but not for men. But the positive effects of central exams on skills are similar across gender, which suggests that the more positive effects on labor market outcomes for women is unlikely to be driven by any differential impact of central exams by gender on skill development.

7.2 Time to effect

The introduction of central exams might not be effective from the onset.[27] On the one hand, this could be the case if teachers and schools initially face problems with implementation or adjustment. On the other hand, it might take some time for the accountability mechanism set in motion by the introduction of central exams to unfold its full impact. If centralized exams indeed make differences in school quality salient, it might be a year or two until parents, teachers, and schools react to this new information and adjust their behavior in ways that will ultimately lead to more effective learning and teaching.

We test for this "time to effect" assertion by estimating variants of Equation 1,

---

[27]In a different policy context, for example, Hanushek and Raymond (2005) show gradual growth of the acceptance of the National Assessment of Educational Progress criteria across U.S. states over the 1993-2002 period.

which allow for heterogenous impacts of central exams by years of existence of the central exam policy. In particular, we interact the central exam dummy with two dummies indicating whether central exams exist in a country for more than two and five years, respectively.[28]

Table 8 shows the results of this analysis. Our findings suggest that the impact of the introduction of central exams unfolds with a delay of about two years. The estimates in columns 2 and 3 show that the central exam effect becomes statistically significant two years after the policy change. Thereafter, the effect of central exams does not appear to increase any further. Overall, this finding for the effects on adult skills is consistent with evidence presented in Lüdemann (2011), who studies reforms of test regimes in German federal states in a difference-in-differences design, and shows that central exam effects already affect student skills in the short run.

7.3 Fade-out of central exam effects

Comparing our results with findings of the existing literature on student skills can be informative about the extent of fade-out in the effects of central exams. However, differences in sample populations potentially blur this comparison. We address this concern at least partly by investigating simple correlations between central exams and the skills of both students and adults for samples drawn from the same population. Specifically, PISA 2000 surveyed a representative sample of 15-year-old students in 2000. Thus, the PISA sample should be sampled from about the same population as the subsample of PIAAC participants who were 15 years old in 2000. Similarly, we construct a correspondence data set for the TIMSS Repeat (1998/1999) data.

We report the results of this exercise in Table 9. All estimates stem from simple regressions of skills on a central exam dummy that differs across countries. Unfortunately, there is only a limited set of countries that participated in both PIAAC and PISA 2000 (columns 1 and 2) and countries that participated in both PIAAC and TIMSS Repeat (columns 3 and 4). The estimates are all insignificant, but the estimated coefficients in the student data are large and roughly comparable in size to the analogous estimates reported in studies that use the PIAAC 2000 and TIMSS Repeat data to study the central exam effect on student skills (for example, Woessmann, 2002; Fuchs and Wößmann, 2008; Jürges et al., 2005). Most notably, the estimated coefficients on the

---

[28]For presentation reasons, we chose a model with just two interactions at two and five years after the policy change. A more flexible model with interactions for each year reveals that this split roughly captures the overall pattern of effect heterogeneities quite well. Results of the model with interactions for each year after the policy change are available upon request.

central exam dummy are one third to one half lower if the dependent variable is adult skills rather than student skills. Thus, this crude comparison of simple correlations between central exams and skills based on comparable samples also indicates that the association between central exams and skills becomes weaker over time.

8   Conclusion

This paper is the first to investigate the relationship between the type of exit examination at the end of secondary school and the cognitive skills and labor market outcomes of adults, using a panel framework. We find that central exams increase the cognitive skills of adults by about 7 percent of a standard deviation.

Compared to the results of studies that examine student skills, which typically find central exam effects between 10 and 20 percent of a standard deviation, the estimates of the long-run effects on skills are small. This difference might be the consequence of some fade-out in central exam effects—which, in turn, might occur because part of the central exam effects on students was driven by a higher test-taking ability of students in central exam regimes, an ability that does not affect human capital accumulation in the long run.

In terms of labor market outcomes, we do not detect any significant effects. In particular, our statistical power allows us to rule out effects on earnings above 4,6 percent. But this finding is consistent with recent evidence on the average wage return to increases in adult skills of about 7 percent of a standard deviation, if we assume that central exams affect earnings only through their impact on skill production.

These findings contribute to the growing literature on accountability systems in schools in general, and the literature on central exams in particular. Overall, our results corroborate the notion of the positive effects of central exams on skill development. However, our finding of comparatively small effects on adult skills might suggest that larger estimates of the effect of standardized testing in secondary education are, to some extent, driven by short-run improvements in students' test scores, which have no impact on skill development in the long run.

References

Backes-Gellner, U. and S. Veen (2008). The consequences of central examinations on educational quality standards and labour market outcomes. *Oxford Review of Education 34*(5), 569–588.

Baker, O. and K. Lang (2013). The effect of high school exit exams on graduation, employment, wages and incarceration. Working Paper 19182, National Bureau of Economic Research.

Becker, W. E. and S. Rosen (1992). The learning effect of assessment and evaluation in high school. *Economics of Education Review 11*(2), 107–118.

Bergbauer, A. B., E. A. Hanushek, and L. Woessmann (2018). Testing. Working Paper 24836, National Bureau of Economic Research.

Bethell, G. and A. Zabulionis (2000). *Examination Reform in Lithuania: Background, Strategies and Achievements*. National Examination Centre.

Betts, J. R. (1998). The impact of educational standards on the level and distribution of earnings. *American Economic Review 88*(1), 266–75.

Bishop, J. H. (1997). The effect of national standards and curriculum-based exams on achievement. *American Economic Review: Papers and Proceedings 87*(2), 260–264.

Bishop, J. H. (1998). The effect of curriculum-based external exit exam systems on student achievement. *Journal of Economic Education 29*(2), 171–182.

Bishop, J. H. (1999). Are national exit examinations important for economic efficiency? *Swedish Economic Policy Review 6*(2), 349–401.

Bishop, J. H. (2006). Drinking from the fountain of knowledge: Student incentive to study and learn - externalities, information problems, and peer pressure. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*. Amsterdam: North-Holland. 909–944.

Bishop, J. H. and F. Mane (2001). The impacts of minimum competency exam graduation requirements on high school graduation, college attendance and early labor market success. *Labour Economics 8*(2), 203–222.

Bishop, J. H., J. Y. Moriarty, and F. Mane (2000). Diplomas for learning, not seat time: The impacts of new york regents examinations. *Economics of Education Review 19*(4), 333–349.

Bishop, J. H. and L. Wößmann (2004). Institutional effects in a simple model of educational production. *Education Economics 12*(1), 17–38.

Boezerooy, P. and J. Huisman (2000). From secondary to tertiary education in the Netherlands. *Journal of Institutional Research 9*(1), 54–64.

Caves, K. and S. Balestra (2018). The impact of high school exit exams on graduation rates and achievement. *Journal of Educational Research 111*(2), 186–200.

Chang, B.-M. (2009). Korea's english education policy innovations to lead the nation into the globalized world. *Journal of Pan-Pacific Association of Applied Linguistics 13*(1), 83–97.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and taxation: Theory and evidence. *American Economic Review 99*(4), 1145–77.

Costrell, R. M. (1997). Can centralized educational standards raise welfare? *Journal of Public Economics 65*(3), 271–293.

Dee, T. S. and B. A. Jacob (2007). Do high school exit exams influence educational attainment or labor market performance? In A. Gamoran (Ed.), *Standards-based reform and the poverty gap: Lessons for no child left behind*. Washington, D.C.: Brookings University Press.

Eckstein, M. A. and H. J. Noah (1989). Forms and functions of secondary-school-leaving examinations. *Comparative Education Review 33*(3), 295–316.

Figlio, D. and S. Loeb (2011). School accountability. *Handbook of the Economics of Education 3*, 383–421.

Fuchs, T. and L. Wößmann (2008). What accounts for international differences in student performance? A re-examination using PISA data. In C. Dustmann, B. Fitzenberger, and S. Machin (Eds.), *The Economics of Education and Training*. Heidelberg: Physica-Verlag HD. 209–240.

Garrouste, C. (2010). 100 years of educational reforms in Europe: A contextual database. Working Paper 31853, Munich Personal RePEc Archive.

Goulas, S. and R. Megalokonomou (2015). Knowing who you are: The effect of feedback information on short and long term outcomes. Working Paper 1075, Warwick Economics Research Paper Series.

Graham, A. E. and T. A. Husted (1993). Understanding state variations in SAT scores. *Economics of Education Review 12*(3), 197–202.

Greger, D. and E. Kifer (2012). Lost in translation: Ever changing and competing purposes for national examinations in the Czech Republic. *Journal of Pedagogy / Pedagogický Časopis 3*(1), 43–81.

Hall, C. (2000). National certificate of educational achievement: Issues of reliability, validity and manageability. *New Zealand Annual Review of Education 9*, 173–196.

Hanushek, E. A. and M. E. Raymond (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management 24*(2), 297–327.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review 73*(1), 103–130.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2017). Coping with change: International differences in the returns to skills. *Economics Letters 153*, 15–19.

Hanushek, E. A. and L. Woessmann (2011). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*. North Holland, Amsterdam. 3, 89-200.

Hörner, W., H. Döbert, B. von Kopp, and W. Mitter (2007). *The education systems of Europe*. Springer.

Hvidman, U. and H. H. Sievertsen (2019). High-stakes grades and student behavior. *Journal of Human Resources*, forthcoming.

Ilc, G., V. R. Gabrovec, and A. Stopar (2014). Relating the Slovenian secondary school English language national examinations to the CEFR: Findings and implications. *Linguistica 54*(1), 293–308.

Jacob, B. A. and S. D. Levitt (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics 118*(3), 843–877.

Jürges, H., W. F. Richter, and K. Schneider (2005). Teacher quality and incentives: Theoretical and empirical effects of standards on teacher quality. *Finanz Archiv: Public Finance Analysis 61*(3), 298–326.

Jürges, H. and K. Schneider (2010). Central exit examinations increase performance... but take the fun out of mathematics. *Journal of Population Economics 23*(2), 497–517.

Jürges, H., K. Schneider, and F. Büchel (2005). The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. *Journal of the European Economic Association 3*(5), 1134–1155.

Jürges, H., K. Schneider, M. Senkbeil, and C. H. Carstensen (2012). Assessment drives learning: The effect of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review 31*(1), 56–65.

Klein, E. D., S. M. Kühn, I. v. Ackeren, and R. Block (2009). Wie zentral sind zentrale Prüfungen? Abschlussprüfungen am Ende der Sekundarstufe II im nationalen und internationalen Vergleich. *Zeitschrift für Pädagogik 55*(4), 596–621.

Koretz, D. M. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources 37*(4), 752–777.

Kühn, S. M. (2010). Zentrale Abschlussprüfungen in der Diskussion. In *Steuerung und Innovation durch Abschlussprüfungen?* Springer. 25–26.

Lamprianou, I. (2012). Unintended consequences of forced policy-making in high stakes examinations: The case of the republic of Cyprus. *Assessment in Education: Principles, Policy & Practice 19*(1), 27–44.

Lavy, V. (2008). Does raising the principal's wage improve the school's outcomes? Quasi-experimental evidence from an unusual policy experiment in Israel. *Scandinavian Journal of Economics 110*(4), 639–662.

Lim, E. P. Y. and A. Tan (1999). Educational assessment in Singapore. *Assessment in Education: Principles, Policy & Practice 6*(3), 391–404.

Lüdemann, E. (2011). Intended and unintended short-run effects of the introduction of central exit exams. In E. Lüdemann (Ed.), *Schooling and the formation of cognitive and non-cognitive outcomes*, Ifo Beiträge zur Wirtschaftsforschung 39. Ifo Institut.

Lundahl, L., I. E. Arreman, U. Lundström, and L. Rönnberg (2010). Setting things right? Swedish upper secondary school reform in a 40-year perspective. *European Journal of Education 45*(1), 46–59.

Nardi, E. (2001). The transition from school to university in italy: Examination reform and outstanding issues. *Assessment in Education: Principles, Policy & Practice 8*(3), 339–351.

OECD (2002). Reviews of national policies for education: Lithuania. DOI: https://doi.org/10.1787/9789264193475-en.

OECD (2010). Oecd economic surveys: Chile. DOI: https://doi.org/10.1787/eco_surveys-chl-2010-en.

OECD (2013a). Oecd skills outlook 2013: First results from the survey of adult skills. DOI: http://dx.doi.org/10.1787/9789264204256-en.

OECD (2013b). Technical report of the survey of adult skills (PIAAC). Available at www.oecd.org/site/piaac/_Technical%20Report_17OCT13.pdf.

OECD (2016). Skills matter: Further results from the survey of adult skills. DOI: http://dx.doi.org/10.1787/9789264258051-en.

Piopiunik, M., G. Schwerdt, and L. Woessmann (2013). Central school exit exams and labor-market outcomes. *European Journal of Political Economy 31*(1), 93–108.

Pischke, J.-S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. *Economic Journal 117*(523), 1216–1242.

Pischke, J.-S. and T. von Wachter (2008). Zero returns to compulsory schooling in Germany: Evidence and interpretation. *The Review of Economics and Statistics 90*(3), 592–598.

Polydorides, G. (1986). The determinants of educational achievement at the end of secondary schooling: The case of Greece. Conference paper, ERIC.

Popham, W. J. (2001). Teaching to the test? *Educational Leadership 58*(6), 16–21.

Schwerdt, G. and L. Woessmann (2017). The information value of central school exams. *Economics of Education Review 56*(1), 65–79.

Stephens, Melvin, J. and D.-Y. Yang (2014). Compulsory education and the benefits of schooling. *American Economic Review 104*(6), 1777–92.

Volante, L. (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy* (35).

Watanabe, Y. (2013). The national center test for university admissions. *Language Testing 30*(4), 565–573.

Woessmann, L. (2002). Central exams improve educational performance: International evidence. Kiel Discussion Paper 397, Kiel Institute for the World Economy.

Woessmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics 65*(2), 117–170.

Woessmann, L. (2005). The effect heterogeneity of central examinations: Evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics 13*(2), 143–169.

Woessmann, L., E. Luedemann, G. Schuetz, and M. R. West (2009). *School Accountability, Autonomy, and Choice Around the World*. Edward Elgar, Cheltenham, UK.

Yemini, M., H. Bar-Nissan, and O. Yardeni (2014). Between "us" and "them": Teachers' perceptions of the national versus international composition of the Israeli history curriculum. *Teaching and Teacher Education 42*(1), 11–22.

Yildirim, I., T. Ergene, and K. Munir (2007). High rates of depressive symptoms among senior high school students preparing for national university entrance examination in Turkey. *The International Journal on School Disaffection 4*(2), 35–44.

Figure 1: Central exam regimes in PIAAC countries (1960-2015)

Central exams

High school graduation year
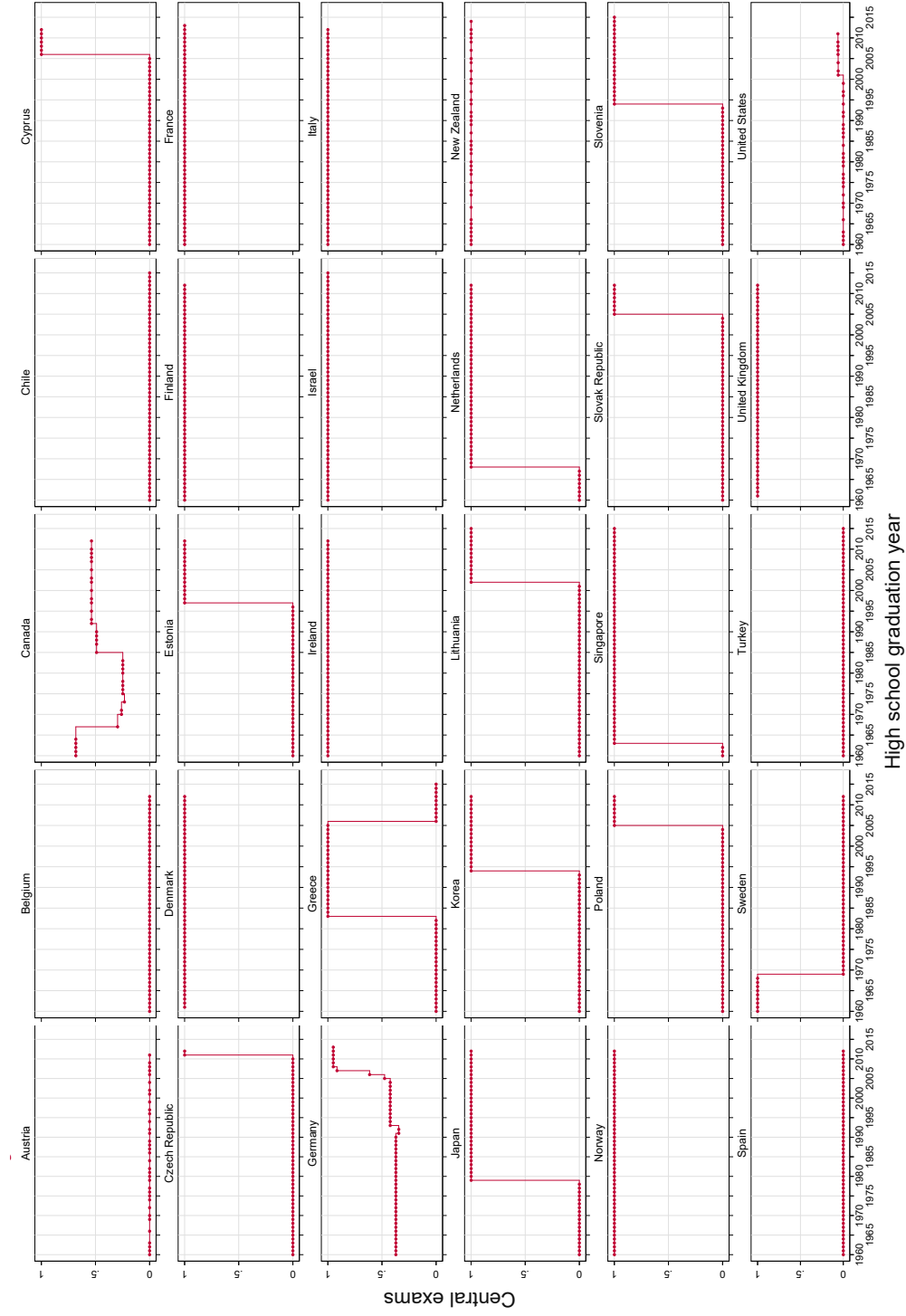
Note: Figure displays the evolution of the exposure to central exams in 30 countries by high school graduation cohorts. In the three countries with federal education systems, namely Canada, Germany, and United States, the exposure to central exams is measured as the share of the population exposed to central exams.

Figure 2: Distribution of central exam effect estimates from placebo randomizations
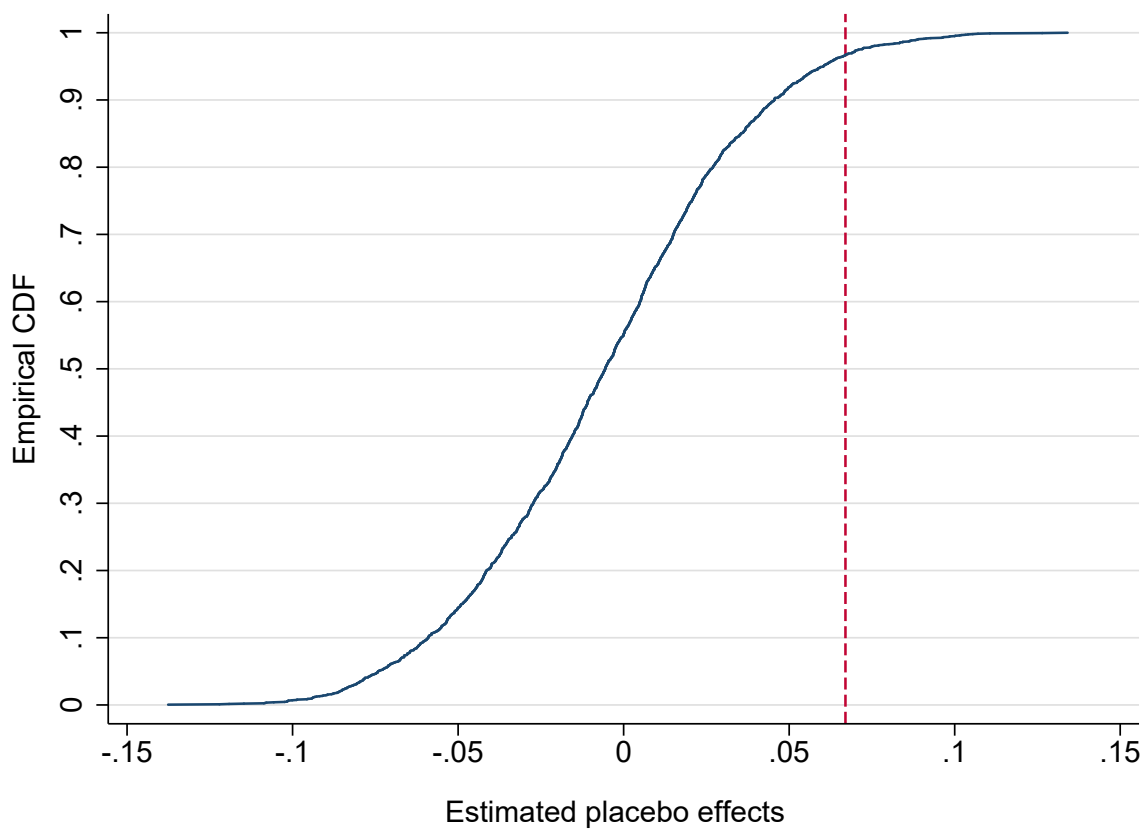
Note: The figure displays the empirical distribution of placebo estimates of central exam effects based on 2,000 randomizations of central exam change. The randomizations are based on the 27-countries sample that excludes the three federal education systems, and use the specification (4) of Table 2. The vertical line corresponds to the effect estimate of .067 for the sample without the three federal education systems reported in Table 5.

Table 1: Descriptive statistics

	No central exams		Central exams		
	Mean	Std. dev.	Mean	Std. dev.	Mean diff.
Numeracy	.053	(.962)	.227	(.932)	.174
Literacy	.017	(.935)	.235	(.916)	.218
Problem-solving	−.118	(.989)	.132	(.938)	.25
Earnings	2.44	(.754)	2.8	(.636)	.361
Employment	.553		.619		.066
College degree	.299		.371		.072
Age	42.8	(13.8)	36.6	(14.3)	−6.2
Age at graduation	18.5	(3.38)	18.7	(4.22)	.226
Female	.495		.498		.0034
Second-gen. migrant	.088		.14		.051
Parent education:					
Primary	.395		.324		−.071
Secondary	.368		.349		−.019
Above secondary	.2		.27		.07
Missing	.037		.057		.02
Observations	76,779		84,573		

Notes: Means and standard deviations by exam regime, for variables used in empirical analysis. Individuals from Germany, Canada, and the USA are included in the central exam group if their central exam factor is > 0.5. Otherwise they are classified as non-central exam. Observations are weighted by adjusted sampling weights that give same weight to each country. Baseline sample: high-school graduates aged 16-65. Numbers of observations are smaller, than those indicated, for the following variables: for earnings, employment, and college degree we exclude those still studying; earnings and employment sample additionally exclude those younger than 25; problem-solving skills are not available for Cyprus, France, Italy, and Spain.

Table 2: The effect of central exams on adult skills

Outcome	Numeracy skills				Literacy		Prob. solv.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Central exams	.205***	.168***	.065***	.067***	.049***	.068***	.135***	.091***
	(.019)	(.017)	(.015)	(.018)	(.017)	(.020)	(.021)	(.020)
Female		-.188***	-.188***	-.188***	-.017***	-.017***	-.079***	-.080***
		(.006)	(.006)	(.006)	(.005)	(.005)	(.007)	(.007)
Second-gen. migrant		-.023*	-.026**	-.009	-.031***	-.011	-.028**	-.008
		(.012)	(.010)	(.011)	(.010)	(.011)	(.012)	(.012)
Parent education:								
Secondary		.276***	.230***	.229***	.231***	.232***	.193***	.199***
		(.009)	(.008)	(.008)	(.008)	(.008)	(.010)	(.010)
Above secondary		.570***	.520***	.520***	.516***	.515***	.500***	.493***
		(.010)	(.010)	(.010)	(.010)	(.009)	(.012)	(.012)
Educ. missing		-.216***	-.204***	-.201***	-.164***	-.161***	-.053***	-.051***
		(.018)	(.015)	(.015)	(.015)	(.015)	(.018)	(.018)
Grad. year FE	yes	yes	yes	yes	yes	yes	yes	yes
Grad. age FE	no	yes	yes	yes	yes	yes	yes	yes
Country FE	no	no	yes	yes	yes	yes	yes	yes
Country-specific linear trend in grad. year	no	no	no	yes	no	yes	no	yes
Countries	30	30	30	30	30	30	26	26
Observations	161,352	161,352	161,352	161,352	161,352	161,352	129,519	129,519
Clusters	1,478	1,478	1,478	1,478	1,478	1,478	1,453	1,453
R^2	.100	.185	.253	.259	.246	.253	.203	.222

Notes: Least squares regressions weighted by adjusted sampling weights that give same weight to each country. Dependent variables are the globally standardized test scores indicated in first row. Sample: high school graduates aged 16-65. Robust standard errors clustered by graduation year × country in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 3: The effect of central exams on labor market and education outcomes

Outcome	Earnings	Employment	College degree	High school
	(1)	(2)	(3)	(4)
Central exams	.008	−.003	.027*	−.009
	(.019)	(.006)	(.016)	(.006)
Female	−.176***	.004*	.051***	.013***
	(.006)	(.002)	(.005)	(.002)
Second-gen. migrant	.009	−.014***	−.002	.003
	(.009)	(.004)	(.007)	(.004)
Parents education:				
Secondary	.110***	.011***	.095***	.039***
	(.007)	(.003)	(.005)	(.003)
Above secondary	.229***	.013***	.313***	.045***
	(.008)	(.003)	(.007)	(.003)
Educ. missing	−.028*	−.029***	−.039***	−.016*
	(.015)	(.007)	(.009)	(.009)
Grad. year FE	yes	yes	yes	yes
Grad. age FE	yes	yes	yes	yes
Country FE	yes	yes	yes	yes
Country-specific linear trend				
in grad. year	yes	yes	yes	yes
Countries	30	30	30	30
Observations	70,915	76,025	62,485	76,025
Clusters	1,383	1,392	1,313	1,392
R^2	.457	.087	.323	.547

Notes: Least squares regressions weighted by adjusted sampling weights that give same weight to each country. Dependent variables are log gross hourly wage (column 1), an employment dummy (column 2), a dummy for tertiary education degree (column 3), and a high school attainment dummy (column 4). Sample: individuals aged 25-65 with non-zero earnings (column 1), individuals aged 25-65 with non-zero earnings and unemployed (columns 2, 3, 4). Robust standard errors clustered by graduation year × country in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 4: Sensitivity analysis

Outcome	Adult skills			Labor market and education		
	Numeracy (1)	Literacy (2)	Prob. solv. (3)	Earnings (4)	Empl. (5)	College deg. (6)
Baseline	.067***	.068***	.091***	.008	−.003	.027*
	(.018)	(.020)	(.020)	(.019)	(.006)	(.016)
No weights	.079***	.089***	.087***	.018	−.001	.048***
	(.017)	(.018)	(.018)	(.018)	(.007)	(.014)
Median grad age	.068***	.058***	.073***	.007	.007	.010
	(.015)	(.018)	(.019)	(.015)	(.005)	(.014)
Incl. early leavers	.061***	.065***	.097***	.013	−.006	.031**
	(.018)	(.020)	(.020)	(.018)	(.006)	(.015)

Notes: Each cell reports an estimate from a separate regression weighted by adjusted sampling weights that give same weight to each country, except for the second row, where we forgo using any weight. Only the estimate of the central exam effect based on variants of Equation 1 is reported. Dependent variables are globally standardized test scores in numeracy (column 1), literacy (column 2), problem-solving (column 3), log gross hourly wage (column 4), an employment dummy (column 5), and a dummy for tertiary education degree (column 6). "Incl. early leavers" row additionally includes individuals who, based on the calculatory age of graduation (highest attained degree), is below 15, i.e. earlier than the compulsory leaving age in OECD countries. Robust standard errors clustered by graduation year × country in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 5: Piecewise deletion of countries

	Numeracy (1)	Earnings (2)		(1)	(2)
Baseline	.067***	.008		cont.	cont.
	(.018)	(.019)			
Excluded country:					
Austria	.067***	.009	Belgium	.066***	.009
	(.018)	(.019)		(.017)	(.019)
Canada	.067***	.008	Chile	.066***	.009
	(.017)	(.019)		(.017)	(.019)
Cyprus	.060***	.007	Czech Republic	.071***	.010
	(.019)	(.019)		(.018)	(.019)
Denmark	.066***	.009	Estonia	.059***	.011
	(.017)	(.019)		(.019)	(.021)
Finland	.066***	.009	France	.067***	.009
	(.017)	(.019)		(.017)	(.019)
Germany	.067***	.007	Greece	.127***	.035*
	(.018)	(.019)		(.018)	(.020)
Ireland	.067***	.007	Israel	.065***	.008
	(.018)	(.019)		(.017)	(.019)
Italy	.066***	.009	Japan	.070***	-.011
	(.018)	(.019)		(.019)	(.020)
Korea	.070***	.002	Lithuania	.053***	.004
	(.019)	(.020)		(.018)	(.021)
Netherlands	.071***	.013	New Zealand	.067***	.009
	(.019)	(.020)		(.017)	(.019)
Norway	.064***	.008	Poland	.055***	.014
	(.017)	(.019)		(.018)	(.020)
Singapore	.068***	.007	Slovak Republic	.051***	.008
	(.017)	(.019)		(.018)	(.020)
Slovenia	.071***	.012	Spain	.065***	.010
	(.019)	(.021)		(.017)	(.019)
Sweden	.069***	.002	Turkey	.067***	.008
	(.018)	(.020)		(.018)	(.019)
United Kingdom	.068***	.007	United States	.067***	.008
	(.017)	(.019)		(.018)	(.019)
Excluded country groups:					
Japan and Korea	.075***	-.020	Federal education	.067***	.007
	(.020)	(.021)	systems	(.018)	(.019)

Notes: Each estimate stems from a separate regression excluding one or more countries. The regressions are weighted by adjusted sampling weights that give same weight to each country. We report here only the estimates of the central exam effect on numeracy (column (1)), which corresponds to the specification of Table 2, column (4), and the effect on earnings (column (2)), which corresponds to the specification of Table 3, column (2). Dependent variable is the globally standardized test scores in numeracy, or log gross hourly wage, respectively. Robust standard errors clustered by graduation year $\times$ country in parentheses.
* p<0.10, ** p<0.05, *** p<0.01

Table 6: Central exams and adult outcomes in the German subsample

Outcome:	Adult skills			Labor market outcomes		
	Numeracy (1)	Literacy (2)	Prob. solv. (3)	Earnings (4)	Empl. (5)	College deg. (6)
Central exams	.112***	.100***	.056*	−.005	.053***	.017
	(.030)	(.032)	(.033)	(.023)	(.016)	(.013)
Backgr. char.	yes	yes	yes	yes	yes	yes
Grad. age FE	yes	yes	yes	yes	yes	yes
Grad. year FE	yes	yes	yes	yes	yes	yes
Observations	4,086	4,086	3,646	2,098	2,949	3,250
Clusters	570	570	555	449	499	543
R^2	.288	.279	.263	.255	.168	.506

Notes: Dependent variables: for numeracy, literacy and problem solving skills we use German-wide standardized scores. For earnings we use log gross hourly wage, employment (Empl.) is a 0/1 variable, where 1 means currently employed. The College deg. equals 1 if an individual completed any college. The College deg. estimation excludes those currently studying, the Earnings and Employed estimations additionally exclude those younger than 25. The set of controls used for the skill outcomes (column 1-3) is identical to our preferred specification of Table 2 (column 4), the set of controls used for labor market outcomes (column 4-6) is identical to Table 3. Estimations are weighted by sampling weights. Robust standard errors (in parentheses) clustered at graduation year × federal state. * p<0.10, ** p<0.05, *** p<0.01

Table 7: Subgroup analysis

Outcome	Adult skills			Labor market and education		
	Numeracy (1)	Literacy (2)	Prob. solv. (3)	Earnings (4)	Empl. (5)	College deg. (6)
Baseline:						
Central exams	.067***	.068***	.091***	.008	.014	.014
	(.018)	(.020)	(.020)	(.019)	(.011)	(.011)
With interactions by parental education:						
Central exams	.052***	.036*	.114***	.051**	.017	.018
	(.020)	(.021)	(.024)	(.021)	(.013)	(.012)
Secondary	.011	.037**	−.040**	−.053***	−.007	.001
× Central exams	(.015)	(.015)	(.019)	(.014)	(.009)	(.008)
Tertiary	.037*	.065***	−.025	−.094***	−.003	−.027**
× Central exams	(.019)	(.019)	(.025)	(.017)	(.011)	(.011)
With interactions by gender:						
Central exams	.082***	.073***	.095***	−.013	.006	.003
	(.019)	(.021)	(.022)	(.021)	(.012)	(.012)
Female	−.031***	−.008	−.009	.048***	.017*	.022***
× Central exams	(.012)	(.011)	(.014)	(.013)	(.009)	(.007)
With interactions by migration background:						
Central exams	.064***	.066***	.086***	.006	.013	.013
	(.018)	(.020)	(.020)	(.019)	(.011)	(.011)
Second-gen. migrant	.045**	.030	.063**	.027	.018	.009
× Central exams	(.022)	(.022)	(.027)	(.020)	(.013)	(.011)

Notes: Each cell reports an estimate from a separate regression for the outcome measure weighted by adjusted sampling weights that give same weight to each country. Only the estimate of the central exam effect, which corresponds to the specification of Table 2, column (4), is reported. Dependent variables are globally standardized test scores in numeracy (column 1), literacy (column 2), problem-solving (column 3), log gross hourly wage (column 4), an employment dummy (column 5), and an indicator for tertiary education degree (column 6). Robust standard errors clustered by graduation year × country in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 8: Time to effect

Outcome	Numeracy skills		
	(1)	(2)	(3)
Central exams	.067***	.023	.022
	(.018)	(.024)	(.024)
Central exams 2+		.058**	.066**
		(.026)	(.032)
Central exams 5+			−.012
			(.031)
Countries	27	27	27
Observations	131,398	131,398	131,398
Clusters	1,361	1,361	1,361
R^2	.265	.265	.265

Notes: Estimates in column (1) correspond to our baseline specification, of Table 2, column (4) for our sample without the three countries with federal education systems, i.e. Canada, Germany, and the United States. In column (2), an additional effect of "2 years after the introduction" is amended (Central exams 2+), additionally in column (3) an effect of "5 years after the introduction" is included (Central exams 5+). All estimates are weighted by adjusted sampling weights that give same weight to each country. Robust standard errors clustered by graduation year × country in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 9: Comparison with correlations in PISA and TIMSS

	PIAAC (1)	PISA (2)	PIAAC (3)	TIMSS (4)
Central exams	.074	.175	.262	.361
	(.140)	(.124)	(.244)	(.290)
Countries	18	18	13	13
Observations	1,899	46,856	1,067	51,957
R^2	.018	.013	.033	.036

Notes: Dependent variable: globally standardized numeracy score for PIAAC and standardized math scores for PISA and TIMSS. PISA data refers to the PISA 2000 wave. TIMSS data refers to the 1999 TIMSS-Repeat study. The estimations in the PIAAC columns are based on a sample of PIAAC individuals who were 15 years old in 2000. The estimations in all columns except (1) are limited to the selection of countries which are common to PIAAC/PISA and PIAAC/TIMSS. All estimations are weighted by adjusted weights, where each country's weights sum up to 1. Robust standard errors (in parentheses) clustered by country. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A1: Introduction of central exit examinations (CEE) at the upper secondary level (ISCED 3) across PIAAC countries

Country	Year of CEE-Introduction	Source
Austria	2016	Ministerium für Bildung und Frauen Österreich (BMBF):
		https://www.bmbf.gv.at/schulen/unterricht/ba/reifepruefung.html
Belgium	-	Hörner et al. (2007), Centre for Educational research, London School of Economics and Political Science:
		http://www.leeds.ac.uk/educol/documents/00001195.htm
Canada⋆	1929-1984	General Accounting Office, Washington, 1993:
		http://files.eric.ed.gov/fulltext/ED361377.pdf,
		https://archive.org/details/ERIC_ED361377
Chile	-	OECD (2010)
Cyprus	2006	Lamprianou (2012), Hörner et al. (2007),
		Ministry of Education Cyprus: http://www.highereducation.ac.cy/en/examinations.html
Czech Republic	2011	Greger and Kifer (2012), expert based evaluation
Denmark	1908	Hörner et al. (2007), Bishop (1999)
Estonia	1997	Education fact sheets on http://archimedes.ee,
		http://www.innove.ee/en/general-education/final-examinations/secondary-school
Finland	1919	Matriculation Examination Board Finland:
		https://www.ylioppilastutkinto.fi/fi/english

Notes: ⋆Countries with federal school systems. CEE introduction years differ across federal states. See Table A2 for more detailed information.
The term "personal communication/expert based evaluation" refers to information provided by experts of the respective national education system. This comprises qualitative interviews or e-mail communication with national centers responsible for the execution and evaluation of central exams.

Table A1: *Continued*

Country	Year of CEE-Introduction	Source
France	1808	Hörner et al. (2007), Ministry of Education France:
		http://www.education.gouv.fr/cid60987/bac-2015-questions-reponses.html
Germany⋆	1945-2007	Lüdemann (2011), Klein et al. (2009)
Greece	1983, abolished in 2005	Goulas and Megalokonomou (2015), Polydorides (1986), expert based evaluation
Ireland	1960	State examination Commission Ireland:
		https://www.examinations.ie/?l=en&mc=li&sc=li
		Department of Education and Skills:
		http://www.education.ie/en/The-Education-System/Post-Primary/
Italy	1923	Nardi (2001), expert based evaluation
Israel	1923	Yemini et al. (2014), Lavy (2008)
Japan	1979	Eckstein and Noah (1989), Watanabe (2013) expert based evaluation
Lithuania	2002	Bethell and Zabulionis (2000), OECD (2002)
Korea	1994	Chang (2009)
		Korea Institute of Curriculum and Evaluation:
		http://www.kice.re.kr/sub/info.do?m=0205&s=english
The Netherlands	1968	Hörner et al. (2007), Boezerooy and Huisman (2000), Government of the Netherlands:
		https://www.government.nl/topics/secondary-education/contents/secondary-school-leaving-examination
New Zealand	ever since	Hall (2000), New Zealands Qualifications Authority:
		http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/history-of-ncea/, http://www.nzqa.govt.nz/qualifications-standards/results-2/secondary-school-qualifications-prior-to-2002/
Norway	1960	Hörner et al. (2007), Organization for Internationalization in Education:
		https://www.epnuffic.nl/en/publications/find-a-publication/education-system-norway.pdf

Notes: ⋆Countries with federal school systems. CEE introduction years differ across federal states. See Table A2 for more detailed information.
Japan and Korea have national university entrance exams that consider to be equivalent to a central exam in our main specification. The term "personal communication/expert based evaluation" refers to information provided by experts of the respective national education system. This comprises qualitative interviews or e-mail communication with national centers responsible for the execution and evaluation of central exams.
Israel introduced its system of central exams "Bagrut" in 1923, during the British Mandate of Palestine (1920-1948).

Table A1: *Continued*

Country	Year of CEE-Introduction	Source
Poland	2005	The Central Examination Commission:
		http://www.cke.edu.pl/,
		http://www.cke.edu.pl/images/stories/English/the_matura_exam.pdf
Singapore	1975	Statement Singapore Examinations and Assessment Board:
		http://www.seab.gov.sg/pages/nationalExaminations/GAL/general.asp
		Bishop (1997), Lim and Tan (1999)
Slovenia	1994	Ilc et al. (2014)
Spain	-	Bishop (1999), Fuchs and Wößmann (2008), Hörner et al. (2007)
Slovak Republik	2005	National Institute for Certified Educational Measurement:
		http://www.nucem.sk/en/maturita
		http://www.nucem.sk/en/maturita#3/brief-history-of-maturita
Sweden	eliminated in 1968	Bishop (1999), Hörner et al. (2007)
Turkey	-	Yildirim et al. (2007), Measuring, Selection and Placement Center Turkey: http://www.osym.gov.tr/belge/1-2706/tarihsel-gelisme.html
United Kingdom	in 1950's	http://www.a-levels.co.uk/history-of-a-levels.html
United States*	2003	Caves and Balestra (2018)

Notes: *Countries with federal school systems. CEE introduction years differ across federal states. See Table A2 for more detailed information.

Table A2: Central exit examinations at the upper secondary level (ISCED 3) in countries with federal school systems

Country	Federal state	Year of CEE-Introduction
Canada	Alberta	1984
	British Columbia	1984
	Manitoba	eliminated 1970/reintroduced 1991
	Newfoundland/Labrador	1974
	New Brunswick (francophone)	1991
	New Brunswick (anglophone)	-
	Northwest Territories	-
	Nova Scotia	eliminated 1972
	Nunavut	-
	Ontario	eliminated 1967
	Prince Edward Island	-
	Quebec	1929
	Saskatchewan	-
	Yukon	-
Germany[a]	Baden-Württemberg (BW)	1952
	Bavaria (BY)	1946
	Berlin (BE)	2006
	Brandenburg (BB)	2005
	Bremen (HB)	2007
	Hamburg (HH)	2005
	Hesse (HE)	2007
	Mecklenburg-Western Pommerania (MV)	1991
	Lower Saxony (LS)	2006
	North-Rhine Westphalia (NW)	2007
	Rhineland Palantine (RP)	-
	Saarland (SL)	1945
	Saxony (SN)	1993
	Saxony-Anhalt (ST)	1993
	Schleswig-Holstein (SH)	2008
	Thuringia (TH)	1990
USA[b]	New York	2002

Notes: [a]Federal states in the former German Democratic Republic (GDR) were de facto central exam states, before 1990, as in the GDR there was a central examination regime in place (Kühn 2010). Years listed in the table indicate the introduction of central exams after the German reunification. Further, we only refer to central exams at the end of secondary school. Central exams in low school tracks are not considered. According to Lüdemann (2011), time of introduction in the highest track was timely close to the introduction of central exams in the lower school tracks in Germany.
[b]In the USA, only the federal state of New York has a centralized school leaving exam (Regents Examination) that meets our definition of a central exit exam.

Table A3: Coding of central exams in Fuchs and Wößmann (2008) compared to
our coding across PIAAC/PISA and PIAAC/TIMMS countries

Country	F&W	Our	Dataset
Austria	0	0	PISA
Belgium	0	0	PISA/TIMMS
Canada	0.5	0.537	PISA/TIMMS
Chile	-	0	TIMMS
Cyprus	-	0	TIMMS
Czech Republic	1	0	PISA/TIMMS
Denmark	1	1	PISA
Finland	1	1	PISA/TIMMS
France	1	1	PISA
Germany	0.4	0.424	PISA
Greece	0	1	PISA
Ireland	1	1	PISA
Israel	-	1	TIMMS
Italy	1	1	PISA/TIMMS
Japan	1	1	PISA/TIMMS
Korea	1	1	PISA/TIMMS
Lithuania	-	0	TIMMS
Netherlands	1	1	PISA/TIMMS
New Zealand	1	1	PISA/TIMMS
Norway	1	1	PISA
Poland	1	0	PISA
Singapore	-	1	TIMMS
Slovak Republic	-	0	TIMMS
Slovenia	-	1	TIMMS
Spain	0	0	PISA
Sweden	0.5	0	PISA
Turkey	-	0	TIMMS
United Kingdom	1	1	PISA/TIMMS
United States	0.1	0.067	PISA/TIMMS