Halle Institute for Economic Research
Member of the Leibniz Association

Discussion Papers

No. 13

May 2023

# Where Do STEM Graduates Stem from? The Intergenerational Transmission of Comparative Skill Advantages

Eric A. Hanushek, Babs Jacobs, Guido Schwerdt,
Rolf van der Velden, Stan Vermeulen, Simon Wiederhold

## Authors

**Eric A. Hanushek**
Hoover Institution, Stanford University, CESifo, and NBER
E-mail: hanushek@stanford.edu

**Babs Jacobs**
Research Centre for Education and the Labour Market (ROA) at Maastricht University
E-mail: bpja.jacobs@maastrichtuniversity.nl

**Guido Schwerdt**
University of Konstanz, ifo Institute, CESifo, and IZA
E-mail: guido.schwerdt@uni-konstanz.de

**Rolf van der Velden**
Research Centre for Education and the Labour Market (ROA) at Maastricht University
E-mail: r.vandervelden@maastrichtuniversity.nl

**Stan Vermeulen**
Research Centre for Education and the Labour Market (ROA) at Maastricht University
E-mail: c.vermeulen@maastrichtuniversity.nl

**Simon Wiederhold**
Halle Institute for Economic Research (IWH) – Member of the Leibniz Association, Department of Structural Change and Productivity, Martin Luther University Halle-Wittenberg, ifo Institute Munich, and CESifo
E-mail: simon.wiederhold@iwh-halle.de
Tel +49 345 7753 840



Comments and suggestions on the methods and results presented are welcome.

IWH Discussion Papers are indexed in RePEc-EconPapers and in ECONIS.

## Editor

Halle Institute for Economic Research (IWH) – Member of the Leibniz Association

Address: Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany
Postal Address: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

# Where Do STEM Graduates Stem from? The Intergenerational Transmission of Comparative Skill Advantages*

## Abstract

The standard economic model of occupational choice following a basic Roy model emphasizes individual selection and comparative advantage, but the sources of comparative advantage are not well understood. We employ a unique combination of Dutch survey and registry data that links math and language skills across generations and permits analysis of the intergenerational transmission of comparative skill advantages. Exploiting within-family between-subject variation in skills, we show that comparative advantages in math of parents are significantly linked to those of their children. A causal interpretation follows from a novel IV estimation that isolates variation in parent skill advantages due to their teacher and classroom peer quality. Finally, we show the strong influence of family skill transmission on children's choices of STEM fields.

*Keywords: causality, intergenerational mobility, parent-child skill transmission, STEM*

*JEL classification: I24, I26, J12, J24, J62*

* Acknowledgements: This paper is part of the Intergenerational Transmission of Skills (ITS) project carried out at the Research Centre for Education and the Labour Market (ROA: https://www.roa.nl/research/research-projects/intergenerational-transmission-skills-its-research-project). We gratefully acknowledge a grant received from the Dutch Ministry of Education, Culture and Science and the Netherlands Initiative for Education Research (NRO: grant 405-17-900). Schwerdt acknowledges that his research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2035/1 – 390681379. Helpful comments on earlier drafts were provided by David Figlio, Edwin Leuven, Mikael Lindahl, Lance Lochner, Markus Nagler, Gary Solon, Ludger Woessmann, Ulf Zoelitz, and participants at conferences and seminars in Aarhus, Lund, Maastricht, Milan, Minneapolis, Munich, Padova, and Zurich.

# 1. Introduction

*The pace of innovation is accelerating globally, and with it the competition for scientific and technical talent. Now more than ever the innovation capacity of the United States—and its prosperity and security—depends on an effective and inclusive STEM education ecosystem.*

Committee On Stem Education (2018), p. v

Advancing STEM education is a policy objective not only in the U.S. but also in many developed countries, and this quite commonly leads to a government focus on ensuring that schools produce sufficient math skills to support STEM careers. This policy focus, however, generally ignores the fundamental role of comparative advantage in occupational choice. The key foundation of occupational choice as developed by Roy (1951) has been challenging to implement empirically. Economists have not been entirely successful in adding the underlying structural detail that supports notions of comparative advantage empirically. Specifically, the important issues of the underlying source of differences in comparative advantages across individuals and of the malleability of these advantages remain largely unanswered.

Comparative advantage based on differential skills has, for example, been used to explain why some choose college attendance and others do not (Willis and Rosen (1979)), how selection affects wages in different economic sectors (Heckman and Sedlacek (1985)), and why college students choose different fields of study (Kirkeboen, Leuven, and Mogstad (2016)).[1] With some exceptions, these analyses consider ex post realizations of outcomes and then address underlying selection issues that play into these realizations. The inferences are based on a combination of observed ability differences and model-based interpretations of the empirical results. Invariably, however, the conclusions about the role of comparative advantage are based on assumptions of fixed and exogenous ability differences. As such, they provide little basis for considering what causes differences in comparative advantage. Importantly, it also precludes consideration of what

[1] As indicated in the review by Altonji, Arcidiacono, and Maurel (2016), comparative advantage is not the only approach to modeling choice of fields of study. In the analysis of fields of study and in the more general labor market analyses, a frequent alternative is a single-factor model, implicitly built on absolute advantage in one dimension of ability. Nonetheless, as shown in the cited examples, this tends to be rejected in models where the empirical strategy incorporates various approaches to comparative advantage.

policies might be employed if, for example, a country wants to increase the prevalence of STEM-trained workers.

Our analysis focuses on the ability differences that are assumed fixed in the outcome-based analyses of schooling and labor market realizations. This paper directly addresses the measurement and source of comparative advantages based on cognitive skill differences within individuals. While analyses of comparative advantage from a labor market perspective have focused on various skill differences such as brain v. brawn or cognitive v. noncognitive skills, we focus on different dimensions of cognitive skills – math v. language. We show that these dimensions are separately important for future economic outcomes and that comparative advantage across these dimensions is key to the academic choices leading to STEM careers.

The core analysis shows that the parental comparative advantage in math (vs. language) skill is a strong determinant of children's comparative skill advantage. Importantly, this intergenerational transmission of comparative math skills is malleable through the education system.

The analysis builds on a unique data set that provides comparable measures of different domains of cognitive skills for parents and their children. Our data come from linking extensive Dutch survey data on parent skills in math and language around the end of primary education to register data on their children's skills in the same subjects elicited on similar tests at a similar age.[2] The parental survey data cover three cohorts of parents sampled when they were students in the first year of secondary education (1977 and 1989) or the last year of primary education (1982). The surveys are nationally representative covering 8–15 percent of all students entering Dutch secondary education. In total, the combined dataset includes more than 25,000 parents and 40,000 of their children.

Our assessments of parental and child skills all occur around the end of primary school, leading us to begin by ensuring that the early test scores represent skills that have long-term economic value. By linking the parental survey data to the parents' adult education and labor market outcomes, we can demonstrate the economic importance of our early-life measures of cognitive skills. Consistent with prior work (e.g., Chetty, Friedman, and Rockoff (2014), Aucejo

[2] The linkage to registry data minimizes the problems of sample attrition that plague attempts to investigate intergenerational linkages with survey-based panels (e.g., Brown, McIntosh, and Taylor (2011); de Coulon, Meschi, and Vignoles (2011)).

and James (2021)), skills in math and language measured early in the education system prove to be highly correlated with later education choices, hourly wages, income, and wealth. But, importantly, we also show that the separate dimensions of math and language skills, while often unobserved or ignored in more general labor market studies, are independently significant in explaining economic outcomes some three decades later. These results support our focus on the intergenerational transmission of *comparative* skills as measured by early test scores.

To investigate the sources of comparative advantage that has been central to many economic investigations, we study how comparative skill advantages are transmitted from parents to children. To do so, we begin with a composite conceptual model that combines the Galton-inspired intergenerational transmission model with an educational production function that considers how various inputs affect the cognitive skills of children. Empirically, we exploit within-family between-subject variation in cognitive skills, asking how differences in parents' skills between math and language relate to differences in math and language skills of their children. In this analysis, all observed and unobserved influences of family, school, and neighborhoods that do not differentially affect the two skill domains are eliminated.

We find that parents with a comparative advantage in math skills are significantly more likely to have children with a similar math skill advantage. In terms of magnitude, a difference of 10 percentile ranks between skills in math and language in the parent generation translates into a one-rank difference in the child generation. The strength of transmission remains virtually unchanged when we allow for the possibility that various grandparent characteristics (i.e., education and occupational status) and detailed regional factors, all measured at the time of parents' skill assessment, affect math and language skills differently.

To go deeper both into the source of comparative skill advantage and into its malleability, we employ a novel instrumental variable (IV) estimation. We exploit differences between math and language skills of the parents' classroom peers to isolate variation in parents' comparative skill advantage developed outside the family. These differences in comparative skill advantages of peers reflect differences in the subject-specific quality of the early formal education environment of parents. Our IV estimates indicate clearly that nonfamily inputs in the production of skills affect comparative skill advantages that then carry over to future generations. This would not be the case if the observed skill transmission patterns just reflected innate differences in talent (e.g., a "math gene") or dynastic predispositions for specific subjects (e.g., arising through occupational legacies).[3] Overall, our IV results indicate that any policy that shifts focus from one skill domain to another not only affects the comparative skill advantage of the current students but also has lasting impacts on subsequent generations.

Given our findings on the sources of comparative skill advantages, we return to how parents influence the long-run path of children. In particular, although academic and policy attention has focused on increasing the number of individuals entering STEM fields of study and occupations (e.g., UNESCO (2017), Stoet and Geary (2018)), the role of families in influencing STEM choices has received little attention (see, for example, the review in Altonji, Arcidiacono, and Maurel (2016)).[4] From the registry data for children, we observe patterns of course taking in secondary schooling and of choice of field of study in post-secondary or tertiary education.[5] We show that children of parents with relatively higher math skills are more likely to choose STEM fields both at school and after school. Put differently, parents with comparative advantage in math (language) produce children who opt for STEM (non-STEM) fields, just as would be suggested by a simple Roy model of occupational choice. Parents influence the comparative skill advantages of both boy and girl offspring with no gender bias, leading to similar course patterns in secondary school. But ultimately comparative skill advantages have less influence on girls' choices of STEM field of study than on boys' choices, potentially contributing to the frequently-observed lower participation in STEM fields by girls.

Our results contribute to five strands of prior literature. First, we add evidence on the sources of comparative advantage to the well-established theoretical literature on the importance of comparative advantage in the labor market (e.g., Roy (1951), Lazear (2009), Acemoglu and Autor (2011), Eisenhauer, Heckman, and Vytlacil (2015)).[6]

[3] In addition to the substantive interpretation of the IV estimates, they provide a correction for any measurement error in the comparative skill advantages. Measurement error concerns are further addressed by a series of alternative corrections of such error.

[4] An exception is Altmejd (2023), who considers the intergenerational transmission of field of study in Sweden. While his design allows identification of how parental choices of fields of study lead to those of children, it does not consider the underlying sources of parental or child choices beyond the familial consistency.

[5] Note that there is also other recent analysis of course taking in Dahl, Rooth, and Stenberg (2023).

[6] The idea of comparative advantage has also been deeply embedded in a range of studies of other individual choice behavior, such as educational investment decisions (Willis and Rosen (1979)), immigration decisions (Borjas (1987)), the division of labor within households (Becker (1981)), and social interactions (Cicala, Fryer, and Spenkuch (2018)).

Second, we broaden the perspective of the large literature on the intergenerational transmission of human capital by providing the first evidence on how comparative advantages in cognitive skills are transmitted from parents to children. This literature has made important advances in understanding overall influences of families (e.g., Black and Devereux (2011), Adermon, Lindahl, and Palme (2021)) but has stopped short of addressing the important role of multiple skill dimensions. Our IV results also speak to the nature-nurture debate by showing that early comparative skill advantages do not just arise from genetic configurations but are shaped by pre-birth factors outside the family.[7]

Third, we directly insert the idea of comparative skill advantages into the growing literature on labor market returns to skills. Several recent studies suggest substantial wage returns to tested numeracy and literacy skills (e.g., Hanushek et al. (2015, 2017)), but they typically treat alternative tests as separate measures of a common cognitive factor. Other research emphasizes the economic importance of specific skills, such as social skills (e.g., Deming (2017), Piopiunik, Schwerdt, Simon, and Woessmann (2020)), digital skills (e.g., Falck, Heimisch-Roecker, and Wiederhold (2021), Kiener, Gnehm, Clematide, and Backes-Gellner (2022)), or technical skills (Barrera-Osorio, Kugler, and Silliman (2020)).[8] However, this literature either considers these skills in isolation or estimates returns as a horserace between different skill domains without recognizing the fundamental role of comparative advantage.

Fourth, by adding findings about the underlying structure of skill production, we provide new information for the persistent debates on STEM education. We show that comparative skill advantage significantly enters into STEM field preparation and choices. In a reduced-form analysis across generations, we also show that the comparative skills of parents are directly related to children's STEM education patterns. These results imply that changes in relative skills to today's generation, whether related to policy or otherwise, have additional ramifications for future generations.

[7] A variety of prior papers consider identifying under varying assumptions the effects of various pre-birth components and environmental components (e.g., Björklund, Lindahl, and Plug (2006), Sacerdote (2011), Lundborg, Plug, and Rasmussen (2021)) and into the direct influence of genetics (Houmark, Ronda, and Rosholm (2020)).

[8] A complete description of individuals' early-career human capital is provided by Langer and Wiederhold (2023 ), who consider all skills developed through the German apprenticeship system. Aggregating more than 13,000 different skills to six broad skill categories, they show that cognitive, social, and digital skills have higher returns than manual, management, or administrative skills.

Fifth, we contribute to the methodological discussion about how to measure relevant cognitive skills. We focus throughout on the ordinal properties of the math and language assessments by analyzing child and parent skills as percentile ranks in the overall skill distributions. This guards against concerns about assuming cardinal properties for standard assessments as found in most economic analyses that include test scores (Ho and Reardon (2012), Bond and Lang (2013), Nielsen (2015)). The results are nevertheless robust to the more conventional analysis of scale scores.

The remainder of the paper is structured as follows: The next section describes our data and the Dutch institutional background. Section 3 documents the differential predictive power of our early measures of math and language skills for parents' lifetime outcomes. In section 4, we outline a conceptual framework describing how we think about the production of comparative skill advantages and their intergenerational transmission. Section 5 shows how we implement this framework empirically. In section 6, we present our results on intergenerational transmission of comparative skill advantages. Section 7 shows that parents' comparative skill advantages affect children's actual STEM choices. Section 8 concludes.

## 2. Data and Institutional Background

### 2.1 The Dutch Education System

The Dutch education system is an early stratifying system (Bol and van de Werfhorst (2013)), where students are allocated to different tracks (low, middle, or high) after primary education (grade 6, at age 12). This allocation is largely based on the performance of students on a national test at the end of primary education, the CITO (Central Institute for Test Development (CITO)) test.[9]

The CITO test is a national high-stakes test measuring school performance in math and language (along with other subjects).[10] This test, first employed in 1970, was introduced to

[9] The other component that determines track allocation is the primary school teacher's advice, which is partly based on the objective results of the CITO test, and partly on the teacher's subjective expectations of students' success in secondary education.

[10] Before the 2014/15 school year, participation in the national test was not mandatory. However, around 85 percent of the schools in primary education have participated in the CITO test since its introduction. From 2014/2015 onwards, it is compulsory for students in grade 6 to take a final test. The government makes the CITO test available to all schools. Even though schools can also choose another final test approved by the Ministry of Education, most schools participate in the CITO test (Jacobs, van der Velden, and van Vugt (forthcoming)).

ensure an objective, merit-based assignment to different tracks in subsequent schooling. The testing is done over a three-day period in spring of the final year of primary schooling. The test involves multiple choice items and is centrally scored.

After having been in secondary school for two years (for students attending the low track) or three years (for students attending the middle or high track), students have to decide on a course profile that will determine the type of courses they can take in upper-secondary or tertiary education.[11] After finishing secondary school, students can choose, depending on their track in secondary education, to enter upper secondary vocational education, tertiary vocational education, or university. They can also directly enter the labor force without additional schooling.

### 2.2 The Intergenerational Transmission of Skills (ITS) Database

For this paper, we developed the Intergenerational Transmission of Skills (ITS) database, which provides CITO test scores for parents and their children. This database was constructed to be the foundation of an extensive research program on the intergenerational transmission of cognitive skills (Jacobs, Vermeulen, and van der Velden (2021)).[12]

The ITS dataset combines extensive survey data gathered for three cohorts of students in the 1970's and 1980's and linked to more recent register data on their children available at Statistics Netherlands. The survey data contain cognitive skill measures of the parent generation along with other descriptive information about the families. The register data contain cognitive skill measures of the children's generation as well as other information on their secondary schooling. Two cohorts of parents were sampled in the first year of secondary education (1977 and 1989),

[11] In the low track (called in Dutch 'VMBO'), students can choose between four profiles: Technical, Agriculture, Economics, and Health & Welfare, or a combination thereof. In the middle and high tracks (called in Dutch 'HAVO' and 'VWO', respectively), students can choose between Nature & Technical, Nature & Health, Economics & Society, Culture & Society, or a combination thereof.

[12] For more information on this research program and details of the construction of this database, see https://www.roa.nl/research/research-projects/intergenerational-transmission-skills-its-research-project. The inaugural papers in this project were Jacobs and van der Velden (2021) and our initial investigation of comparative cognitive skills, Hanushek et al. (2021). Jacobs and van der Velden (2021) estimate structural equation models to investigate the relative contribution of three mechanisms that underlie the intergenerational transmission of education from parents to children: human capital, cultural capital, and financial capital. Our previous analysis considered comparative cognitive skills in a different context and did not see the implications of comparative skills for testing the Roy model and for addressing the STEM policy debates.

and one cohort was sampled in the last year of primary education (1982).[13] Each of these longitudinal surveys is a nationally representative panel of students: in the 1977 cohort, 37,280 students from 1,275 schools participated (15 percent of the student population at that time); in the 1982 cohort, 16,813 student from 669 schools participated (8 percent of the student population); and in the 1989 cohort, 19,524 students from 381 schools participated (10.5 percent of the student population).

Individual classrooms were selected within sampled schools, and all students in that classroom were surveyed. The math and language skills of the surveyed cohorts were assessed during the school year using a shortened version of the CITO test.[14] In addition, background information on their parents (the grandparent generation in our analysis) such as their highest level of education, socio-economic status, and number of children living at home was collected. After the initial survey and assessment, individuals were followed annually over the course of their school career until leaving education. For most students in the original cohorts, basic identifying information is available including name and address at the time of the survey, allowing us to link these cohort data to register data from Statistics Netherlands. The data could be linked successfully in 80 to nearly 100 percent of cases, depending on the cohort (1977 cohort: 81 percent; 1982 cohort: 88 percent; 1989 cohort: 98 percent, due to the availability of a unique personal identifier). Unless both parents participated in one of the three surveys, we have one parent in each matched family.[15]

The combined dataset contains information on the math and language skills of 25,483 parents and 41,774 of their children. The sample sizes and average skills of parents and children differ by cohort, as can be seen in Table 1. The sample size differences across cohorts partly

[13] In the 1977 and 1989 cohort, parent cognitive skills were tested after tracking. Our results are robust to including controls for the school track attended and also hold within each cohort (see below), implying that they are not simply driven by track effects.

[14] Note that surveyed students took the full CITO tests for placement purposes, but the surveys were given at different times during the year and the official CITO scores were not linked to the surveys. In the 1977 and 1982 cohorts, the survey tests were taken at the start of the school year. In the 1989 cohort, students took the test 5–7 months after the start of the school year, during the first months of the 1990 calendar year.

[15] The fact that we usually observe the cognitive skills of only one of the parents in the ITS data potentially induces measurement error in the parent skill variables. To address this, we make use of the fact that we observe both parents for 365 children in our data. When randomly dropping one of the parents and estimating the relationship between child and parent skills, results are very similar as in the two-parent sample (see Figure A1). This indicates that our main findings are unlikely to be affected by just having skill information for one of the parents in most of our data.

reflect the window for observed test-taking by children. Statistics Netherlands has register data of all schools that participated in the CITO test from school year 2005/2006 onwards. Because of COVID-19, our observation window concludes at the end of the 2018/2019 school year.[16] Thus, we only observe those parents whose children took the CITO test at the end of primary school between 2006 and 2019.[17] This implies that for the 1977 cohort, we observe parents who are relatively old when they had children, while for the 1989 cohort we observe relatively young parents.[18] The selectivity of our sample with respect to age also has implications for parent education and skills. Because more highly educated people tend to enter parenthood at a later age, the parents from the 1977 cohort whose children we can observe in our data are positively selected in terms of their education and skills. The parents from the third cohort entered parenthood relatively young and therefore tend to have slightly lower educational attainment and skills, while the parents from the second cohort (around age 12 in 1982) fall somewhere in between. However, since our main estimation model relies on variation in cognitive skills within-parent between-subjects and because our results hold in each individual cohort, this sample selectivity has no major implications for our results.

Data on grandparent education, which we derive from the parent questionnaire in the original cohort studies, provide additional information about the long-run transmission of skills (e.g., Adermon, Lindahl, and Palme (2021)). In Table 1, we again observe that our parent subsample in the 1977 cohort is positively selected, with a relatively high share of tertiary educated grandparents.  However, there are no glaring differences by cohort in the social background of grandparents, measured by the type of occupation that the main breadwinner in the household held when parents took the skill test.

In addition to test scores, the registry data also provide detailed information on children's educational careers, allowing us to observe children's STEM choices at school. These in-school choices have important long-term consequences, as enrollment into most upper-secondary or tertiary education programs is only possible with specific backgrounds in terms of courses taken. We also observe STEM choices in upper secondary vocational or tertiary education directly. We

[16] The CITO test was not taken in the COVID-19-year 2019/2020.

[17] At the time of test taking, 91.8 percent of children live in the same household as the parent whose cognitive skills we observe.

[18] In the year of birth of the children, the parents were on average 31.7 years old (33.6 years in the 1977 cohort, 30.7 years in the 1982 cohort, and 27.0 years in the 1989 cohort).

separately code outcomes as either STEM or non-STEM based on the type of courses taken at school and the subsequent field of study. We observe that 34 percent of children choose a STEM profile at school, while 23 percent study a STEM field in upper secondary vocational or tertiary education (Table 1).[19]

### 2.3 Measuring Comparative Skill Advantage

Based on the test score data in math and language, we construct a straightforward measure of comparative skill advantage for each individual. Test scores of children in a subject are measured in percentile ranks within each test year based on the universe of test data from administrative records.[20] Parent test scores in each subject are measured in percentile ranks within each cohort, using the complete survey data (i.e., including parents and the unmatched survey-takers). Within each generation, we interpret the difference between the percentile ranks in math and language as measuring comparative skill advantage.[21]

Our measure of comparative skill advantage does not permit an absolute interpretation as there is no natural metric that would allow measurement of levels of math and language skills on the same scale. We define the comparative skill advantage in relative terms by anchoring the skills of an individual in each subject to the distribution of the entire population participating in the same skill assessment.

Figure 1 provides a histogram of the comparative skill advantage, separately for children and parents. The figure shows a wide dispersion of math-language skill differences despite the high underlying correlation of math and language skills in each generation (0.67 for children and 0.61 for parents). Comparative skill advantages reach plus and minus 50 percentile points with a

[19] See section 7 for an analysis of STEM outcomes of children. There, we also show that our results are robust to applying different definitions of STEM.

[20] After the 2014/2015 school year, test suppliers other than CITO became available. As it might not be random which schools switched to a different test supplier (Jacobs, van der Velden, and van Vugt (forthcoming)), the calculation of rank positions is done based on the schools that participated in the CITO test throughout the entire period of observation. Results are robust to an alternative calculation of percentile ranks based on the universe of schools.

[21] The choice of calculating the math skill advantage or the language skill advantage has no impact on the analysis. Other plausible formulations of the comparative skill advantage include a simple binary measure (i.e., 1 if math skills exceed language skills, 0 otherwise) and a math-language skill ratio (see Goulas, Griselda, and Megalokonomou (forthcoming)). Our results are robust to these alternative formulations.

standard deviation around 25 percentile points in the pooled sample and each cohort samples (Table 1).

Relatively early test scores are particularly well suited for assessing comparative skill advantages for the purpose of this analysis, even if later-life test scores were available. First, the comprehensive and unified curriculum in all Dutch primary schools implies that our skill data are less contaminated by other influences including subsequent career paths. This is particularly important for any analysis that relates comparative skill advantages to study or occupational choices, because concerns about reverse causality or omitted variables would be aggravated with skills measured at an adult age. Second, as emphasized in models of field-of-study choice (Altonji, Arcidiacono, and Maurel (2016)), individual beliefs about own comparative advantage may be more important than actual comparative advantage, although perceived and actual comparative advantages can be assumed to be highly correlated.[22] Arguably, primary education is the formative period not only for the production of basic skills in math and language, but also for the formation of individuals' perceptions of whether they are better in math than in language or vice versa.

## 3. Early Life Assessments of Cognitive Skills and Long-Run Outcomes

The importance and interpretation of the measured comparative skill advantages for individuals critically hinges on how reliably these relatively early tests of math and language capture variations in longer term economic outcomes. We build on previous literature that has shown that test scores are closely related to adult earnings across developed countries (Hanushek, Schwerdt, Wiederhold, and Woessmann (2015, 2017)). This is reinforced by research showing that early life assessments of cognitive skills are significant predictors of future educational achievement and of labor market outcomes in various other settings (e.g., Büchner, Smits, and Velden (2012), Chetty, Friedman, and Rockoff (2014), Aucejo and James (2021)).

We validate our specific measures of early life math and language skills with economic performance data from the register database. For the parent generation, we link the test scores in math and language assessed around age 12 to administrative records on wages, household income, and household wealth measured 30 years after testing took place (i.e., 2007 for 1977

[22] Similar ideas have also entered in research on learning about comparative advantage across occupations (Papageorgiou (2014)).

cohort; 2012 for 1982 cohort; 2019 for 1989 cohort). Table 2 reports the results. It contains estimates of three specifications of regression models for six different long-run outcomes in the parental generation. Regression models in panel A (panel B) include only math (language) skills, while both skills are included simultaneously in panel C.[23]

The results demonstrate that the level of both early math and early language skills are strongly and consistently related to long-run success measured by educational attainment, field of study choices, hourly earnings, personal income, household income, and household wealth.[24] Importantly, when both skill domains are used in the analysis, math and language skills are independently significant in determining future educational and labor market outcomes.

These correlations between test scores at school and economic outcomes in adulthood clearly show that our measures of cognitive skills are economically meaningful. An equal-percentile move in math performance systematically has a larger impact on economic outcomes compared to a language move, but both skills independently contribute to outcomes even though they are highly correlated. Since information on later life outcomes is obtained from reliable administrative records, the strong correlations of our test score measures with these outcomes also lessen concerns about measurement error in the parent skill measures.

It is of course not the specific knowledge tested on the CITO test that drives the economic results. In the Heckman sense, skills beget skills, and the tests simply index the learning path that these school children are on and that will ultimately lead to the differential choices and outcomes that are observed in adulthood.

The relationship between the parent's math and language skills and the parent's completion of a STEM field of study (column 2 of Table 2) reinforce our focus on comparative advantage when we subsequently turn to understanding the choice of STEM fields by children.[25] The math skills of the parent are positively related to the parent's pursuit of STEM education but significantly *negatively* related to language skills. In other words, holding constant math skills,

[23] All regression models control for an extensive set of covariates for family background, measured at the time of the skill assessment. These covariates are described in the notes to Table 2 and in Section 5.

[24] Interestingly, the wage returns to math (language) skills are also very similar to the estimates for grade 6 test scores reported in Chetty, Friedman, and Rockoff (2014).

[25] Note that the number of observations in column (2) of Table 2 is reduced because administrative data on the completed field of study are available only after 2002.

an increase in language skills (i.e., a move toward greater comparative advantage in language) leads to less completion of STEM education.[26]

These results motivate several aspects of our modeling efforts. First, it is clear that the single dimension of "cognitive skill" found in most prior labor market analyses distorts the richer picture of skills found in our data.[27] Second, given the importance of the different domains of cognitive skills, our efforts to understand the underlying sources of comparative skills must consider the possibility that the production function underlying these alternative skills also differs, even if the two distinct production processes share some common inputs.[28]

## 4. Conceptual Framework

Our analysis takes the perspective of the child and investigates the sources of comparative skills that lead to alternative career outcomes. The overarching conceptual framework comes from the combination of two separate research traditions: the investigation of intergenerational mobility merged with the investigation of educational production functions. The extensive work on intergenerational persistence of economic and noneconomic outcomes, which started over a century ago by Francis Galton (1889), provides structure to the interaction of parents and children. The educational production function analyses address how parents combine with schools and other factors to affect the skills of their children. The combination of the two not only extends both lines of research but also permits new insights into the influence of comparative skill advantages on STEM education.

[26] One noteworthy exemption to the prior single-factor modeling of educational and economic outcomes is the study by Aucejo and James (2021), which also finds that math skills have a positive effect on enrolling in STEM fields while verbal skills have an effect close to zero.

[27] Typically, if multiple test measures are available, studies simply choose one to emphasize (e.g., Murnane, Willett, Duhaldeborde, and Tyler (2000)) or average the scores to deal with potential measurement errors (e.g., Lazear (2003)). Interestingly, however, when information on multiple test domains is available and is used in the labor market analysis, both math and language are independently significant in determining earnings even though little attention has been drawn to the fact (Hanushek, Schwerdt, Wiederhold, and Woessmann (2015)).

[28] In educational production function analyses, distinct differences by test domain are frequently reported in the results and are sometimes (but not often) included in the modeling. Differences in the portion of student math and language outcomes that is related to schools, for example, have often been noted, and the common finding of smaller impacts of schooling on reading has been generally attributed to the role of families, albeit with little analysis (Hanushek and Rivkin (2010)). A number of past production function studies of teacher quality have, however, emphasized between-subject differences in student outcomes (Metzler and Woessmann (2012), Bietenbeck, Piopiunik, and Wiederhold (2018), Hanushek, Piopiunik, and Wiederhold (2019)).

The following subsections develop our conceptual approach to identifying the roles of parents and of educational environments on comparative skills. We construct a simple linear measure of the difference between math ($T_m$) and language ($T_l$) skills that we interpret as a measure of comparative (math) skill advantage (CS):

$$CS = T_m - T_l \tag{1}$$

Our initial objective is to investigate the importance of intergenerational persistence in $CS$ along with the role of nonfamily factors. This then leads into our consideration of how comparative skills enter into STEM choices.

**4.1 Intergenerational Transmission**

A large literature focuses on intergenerational persistence of economic outcomes including income (Solon (1999), Björklund and Jäntti (2011)), educational attainment (Björklund and Salvanes (2011), Black and Devereux (2011)), and more recently cognitive skills (Adermon, Lindahl, and Palme (2021)). These generally follow the linear statistical approach begun by Galton (1889) but with increased sophistication in dealing with a variety of estimation issues including measurement error, the identification of genetic effects, and the influence of extended families.[29]

We borrow the general framework of this prior work to study the intergenerational transmission of comparative skill advantages with the following model:

$$CS^c = \alpha + \beta CS^p + \varepsilon \tag{2}$$

where $CS^c$ and $CS^p$ denote comparative skill advantages of children and parents, respectively.

The key parameter of interest is $\beta$, the measure of intergenerational persistence. Heuristically, the larger $\beta$, the more the family determines child outcomes, leading the prior empirical analyses to focus on obtaining consistent estimates of $\beta$.

This model allows us to measure the strength of the transmission of comparative skill advantages across generations in the standard framework of the literature on intergenerational

[29] There is a parallel, more theoretical line of research following Becker and Tomes (1976, 1979). See the overview in Mogstad (2017) and related empirical analysis in Houmark, Ronda, and Rosholm (2020). Also related is structural modeling of intergenerational effects (e.g., Lee and Seshadri (2018)) including analysis of multiple types of ability (e.g., Guo and Leung (2021)).

mobility. But it is not informative with respect to the question of how this transmission comes about. Analyzing this latter question requires a richer conceptual model of how comparative skill advantages emerge.

### 4.2 Skill Production

The central focus of this analysis is the formation of comparative advantage. The economic literature lacks a common framework for modeling the production of comparative skill advantages, and studies that analyze the impacts of comparative advantage on economic decisions typically take the basic ability differences as exogenously given. We develop a simple framework that characterizes the underlying production function for comparative skills. This provides structure for our thinking about potential confounders in the estimation of the intergenerational transmission of comparative skill advantages.

An important line of inquiry in the economics of education investigates education production functions and how families affect the skills of children. Beginning with the Coleman Report (Coleman et al. (1966)), the first large-scale quantitative study of skill formation in children, there has been ubiquitous recognition of the important role of family background in affecting student achievement. Existing studies have not, however, provided clear evidence on the causal structure of family inputs, and they have not considered the role of family inputs for the formation of relative skills of students, i.e., their comparative advantage.

The general form of a production function formulation of math or language skills that relates closely to our empirical analysis is:

$$\begin{aligned} T_i^c &= \lambda F_i + \varphi S_i + \eta_i \\ &= \lambda_1 G_i + \lambda_2 B_i + \varphi S_i + \eta_i \end{aligned} \tag{3}$$

Test scores of child $i$, $T_i$, are explained by family background factors ($F_i$) and environmental factors that we refer to for expositional purposes simply as school factors ($S_i$). As argued in Björklund, Lindahl, and Plug (2006) and further developed in Adermon, Lindahl, and Palme (2021), it is insightful to partition the family background inputs further into pre-birth factors ($G_i$), i.e., factors that are determined before the child was born, and post-birth factors ($B_i$), i.e., inputs to educational production that are not fully determined at the time of birth. The error term, $\eta_i$, contains all other influences on child test scores and is assumed to be i.i.d. with mean zero.

To streamline our exposition, we focus on the pre-birth factors ($G_i$), which include among other things the cognitive skills of parents. To this end, we subsume in our presentation all post-birth factors (contemporaneous family inputs, $B_i$, and environmental factors, $S_i$ into a new composite error term, $\mu_i$.

$$T_i^c = \lambda_1 G_i + \mu_i \quad where\, \mu_i = \lambda_2 B_i + \varphi S_i + \eta_i \tag{4}$$

We interpret eq. 4 as the reduced form effects of pre-birth factors. Most studies in the literature on education production are primarily interested in studying the causal effects of either school factors or post-birth family inputs, considering pre-birth factors to be simply further covariates in the empirical model. Our interest in intergenerational transmission, however, leads us to study the reduced form causal effect of inputs to educational production of a child that are already determined before the child was born. We think of these as primitives in the production of learning that is captured in the later test scores. With this focus any measured post-birth factors in our empirical model become potentially endogenous.

We extend the one-dimensional skill production model to a two-dimensional model of the production of separate skills in math and language as follows:

$$T_{idm}^c = \rho_{1m} T_{idm}^p + \rho_{2l} T_{idl}^p + \delta_m \psi_{id} + \mu_{idm} \tag{5}$$

$$T_{idl}^c = \rho_{1l} T_{idl}^p + \rho_{2m} T_{idm}^p + \delta_l \psi_{id} + \mu_{idl} \tag{6}$$

In this framework, the domain-specific test scores, $T_{ida}^c$, of child $i$ of dynasty $d$ in domain $a$ (either math or language) are explained by pre-birth factors, which we have further decomposed into parent skills in math, $T_{idm}^p$, and language, $T_{idl}^p$, and other pre-birth factors, $\psi_{id}$.[30] Note that in this general framework, all inputs to educational achievement of child $i$ can potentially affect the production of both skills, and their effects may be different across the two skill domains. In particular, the framework allows for different main effects of parental skills on child skills across domains, $\rho_{1m}$ and $\rho_{1l}$, and also for spill-over effects in the sense that parental math (language) skills can also impact the production of a child's language (math) skills, $\rho_{2m}$ and $\rho_{2l}$.

[30] In the empirical analysis, we can link families over time going back to grandparents, as suggested by Adermon, Lindahl, and Palme (2021) and Moreno (2021).

By differencing eq. 5 and 6, we arrive at a framework for the production of a comparative skill advantage that has its root in a standard educational production model of specific skills:

$$\begin{aligned} CS_i^c &= T_{idm}^c - T_{idl}^c \\ &= (\rho_{1m} - \rho_{2m})T_{idm}^p - (\rho_{1l} - \rho_l)T_{idl}^p + (\delta_m - \delta_l)\psi_{id} + (\mu_{idm} - \mu_{idl}) \end{aligned} \tag{7}$$

Eq. 7 shows that the comparative skill advantage depends on the net effects of the two subject-specific skills. The net effect, $\rho_{1a} - \rho_{2a}$, is defined as the direct effect of parent skills on child skills in the same subject, $\rho_{1a}$, minus the spill-over effect of parent skills in subject $a$ on child skills in the other subject.

A simple Galton-inspired intergenerational transmission model of comparative skill advantages as in eq. 2 can be readily derived from this model by making further assumptions about the effects of parent skills on the production of child skills. In particular, if the net effects are assumed to be constant across domains, i.e., $\rho_{1a} - \rho_{2a} = \rho_1 - \rho_2 = \beta^*$, eq. 7 further simplifies to:

$$CS_i^c = \beta^* CS_i^p + (\delta_m - \delta_l)\psi_{id} + (\mu_{idm} - \mu_{idl}) \tag{8}$$

where $\beta^*$ measures the effect of parents' comparative skill advantage on the comparative skill advantage of their children.[31]

### 4.3 Causality

The framework laid out in eq. 8 clarifies the identification problems that surround a simple Galton regression of child comparative skill advantage on parent comparative skill advantage.

First, any input to skill production that has the same impact on the production of both skills cancels out in eq. 8. Thus, elements of the vector of pre-birth factors ($\psi_{id}$), such as genetic factors or characteristics of grandparents, do not confound the estimation of the intergenerational transmission of comparative skill advantages as long as these factors influence the production of math and language skills in the same way. Similarly, any post-birth inputs to education with

[31] In most studies exploiting within-student across-subject variation in test scores (e.g., Bietenbeck, Piopiunik, and Wiederhold (2018)), it is assumed that no spill-over effects exist and that direct effects are constant across subjects. Under these stronger assumptions, $\beta^*$ identifies the direct effect of parent skills on child skills.

constant effects across skill domains included in the composite error term, $\mu_{idm} - \mu_{idl}$, cancel out. Thus, important determinants of education production such as school quality will also not confound the estimation of the intergenerational transmission of comparative skill advantages as long as children's subject-specific school quality is not correlated with the relative skills of the parents (see below).

Second, the conceptual model in eq. 8 suggests that any bias in the estimation of $\beta^*$ arises because parent comparative skill advantage is correlated either with other pre-birth factors, $\psi_{id}$, or with post-birth factors, $(\mu_{idm} - \mu_{idl})$. In terms of post-birth factors, variation in subject-specific school or teacher quality might be a potential confounder. However, that is only a problem if differences in subject-specific school or teacher quality are correlated with, but not caused by, parent comparative skill advantage. If, for example, parents with a comparative skill advantage in one subject deliberately send their children to schools with higher subject-specific quality in this subject, this is simply a mediator of the reduced-form effect of parent comparative skill advantage, implying no bias. But a bias could arise if, for example, the correlation between parent comparative skill advantage and subject-specific school quality exists because of regional immobility of parents combined with persistent differences in subject-specific school quality across regions.

Pre-birth factors could reflect dynastic predispositions for specific subjects (e.g., arising through occupational legacies) or genetic differences in talent for a specific subject (e.g., a "math gene"). It is an open question whether genetically derived and biologically inherited differences in talent for a specific subject exist,[32] but, if they exist, they would affect the development of skill advantages of both parents and children. This would be part of the direct relationship of comparative skill advantages across generations, and such a genetic link may be an integral part of the intergenerational transmission of cognitive skills. In this case, the genetic component or the dynastic predisposition can simply be viewed as the mechanism for intergeneration transmission.

Establishing the causal relationships of comparative skills is a central part of this analysis, but it is not the only important issue.  For policy purposes, when we wish, say, to affect the

[32]Summarizing the state of the literature, Holden (2008) concluded that "…genius-type alleles, particularly for specific skills such as math ability, don't seem to exist". However, recent studies suggest that math ability might be at least moderately heritable (e.g. Davis and al. (2014), Zhang et al. (2023)).

availability of STEM-trained individuals by addressing comparative skill advantages, we want to know if these comparative skill advantages are malleable. In other words, we are interested in the question whether any "shock" due to post-birth factors in the production of the comparative skill advantage of parents also spills over to the next generation. The extreme alternative scenario would be that any observed correlation in comparative skill advantages across generations is entirely predetermined.

## 5. Empirical Strategy

Our empirical strategy follows directly from the conceptual model outlined in the previous section. We start by estimating the simple regression model of eq. 2 to measure the strength of this transmission. Since we rely solely on between-subject test score variation within children and within parents, observed or unobserved characteristics of children, parents, classrooms, or schools do not confound the estimate on parents' comparative skill advantage as long as they have a similar impact on math and language skills. However, to account for the possibility that covariates affect math and language skills differently, we make multivariate adjustments of the simple Galton correlational model using our parent survey data:

$$CS_I^c = \alpha + \beta CS_i^p + \gamma X_i + \varepsilon_i \tag{9}$$

The vector of covariates, $X_i$, in eq. 9 contains a set of parent and grandparent background characteristics, measured at the time when the observed parent took the skill test, i.e., around the end of primary education (see Table 1).[33] For parents, we include gender, migration background, and number of siblings. For grandparents, we include the age of either grandparent (measured in seven age categories)[34], educational attainment (measured by four categories of the highest level of education of both grandparents)[35], social background (measured by seven categories of occupational status of the main breadwinner in the parent household), and a total of 799 municipality-of-residence fixed effects. By including these variables, we control for some of the

[33] The only exception is grandparent age, which is measured at the birth of the parent.

[34] The age groups are: below 21, 21-25, 26-30, 31-35, 36-40, 41 and above.

[35] Results are robust to using grandfather's or grandmother's level of education or when including both jointly.

pre- and post-birth factors that might also influence the formation of children's comparative skill advantage in eq. 8.[36]

To address questions of causality and malleability discussed in the last section more thoroughly, we additionally pursue an instrumental variable (IV) strategy. This strategy exploits variation in parents' comparative skill advantage driven by pre-birth factors that are arguably exogenous with respect to the formation of children's comparative skill advantage. Specifically, we consider the portion of parents' comparative skill advantage that is driven by between-subject differences in teachers or peer quality during the parent's early formal education.

Our IV approach capitalizes on a unique feature of the data: the sampling design of the parent cohort surveys uses classroom within school as the primary sampling unit. This yields information on math and language test scores for (almost) all classmates of parents around age 12 for two of the three survey cohorts.[37]

Formally, we instrument $CS_i^p$ in eq. 9 by the comparative skill advantage of parents' classroom peers:

$$CS_j^p = \theta + \pi \overline{CS}_{-j}^{class} + \vartheta_j \tag{10}$$

where $\overline{CS}_{-j}^{class}$ is measured as the difference between the percentile ranks in math and language of parents' classroom peers.

In our baseline specification, we construct the instrument by first calculating the percentile rank of average skills of the parents' classmates (leave-out mean) in the country-wide skill distribution for the respective cohort separately for math and language. We then use the difference in the classroom ranks between math and language as our instrument, $\overline{CS}_{-j}^{class}$ in eq. 10.[38] The between-subject difference in classroom ranks measures the relative quality of the

[36] All regressions also control for parent survey indicators and children test year fixed effects.

[37] For more details on the assignment of classrooms in the survey data for the 1982 and 1989 cohort, see Appendix A.3. A small number of observations is missing (1 percent in the 1982 cohort and 5 percent in the 1989 cohort) because not all classmates were tested or were tested but could not be linked in the original dataset. We cannot construct the instrument for the 1977 cohort as the school and class identifiers in that dataset were removed by Statistics Netherlands and could not be restored. In total, the sample in the IV analysis consists of 8,011 parents and 12,268 children.

[38] While we consider differences in classmates' performance ranks between math and language to be the most straightforward measure of the quality of parents' classroom environments, there are also other plausible ways of

formal education environment in math vs. language – whether from teachers, peers, or other elements of schools.[39]

Our IV approach isolates variation in the comparative skill advantage of parents that is independent of dynastic factors potentially impacting the formation of their children's skill advantage. The exclusion restriction is that our instrument is only correlated with children's comparative skill advantage because of its association with the comparative skill advantage of the parents. We address possible concerns with the exclusion restriction in section 6.3.

The IV estimator addresses two potential issues. First, measurement error in the comparative advantage of parents could bias the estimates of intergenerational persistence. Second, residual factors that differentially impact either math or language skills (and are not simply mechanisms by which parents influence children's comparative advantage) may bias the estimated influence of parents. In both cases, our instrument will serve to correct for the potential biases.

## 6. Intergenerational Transmission of Comparative Skill Advantages

Parents directly transmit individual skills to their children. As easily shown in our data, parents with greater math skills have children with greater math skills, and the same subject specific relationship also holds for language (see Figure A2 and Table A1).[40] But our interest goes beyond the separate factors to look at whether comparative skill advantages are transmitted to children.[41]

---

operationalizing the core idea behind this identification strategy. In Appendix A.3, we show that our IV results are robust to several alternative ways of constructing an instrument based on peer performance in math and language.

[39] The 1982 cohort has students in the last year of primary school where the classmates indicate relevant peer and school quality. In the 1989 cohort, students were sampled about halfway through their first academic year in secondary school. Thus, students had 5–7 months of exposure to their teachers and peers in secondary school. Moreover, primary schools often feed into secondary schools, with the consequence that primary school students stay together with at least some of their classmates when entering secondary school. In fact, in the period 2006–2019, where we can observe school transitions in our administrative CITO data, a median share of 19 percent of a student's primary school peers attends the same secondary school-track combination. This share has been slightly decreasing over time, potentially reflecting more school choice.

[40] The patterns of the two subject-specific relationships are remarkably similar: An increase in parent skills by one percentile is associated with an increase in child skills of 0.28 percentiles in math and 0.30 percentiles in language. These estimates are in the same ballpark as the parent-child human capital persistence parameter of 0.361 estimated in Adermon, Lindahl, and Palme (2021).

[41] An alternative interpretation of the single-subject relationships might be that there is a single latent factor (general cognitive ability) and that each of the subject measures is the true latent factor plus random error. If that were the case, however, one would not expect the close relationship of parent-child math and parent-child language to be significantly larger than that for the alternative parent skill (panel C of Table A1).

We establish the basic character of comparative advantage transmission by providing visual evidence of how math-language skill differences are linked across generations. As can be seen in Figure 2, the skill differences of parents and their children are strongly related. Put differently, parents who perform relatively better in math than in language are significantly more likely to have children who are relatively better at math compared to language (and vice versa). The relationship between the comparative skill advantages of parents and their children is linear.

Of course, this bivariate portrayal of intergenerational persistence in comparative skill advantages may be affected by unobserved confounders. To address this, we move to the multivariate specification of eq. 9. We begin in the next subsection with the OLS results that provide the basic persistence estimates. This is followed by the IV estimates that address causality even more rigorously and point to the malleability of parental comparative advantage.

### 6.1. Persistence of Comparative Skill Advantage – Baseline Estimates

Table 3 presents the results of the multivariate regression model from eq. 9.[42] We observe a strong intergenerational transmission of comparative skill advantages even after conditioning on a range of plausible inputs.[43] Accounting only for basic sociodemographic characteristics of parents and grandparents, we find that a difference of 10 percentile ranks between skills in math and language in the parent generation translates into a one-rank difference in the child generation (column 1).

As we progressively add more controls for family background, we find remarkably stable estimates of the key transmission parameter. In column (2) of Table 3, we additionally control for grandparent education. Column (3) adds controls for grandparents' social status, and column (4) further accounts for detailed regional variation by including fixed effects for the municipality of residence when parents took the skill test.[44]

[42] Results for each cohort individually are reported in panel C of Table A2. Estimates are statistically significant in each cohort. Consistent with the subject-specific results in panels A and B, the estimate of parents' comparative skill advantage is largest in the first cohort.

[43] Coefficients on the control variables in the full model are shown in Table A3.

[44] The estimated strength of the intergenerational transmission is very similar when we use the difference between standardized math and language test scores to measure comparative skill advantages instead of percentile ranks (Table A4). This suggests that, at least with high-quality tests such as CITO, the standard implicit assumption of cardinality of previous studies does not distort the results.

Intriguingly, the strength of the intergenerational transmission of comparative skill advantages does not vary by the gender match of parents and their children (column 5 of Table 3). This result differs from a number of previous papers on the intergenerational transmission of educational attainment that have tended to suggest a stronger influence of mothers, particularly for sons (e.g., Black, Devereux, and Salvanes (2005), Holmlund, Lindahl, and Plug (2011), Piopiunik (2014)). In terms of fields of study, Altmejd (2023) suggests that daughters follow mothers more closely while sons follow fathers.

One obvious concern with estimating the persistence in comparative skill advantages is that the results might be distorted by measurement error. We investigate this in two different ways. First, we consider alternative ways to measure comparative advantage that might lessen the impact of measurement error. Second, as described in the next subsection, we employ an IV strategy that directly confronts any possible measurement error. These checks suggest that measurement error is unlikely to affect our estimates in a meaningful way.

Any specific test will measure subject-specific skills only with noise, because the reliability of any given item battery is not perfect. Measurement error in parent cognitive skills could then be aggravated in the estimation of our differenced model (see, for example, Angrist and Krueger (1999)). To understand the possible effect of measurement error, we can directly compare alternative ways of characterizing comparative skills (Table 4). First, in our main analysis, we measure cognitive skills of parents and their children in percentile ranks (column 1). While any measurement error that is rank-preserving does not affect our estimates, errors that change ranks will generally lead to attenuation of our estimates of persistence. In the spirit of the classical solution of aggregating the explanatory variable, we can use broader categories when defining rank measures.[45] This aggregation will reduce the likelihood that measurement error in the tests alters the rank positions of individuals. The estimated transmission parameter changes only little when we measure math and language skills in decile ranks (column 2). The extreme of this aggregation is creating a binary measure that indicates whether or not the rank in math is

[45] The classical treatment of errors in variables aggregates data into two groups and yields consistent estimates of the slope as long as observations are not classified into the wrong group; by eliminating observations at the boundary of the groups, any inconsistency of estimates can be reduced (Wald (1940), Cochran (1968)). We apply several variants of this including aggregating the data into two groups by whether the math percentile is greater or less that the language percentile. Note, however, that errors in the binary measure of comparative skills would no longer be classical because the observed classification cannot be greater than one or less than zero, yielding a correlation of the measurement error with the true value. The bias will nonetheless in general lead to attenuation of the coefficient estimates (Bollinger (1996)).

higher than in language (columns 3–6). Potential measurement errors in this binary specification are largest when all observations are used in calculating the comparative skill advantages (column 3) and are reduced when we drop individuals with small differences in rank positions between math and language in order to reduce the possibility for misclassification. Columns (4)–(6) progressively drop those with comparative skill advantages of less than 5, 10, and 15 ranks. The estimated transmission parameter become very close to the baseline estimates as we increase the gap at the boundary.

Several additional heterogeneity analyses are relegated to the appendix. Most noteworthy, we find that skill transmission tends to become stronger as the education level of grandparents increases, perhaps operating through more negative attitudes toward education in lower-educated families (Table A5, column 1). The strength of transmission does not, however, vary systematically with grandparents' social background (Table A5, column 2).

Our estimation of the intergenerational transmission of comparative skill advantages accounts for all factors that similarly affect math and language skills, such as general motivation and ability, access to learning aids and opportunities, as well as the impacts of peers and neighborhoods. In the spirit of Altonji, Elder, and Taber (2005) and Oster (2019), the stability of the coefficient on parents' comparative skill advantage when we add various parent and grandparent characteristics suggests no major role for unobserved variables in confounding our estimates.

### 6.2 Persistence of Comparative Skill Advantage – IV Estimates

We can further establish the identification of the causal impact of parents on comparative advantage with our IV analysis. But, perhaps even more importantly, the IV results provide evidence on the malleability of parental impacts.

We investigate the impact of nonfamily factors on the pattern of comparative skill advantage through the development of an instrument for parental comparative advantage. As indicated, because we know the composition of the parent's classroom around the end of primary school, we use the average comparative advantage of classmates to characterize the educational environment of each parent. We are not able to distinguish between impacts on comparative advantage coming from a particularly good teacher or from the influence of peers per se, but that has no implications for our analysis.

Differences in the parents' classroom environments are strong predictors of individual comparative skill advantage. The first stage relationship, shown in column (2) of Table 5, indicates that a classroom that scores ten percentile ranks higher in math than in language is associated with parents scoring about 3 percentile ranks higher in math than in language.[46] The reduced form effect on the comparative skill advantage of children is also positive and significant (column 3).

When we instrument parental comparative advantage by the peer comparative advantage, we find significant and very stable estimates of the persistence parameter even with a variety of controls in the model. In column (4) of Table 5, the corresponding IV estimate indicates that an increase of relative math (vs. language) skills of parents by 10 percentile ranks leads to an increase in the relative math skills of children of 1.1 percentile ranks. The estimate is hardly affected by adding controls for grandparents' education and social status to the model. This suggests that the variation in classrooms' comparative skill advantage is unrelated to these characteristics of parental background, which makes it more plausible that it is also unrelated to other unobservable characteristics.

Moreover, the similarity of the estimated transmission parameters in OLS and IV suggests only a limited confounding impact of unobserved subject-specific proclivities of families when estimating the effect of parents' comparative skill advantages on comparative skill advantages of their children.

Our IV estimation provides a way of dealing with the possibility of bias from omitted subject-specific proclivities of families, and at the same time, it also deals with issues of measurement error in the parents' comparative advantage. Because these peer scores are correlated with true individual comparative advantage but not with the individual errors, they are valid instruments to deal with the possibility of measurement error. The IV estimates are virtually identical to the OLS estimates. Combined with the evidence from alternative measures of comparative advantage in the last subsection, these results strongly reinforce the conclusion that measurement error from the test-based measures of comparative advantage does not significantly affect the conclusions of this modeling.

[46] The F-statistic on the excluded instrument is large (>200), indicating a strong instrument.

For our purposes the IV analysis of comparative advantage has a larger and more important implication. These estimates provide insight into the malleability of family cognitive skill influences. Our IV results suggest that the intergenerational transmission of comparative skill advantages within families is not entirely genetic in origin and is not immutable but is partially shaped by factors outside the family (in particular, the formal education system). Thus, independent of the findings of a causal relationship in the within-family transmission of comparative skill advantages, these estimates also indicate that there is room for policy to affect performance not only of the current generation but also of future generations.

### 6.3 Persistence of Comparative Skill Advantage – IV Robustness

The key idea of our IV approach is that comparative skill advantages of parents' classroom peers capture between-subject differences in formal education that affect the formation of comparative skill advantages of parents. The exclusion restriction in the IV approach is that the effect on parents is the only reason why the comparative skill advantages of parents' classroom peers are correlated with the skill advantages of children. Given this assumption, two potential threats to identification are particularly noteworthy.

First, one might be concerned about a bias arising because of the selective school attendance of children. Note, however, as discussed in sections 4.2 and 4.3, school choice in the child generation is not a concern *per se*. As long as school choice is based on average (i.e., subject-invariant) school quality, it does not affect the estimated transmission of comparative skill advantage. And, even if school choice were based on subject-specific quality differences across schools, it would simply be a mediating factor and not a bias as long as parents make this choice because of their own comparative skill advantage. However, school choice becomes a concern if a direct correlation of the instrument with children's sorting into schools arises because of other reasons, such as intergenerational stickiness in school choice (e.g., parents and children attending the same school) in combination with persistent relative advantages of schools in math vs. language education.

However, Table 6 demonstrates that the IV estimates are robust to additionally controlling for math-language achievement differences of the *children's* school peers.[47] This generally mitigates concerns about intergenerational persistence in subject-specific peer quality that would violate the exclusion restriction. When controlling for children's peer quality, our IV estimates of the transmission parameter remain significant and close to the least squares estimates. This is true regardless of whether we add further controls or whether we measure children's peer quality in school ranks or in absolute differences in peer scores between math and language (columns 4–6). Interestingly, the reduced form results in column (3) show that, as expected, math-language skill differences of children's peers strongly predict children's math-language skill differences (i.e., their comparative skill advantage), but the first stage results in column (2) reveal that they cannot predict parents' comparative skill advantage. Thus, parents with relatively higher math skills (vs. language skills) are not systematically more likely to choose schools for their children that perform relatively better in math. This provides direct evidence that school choice based on subject-specific performance is unlikely (see below for a further discussion).[48]

The IV results are also very similar when we exclude children who are more likely to know their parents' classroom peers personally, either because they are in the same school as the children of their parents' former classmates or because they still live in the same municipality as their parents did during their early formal education (Table A10). In these restricted samples, it is considerably less likely that parents' classroom peers exert a direct influence on children's skill development, providing further evidence in favor of the exclusion restriction.

A second potential threat to identification is endogenous switching between schools or classrooms in the parent generation. However, it is extremely unlikely that in the 1970s and 1980s schools or classrooms were selected by grandparents based on the specific school or teacher performance in math *relative* to language. This presumption is reinforced by the fact that no information on the school's subject-specific quality is published in the Netherlands, and even

[47] Note that we measure children's peer quality in the entire school (leave-out mean) because we cannot identify classrooms in our administrative child data. Twenty-seven children with missing school information are excluded in this analysis. Column (1) of Table 6 presents estimates of our baseline least squares model for this reduced sample.

[48] Since we cannot identify classroom within schools in the administrative child data, one may be worried about student sorting within schools. In Table A9, we restrict the sample to children in schools with at most 30 grade-six students in a given year, which are likely to have only one classroom. The estimated transmission parameter remains very similar in the one-classroom sample, but standard errors increase due to the substantial reduction in sample size.

indicators of the school's overall quality are published only since 1997 for secondary education and since 2002 for primary education. Moreover, parents can choose a school, but not the individual teacher or classroom within a school.

If such pattern of school selection existed historically, we would expect it to be even stronger today as more data on school quality has become publicly available in the Netherlands. But additional evidence suggests that school choice based on relative school performance in math vs. language is highly unlikely even today. For instance, the results in column (2) of Table 6 show that parents with a comparative advantage in math skills are not systematically more likely to have their children attend schools that perform relatively better in math. Moreover, we find no relationship between schools' comparative skill advantage[49] and the probability that a school receives a rating of "insufficient" by the school inspectorate of the Netherlands, which is a measure of school quality observable to parents.[50] We also directly address concerns of between- and within-school sorting when parents were at school in a series of robustness checks using one classroom and rural schools, all of which confirm our main findings (Table A11).[51]

## 7. STEM Choices: The Role of Parental Comparative Skill Advantages

The Dutch education system provides an intriguing setting for evaluating the role of comparative skill advantages in determining STEM choices and participation. Students in lower secondary school choose a course profile – a set of courses covering specific areas of study – that guides their work in upper secondary school.[52] Because subsequent fields of study in post-secondary education require specific courses for entry, these profiles have a strong influence on fields of study and, ultimately, occupational choices.

We consider the influence of comparative skill advantages on the choice of a STEM profile in school and of a STEM field of study in post-secondary education.[53] The importance of profile

[49] We constructed this measure as the difference between the percentile ranks in math and language of all children in grade 6 within a school in the nationwide distribution of children's test scores.

[50] With our baseline controls: coef. = 0.0004, $p$=0.406. Inspectorate ratings are available for the period 2012–2018. Conditional on having received a rating, the share of schools with an insufficient judgement is 10.7 percent. However, not all schools are visited by the inspectorate, as only 18.4 percent of schools have received a rating.

[51] Appendix A.3 provides details of these additional robustness checks.

[52] Section 2.1 provides a description of profiles and the interaction with the student's track.

[53] Students are designated as following a STEM-course profile if they take the Technical or Agriculture course profile (low academic track) or the Nature & Technical or Nature & Health course profile (middle/high academic track). STEM study choice is determined based on the 1-digit ISCED97 fields of education classification (UNESCO, 2003), where study programs categorized as Science, Mathematics and Computing, Engineering, Manufacturing and

choice is clear by the linkage to subsequent studies. Of people with a STEM profile in school, 61 percent go on to a STEM field of study (compared to just 14 percent of those with other profiles).

The primary objective of this analysis is understanding the role of parental transmission of skills that guide subsequent STEM choices of their children. But we first put our intergenerational analysis into the context of STEM choices within the generations of parents and children, respectively.

Comparative skill advantages are clearly linked to STEM choices, both within a generation and across generations. The first column of Table 7 considers field of study choices for all participants in the three survey cohorts (as in Table 2), and relates them to their comparative skill advantage around the end of primary school.[54] There is a strong influence of comparative skills on field of study choices across our entire sample (Panel A). We observe a somewhat stronger influence of comparative skills for males (Panel B) than for females (Panel C), but there is an unmistakable influence of comparative skills on the educational choices in the older (parent) generation (and, as previously seen in Table 2, on subsequent labor market outcomes).

From the complete registry data for all Dutch students taking the CITO test from 2006–2019, we also clearly observe that a comparative math skill advantage leads to choosing a STEM profile at school and then a STEM field of study (columns 2 and 3 of Table 7). Again, this relationship holds for both boys and girls, but we see that the influence of comparative skills declines for girls between choosing the course profile and actually entering into a STEM field of study (while, for boys, the opposite holds). This decline may be one contributing factor to the lesser overall participation in STEM fields by girls.

The key element of intergenerational transmission of skills is seen directly in the influence of comparative skill advantage of parents on their children's profile choices and field of study choices (columns 4 and 5 of Table 7). This reduced form relationship, estimated for the 1977

Construction, Agriculture, as well as Medicine and Nursery were classified as a STEM choice of study. Table A12 considers a narrower definition of STEM, which defines course profiles and study programs in the agricultural and medical fields as non-STEM. Results are robust to applying this more restrictive definition. While effect heterogeneity by gender gets more pronounced, this partly reflects the lower baseline probabilities of women choosing these narrowly defined STEM fields.

[54] Throughout this analysis we estimate linear probability models that regress a binary choice variable, e.g., an indicator for choosing a STEM field of study, on comparative math skill advantage defined as the difference between math and language skills (each measured in percentile ranks) and control variables.

cohort where we have data on STEM choices for the majority of children,[55] shows the strong influence of parent skills on STEM participation of their children. This finding, of course, is not very surprising, given that we have shown above that comparative skill advantages of parents filter through to their children. However, the results reported here show that this intergenerational transmission has important consequences for children's educational pathways.

Again, while parents strongly affect both profile choice and field of study choice of girls, there is a weaker influence of comparative skill advantages by the time field of study decisions are made. For boys, there is little difference in the strength of the relation between parents' comparative skill advantages and STEM choices throughout the educational career.

## 8. Conclusions

The role of comparative advantage for economic outcomes has been extensively studied. However, it is not well understood where differences in comparative advantages across individuals come from and how malleable they are. Our analysis shows that comparative skill advantages are transferred across generations: Parents who were relatively better at math (vs. language) in childhood are more likely to have children with a similar comparative skill advantage in math. We also find that parents' comparative skill advantage is a strong predictor of their own STEM choices and those of their children.

The new Intergenerational Transmission of Skills (ITS) database that we develop permits matching skills of Dutch parents and children derived from similar tests taken at similar ages. We measure comparative skill advantage as the ordinal difference between math and language skills in the parent and child generation, respectively, each assessed by the percentile position in the nationwide skill distribution. Our empirical strategy exploits within-family between-subject variation in cognitive skills, thus eliminating all family, school, and neighborhood factors that are not specific to either math or language performance. The estimates of the intergenerational transmission of comparative skill advantage prove very robust when subjected to a variety of specification and robustness exercises.

[55] In the 1977 cohort, we can follow two-thirds (66.5 percent) of children in the post-school activities, allowing us to observe both STEM profile choice and STEM field of study choice. In the later cohorts, this share is substantially smaller (1982 cohort: 43.3 percent; 1989 cohort: 12.2 percent).

Importantly, the intergenerational transmission of comparative skill advantages is malleable and not entirely driven by factors that are fixed within family dynasties. We show this in a novel IV estimation strategy based on comparative skills of the parents' classroom peers. The results of this analysis indicate that nonfamily inputs in the production of skills significantly affect comparative skill advantages that then carry over to future generations. Therefore, educational policies that shift focus from one skill domain to another not only affect comparative skill advantages of current students but also have lasting impacts on future generations. Similarly, evaluations of school programs that ignore achievement spillovers on future generations will understate the full program impact.

Parental skills also influence long-run career patterns. Relatively high math skills of parents promote greater choice of STEM paths for both themselves and their children, as predicted by a Roy model of occupational choice. Suggestively, parents' comparative advantage in math is a stronger determinant of STEM field-of-study choice for boys than for girls, potentially contributing to the observed underrepresentation of women in STEM occupations.

Our results carry an important message regarding policies aimed at increasing the number of STEM-trained workers. The importance of skill-based comparative advantages in determining STEM choices, together with its malleability through environmental factors, suggest that any policy changing the relative cognitive skills of students today will also spill-over to future generations, having a lasting impact on the sorting into STEM (and other) fields.

## References

Acemoglu, Daron, and David Autor. 2011. "Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of Labor Economics, Volume 4b,* edited by Orley Ashenfelter and David Card. Amsterdam: North Holland: 1043-1171.

Adermon, Adrian, Mikael Lindahl, and Mårten Palme. 2021. "Dynastic Human Capital, Inequality, and Intergenerational Mobility." *American Economic Review* 111, no. 5 (May): 1523-1548.

Altmejd, Adam. 2023. "Inheritance of fields of study." (mimeo) Stockholm University.

Altonji, Joseph G., Peter Arcidiacono, and Arnaud Maurel. 2016. "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects." In *Handbook of the Economics of Education, Vol. 5*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: 305-396.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of Political Economy* 113, no. 1: 151-184.

Angrist, Joshua D., and Alan B. Krueger. 1999. "Empirical strategies in labor economics." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card. Amsterdam: North Holland: 1277-1366.

Aucejo, Esteban, and Jonathan James. 2021. "The Path to College Education: The Role of Math and Verbal Skills." *Journal of Political Economy* 129, no. 10 (October): 2905-2946.

Barrera-Osorio, Felipe, Adriana D. Kugler, and Mikko I. Silliman. 2020. "Hard and Soft Skills in Vocational Training: Experimental Evidence from Colombia." NBER Working Paper No. 27548. Cambridge, MA: National Bureau of Economic Research (July).

Becker, Gary S. 1981. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Becker, Gary S., and Nigel Tomes. 1976. "Child endowments and the quantity and quality of children." *Journal of Political Economy* 84, no. pt. 2 (August): S143-S162.

Becker, Gary S., and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of Political Economy* 87, no. 6 (December): 1153-1189.

Bietenbeck, Jan, Marc Piopiunik, and Simon Wiederhold. 2018. "Africa’s Skill Tragedy: Does Teachers’ Lack of Knowledge Lead to Low Student Performance?" *Journal of Human Resources* 53, no. 3 (July 1, 2018): 553-578.

Björklund, Anders, and Markus Jäntti. 2011. "Intergenerational Income Mobility and the Role of Family Background." In *The Oxford Handbook of Economic Inequality*, edited by Brian Nolan, Wiemer Salverda, and Timothy M. Smeeding. Oxford: Oxford University Press.

Björklund, Anders, Mikael Lindahl, and Erik Plug. 2006. "The origins of intergenerational associations: Lessons from swedish adoption data." *Quarterly Journal of Economics* 121, no. 3: 999-1028.

Björklund, Anders, and Kjell G. Salvanes. 2011. "Education and family background: Mechanisms and policies." In *Handbook of the Economics of Education, Vol. 3*, edited by Stephen Machin Eric A. Hanushek and Ludger Woessmann. Amsterdam: North Holland: 201-247.

Black, Sandra E., and Paul J. Devereux. 2011. "Recent developments in intergenerational mobility." In *Handbook of Labor Economics, Vol. 4, Part B*, edited by Orley Ashenfelter and David Card. Amsterdam: North Holland: 1487-1541.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2005. "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital." *American Economic Review* 95, no. 1 (March): 437-449.

Bol, Thijs, and Herman G.  van de Werfhorst. 2013. "Educational Systems and the Trade-Off between Labor Market Allocation and Equality of Educational Opportunity." *Comparative Education Review* 57, no. 2 (May): 285-308.

Bollinger, Christopher R. 1996. "Bounding mean regressions when a binary regressor is mismeasured." *Journal of Econometrics* 73, no. 2 (August): 387-399.

Bond, Timothy N., and Kevin Lang. 2013. "The Evolution of the Black-White Test Score Gap in Grades K-3: The Fragility of Results." *Review of Economics and Statistics* 95, no. 5 (Decenber): 1468-1479.

Borjas, George J. 1987. "Self-selection and the earnings of immigrants." *American Economic Review* 77, no. 4: 531-553.

Brown, Sarah, Steven McIntosh, and Karl Taylor. 2011. "Following in Your Parents' Footsteps? Empirical Analysis of Matched Parent-Offspring Test Scores." *Oxford Bulletin of Economics and Statistics* 73, no. 1 (February): 40-58.

Büchner, Charlotte, Wendy Smits, and Rolf van der Velden. 2012. "Education, cognitive skills and earnings of males and females." ROA RM-2012/2. Maastricht: Research Centre for Education and the Labour Market (February).

Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review* 104, no. 9 (September): 2633-2679.

Cicala, Steve, Roland G. Fryer, and Jörg L. Spenkuch. 2018. "Self-Selection and Comparative Advantage in Social Interactions." *Journal of the European Economic Association* 16, no. 4: 983-1020.

Cochran, W. G. 1968. "Errors of Measurement in Statistics." *Technometrics* 10, no. 4: 637-666.

Committee On Stem Education, National Science and Technology Council. 2018. *Charting a course for success: America's strategy for stem education*. Washington, DC: Executive Office of the President of the United States (December).

Dahl, Gordon B., Dan-Olof Rooth, and Anders Stenberg. 2023. "Intergenerational and Sibling Spillovers in High School Majors." NBER Working Paper Series No. 27618. Cambridge, MA: National Bureau of Economic Research (January (revised)).

Davis, Oliver S. P. et al. 2014. "The correlation between reading and mathematics ability at age twelve has a substantial genetic component." *Nature Communications* 5, no. 1 (2014/07/08): 4204.

de Coulon, Augustin, Elena Meschi, and Anna Vignoles. 2011. "Parents' skills and children's cognitive and non-cognitive outcomes." *Education Economics* 19, no. 5 (December): 451-474.

Deming, David J. 2017. "The growing importance of social skills in the labor market." *Quarterly Journal of Economics* 132, no. 4: 1593-1640.

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2019. "Educational Assortative Mating and Household Income Inequality." *Journal of Political Economy* 127, no. 6: 2795-2835.

Eisenhauer, Philipp, James J. Heckman, and Edward Vytlacil. 2015. "The Generalized Roy Model and the Cost-Benefit Analysis of Social Programs." *Journal of Political Economy* 123, no. 2 (April): 413-443.

Falck, Oliver, Alexandra Heimisch-Roecker, and Simon Wiederhold. 2021. "Returns to ICT skills." *Research Policy* 50, no. 7 (2021/09/01/): 104064.

Galton, Francis. 1889. *Natural Inheritance*. London: MacMillan.

Goulas, Sofoklis, Silvia Griselda, and Rigissa Megalokonomou. forthcoming. "Comparative Advantage and Gender Gap in STEM." *Journal of Human Resources*.

Guo, Naijia, and Charles Ka Yui Leung. 2021. "Do Elite Colleges Matter? The Impact of Elite College Attendance on Entrepreneurship Decisions and Career Dynamics." *Quantitative Economics* 12: 1347-1397.

Hanushek, Eric A., Babs Jacobs, Guido Schwerdt, Rolf van der Velden, Stan Vermeulen, and Simon Wiederhold. 2021. "The Intergenerational Transmission of Cognitive Skills: An Investigation of the Causal Impact of Families on Student Outcomes." NBER Working Paper No. 29450. Cambridge, MA: National Bureau of Economic Research (December).

Hanushek, Eric A., Marc Piopiunik, and Simon Wiederhold. 2019. "The value of smarter teachers: International evidence on teacher cognitive skills and student performance." *Journal of Human Resources* 54, no. 4 (Fall): 857-899.

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about using value-added measures of teacher quality." *American Economic Review* 100, no. 2 (May): 267-271.

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2015. "Returns to skills around the world: Evidence from PIAAC." *European Economic Review* 73: 103-130.

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2017. "Coping with change: International differences in the returns to skills." *Economics Letters* 153: 15-19.

Heckman, James J., and Rodrigo Pinto. 2015. "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs." *Econometric Reviews* 34, no. 1-2 (2015/02/07): 6-31.

Heckman, James J., Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103, no. 6: 2052-2086.

Heckman, James J., and Guilherme Sedlacek. 1985. "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *Journal of Political Economy* 93, no. 6: 1077-1125.

Ho, Andrew D., and Sean F. Reardon. 2012. "Estimating Achievement Gaps From Test Scores Reported in Ordinal “Proficiency” Categories." *Journal of Educational and Behavioral Statistics* 37, no. 4: 489-517.

Holden, Constance. 2008. "Wanted: Math Gene." *Science* 322, no. 5903: 894-894.

Holmlund, Helena, Mikael Lindahl, and Erik Plug. 2011. "The causal effect of parents' schooling on children's schooling: A comparison of estimation methods." *Journal of Economic Literature* 49, no. 3: 615-651.

Houmark, Mikkel Aagaard , Victor Ronda, and Michael Rosholm. 2020. "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills." IZA DP No. 13780. Bonn: Institute of Labor Economics (October).

Jacobs, Babs, and Rolf van der Velden. 2021. "Exploring the Uncharted Waters of Educational Mobility: The Role of Key Skills." ROA Research Memoranda ROA-RM-2021/6. Research Centre for Education and the Labour Market (ROA): Maastricht University (October).

Jacobs, Babs, Stan Vermeulen, and Rolf van der Velden. 2021. "The Intergenerational Transmission of Skills dataset." ROA Technical Report ROA-TR-2021/7. Research Centre for Education and the Labour Market (ROA): Maastricht University (October).

Jacobs, Madelon, Rolf van der Velden, and Lynn van Vugt. forthcoming. "High-Stakes Testing and Educational Careers: Exploiting the Differences in Cut-Offs between Test Recommendations in the Netherlands." *Journal of Research on Educational Effectiveness).*

Kiener, Fabienne, Ann-Sophie Gnehm, Simon Clematide, and Uschi Backes-Gellner. 2022. "IT skills in vocational training curricula and labour market outcomes." *Journal of Education and Work* 35, no. 6-7 (2022/10/03): 614-640.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. "Field of study, earnings, and self-selection." *Quarterly Journal of Economics* 131, no. 3: 1057-1111.

Langer, Christina, and Simon Wiederhold. 2023 "The Value of Early-Career Skills." CESifo Working Paper No. 10288. Munich: CESifo Network (February).

Lazear, Edward P. 2003. "Teacher incentives." *Swedish Economic Policy Review* 10, no. 3: 179-214.

Lazear, Edward P. 2009. "Firm-specific human capital: A skill-weights approach." *Journal of Political Economy* 117, no. 5: 914-940.

Lee, Sang Yoon, and Ananth Seshadri. 2018. "On the Intergenerational Transmission of Economic Status." *Journal of Political Economy* 127, no. 2 (April): 855-921.

Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen. 2021. "On the Family Origins of Human Capital Formation: Evidence from Donor Children." IZA Discussion Paper No. 14708. Bonn: IZA Institute of Labor Economics (September).

Metzler, Johannes, and Ludger Woessmann. 2012. "The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation." *Journal of Development Economics* 99, no. 2 (November): 486-496.

Mogstad, Magne. 2017. "The Human Capital Approach to Intergenerational Mobility." *Journal of Political Economy* 125, no. 6: 1852-1868.

Murnane, Richard J., John B. Willett, Yves Duhaldeborde, and John H. Tyler. 2000. "How important are the cognitive skills of teenagers in predicting subsequent earnings?" *Journal of Policy Analysis and Management* 19, no. 4 (Fall): 547-568.

Nielsen, Eric R. 2015. "The Income-Achievement Gap and Adult Outcome Inequality." Finance and Economics Discussion Series 2015-041. Washington, DC: Federal Reserve Board.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37, no. 2 (April): 187-204.

Papageorgiou, Theodore. 2014. "Learning Your Comparative Advantages." *The Review of Economic Studies* 81, no. 3 (288): 1263-1295.

Piopiunik, Marc. 2014. "Intergenerational transmission of education and mediating channels: Evidence from a compulsory schooling reform in Germany." *Scandinavian Journal of Economics* 116, no. 3: 878-907.

Piopiunik, Marc, Guido Schwerdt, Lisa Simon, and Ludger Woessmann. 2020. "Skills, signals, and employability: An experimental investigation." *European Economic Review* 123(April): 103374.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3, no. 2: 135-146.

Sacerdote, Bruce. 2011. "Nature and Nurture Effects On Children's Outcomes: What Have We Learned From Studies of Twins And Adoptees?" In *Handbook of Social Economics*, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson: North-Holland: 1-30.

Solon, Gary. 1999. "Intergenerational mobility in the labor market." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card. Amsterdam: Elsevier: 1761-1800.

Stoet, Gijsbert, and David C. Geary. 2018. "The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education." *Psychological Science* 29, no. 4 (April): 581-593.

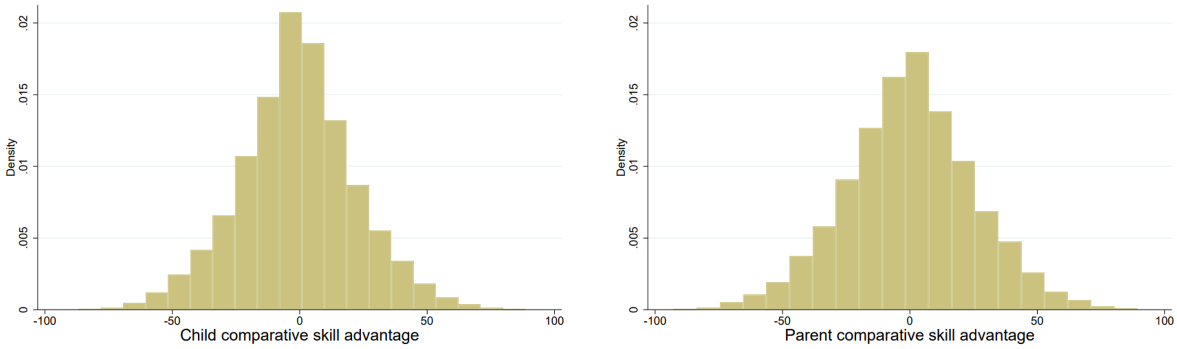
UNESCO. 2017. *Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)*. Paris: United Nations Educational, Scientific and Cultural Organization.

Wald, Abraham. 1940. "The Fitting of Straight Lines if Both Variables are Subject to Error." *The Annals of Mathematical Statistics* 11, no. 3 (September): 284-300.

Willis, Robert J., and Sherwin Rosen. 1979. "Education and self-selection." *Journal of Political Economy* 87, no. 5 (October): S65-S98.

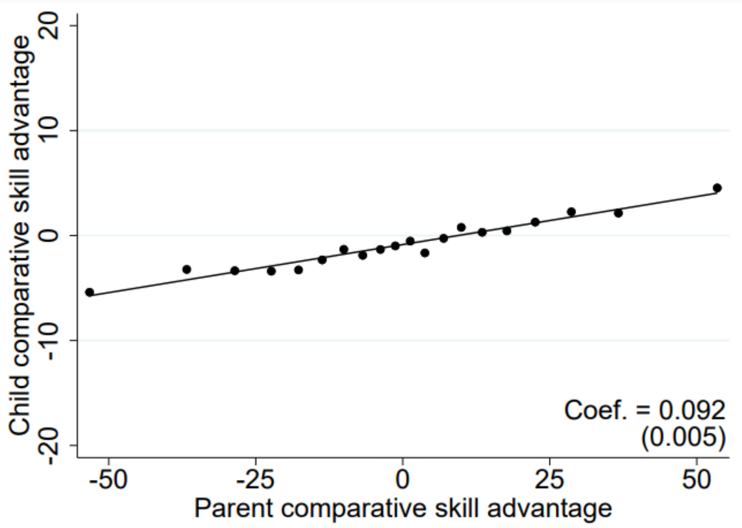
Zhang, Liming, Zhengjun Wang, Zijian Zhu, Qing Yang, Chen Cheng, Shunan Zhao, Chunyu Liu, and Jingjing Zhao. 2023. "A genome-wide association study identified new variants associated with mathematical abilities in Chinese children." *Genes, Brain and Behavior* 22, no. 2: e12843.

**Figure 1: Histogram of comparative skill advantages**



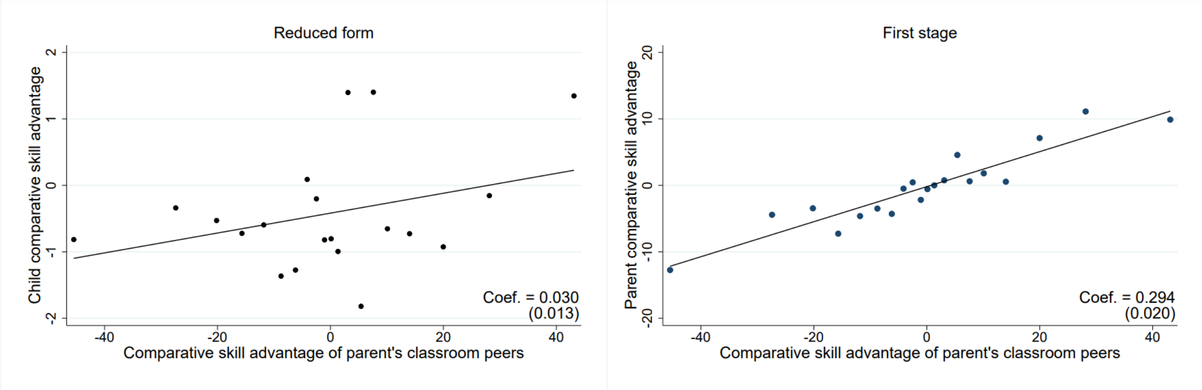

*Notes:* The figure depicts the comparative skill advantage for children (left) and parents (right). The comparative skill advantage of children is measured as the difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. For parents, the comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. *Data sources:* ITS dataset (linked administrative and pooled survey data).

**Figure 2: Intergenerational transmission of comparative skill advantages**




*Notes:* The figure displays a binned scatterplot showing the strength of parent-child transmissions in comparative skill advantages. Child comparative skill advantage is measured as the difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. To construct the figure, we divided the math-language rank difference of parents into 20 ranked equal-sized groups and plotted the mean of the math-language rank difference of children against the mean of the math-language rank difference of parents in each bin. The best-fit line, the coefficient, and the standard error (clustered at the parent level) are calculated from bivariate regressions on the micro data. *Data sources:* Administrative data; pooled ITS survey dataset.

**Figure 3: Comparative skill advantages of parents' classroom peers, parents, and children**




*Notes:* The figure displays two binned scatterplots showing the strength of the relationship between the comparative skill advantage of parents' classroom peers and the comparative skill advantage of children (left) and parents (right), respectively. Child comparative skill advantage is measured as the difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. The comparative skill advantage of parents' classroom peers is measured as the difference between the percentile ranks in math and language test scores of parents' classrooms peers within an education cohort. To construct the scatterplots, we divided the math-language rank difference of parents' classroom peers into 20 ranked equal-sized groups and plotted the mean of math-language rank difference of parents' classroom peers against the mean of the math-language rank difference of children (left figure) or parents (right figure) in each bin. The best-fit line, the coefficient, and the standard error (clustered at the classroom level) are calculated from bivariate regressions on the micro data. *Data sources:* Administrative data; pooled ITS survey dataset.

**Table 1: Descriptive statistics**

| **Variable** | | Pooled | Cohort | | |
|---|---|---|---|---|---|
| | | | 1977 | 1982 | 1989 |
| | | (1) | (2) | (3) | (4) |
| **Child Characteristics** | | | | | |
| Math rank | Mean | 51.71 | 53.80 | 50.61 | 46.62 |
| | SD | 28.06 | 27.87 | 28.05 | 28.00 |
| Language rank | Mean | 52.57 | 55.03 | 51.21 | 46.76 |
| | SD | 28.00 | 27.62 | 28.05 | 28.13 |
| Comparative skill advantage | Mean | -0.86 | -1.23 | -0.60 | -0.14 |
| | SD | 22.86 | 23.30 | 22.41 | 22.20 |
| Course profile | STEM | 0.36 | 0.37 | 0.36 | 0.30 |
| | Non-STEM | 0.48 | 0.47 | 0.49 | 0.52 |
| Field of study | STEM | 0.25 | 0.29 | 0.23 | 0.14 |
| | Non-STEM | 0.51 | 0.59 | 0.47 | 0.33 |
| Gender | Female | 0.50 | 0.50 | 0.51 | 0.51 |
| **Parent Characteristics** | | | | | |
| Math rank | Mean | 50.33 | 53.94 | 47.21 | 44.00 |
| | SD | 28.28 | 27.61 | 28.69 | 27.92 |
| Language rank | Mean | 50.26 | 54.09 | 47.65 | 42.16 |
| | SD | 28.53 | 27.87 | 28.81 | 27.92 |
| Comparative skill advantage | Mean | -0.07 | 0.15 | 0.44 | -1.83 |
| | SD | 25.10 | 23.92 | 27.39 | 24.22 |
| Classroom math rank | Mean | 49.48 | n/a | 54.04 | 45.34 |
| | SD | 28.80 | n/a | 28.79 | 28.18 |
| Classroom language rank | Mean | 49.61 | n/a | 53.07 | 46.46 |
| | SD | 28.33 | n/a | 28.22 | 28.05 |
| Classroom comparative skill adv. | Mean | -0.13 | n/a | 0.97 | -1.12 |
| | SD | 17.88 | n/a | 22.90 | 11.46 |
| Personal income percentile | Mean | 63.29 | 66.36 | 61.67 | 55.72 |
| | SD | 28.84 | 28.77 | 28.65 | 27.79 |
| Household income percentile | Mean | 72.50 | 74.38 | 72.18 | 66.54 |
| | SD | 21.84 | 21.54 | 21.64 | 22.18 |
| Household wealth percentile | Mean | 58.08 | 63.29 | 56.05 | 43.42 |
| | SD | 25.86 | 24.82 | 25.33 | 24.51 |
| Gender | Female | 0.53 | 0.48 | 0.57 | 0.63 |
| Education | Low | 0.24 | 0.21 | 0.25 | 0.30 |
| | Medium | 0.44 | 0.48 | 0.41 | 0.40 |
| | High | 0.25 | 0.28 | 0.25 | 0.17 |
| Migration background | Yes | 0.08 | 0.07 | 0.08 | 0.15 |
| Number of siblings | 0 | 0.06 | 0.06 | 0.05 | 0.05 |
| | 1 | 0.37 | 0.34 | 0.41 | 0.40 |
| | 2 | 0.28 | 0.30 | 0.26 | 0.23 |
| | >2 | 0.23 | 0.25 | 0.21 | 0.19 |

continued on next page

| **Grandparent Characteristics** | | | | | |
|---|---|---|---|---|---|
| Education | Primary education | 0.19 | 0.14 | 0.26 | 0.20 |
| | Lower secondary education | 0.31 | 0.30 | 0.34 | 0.27 |
| | Higher secondary education | 0.29 | 0.33 | 0.19 | 0.34 |
| | Tertiary education | 0.17 | 0.19 | 0.14 | 0.14 |
| Social background | Blue collar worker | 0.28 | 0.28 | 0.29 | 0.28 |
| | Employer – without staff | 0.08 | 0.09 | 0.07 | 0.05 |
| | Employer – with staff | 0.05 | 0.06 | 0.05 | 0.04 |
| | Lower white-collar worker | 0.11 | 0.12 | 0.11 | 0.09 |
| | Middle white-collar worker | 0.19 | 0.21 | 0.16 | 0.17 |
| | Professionals | 0.12 | 0.13 | 0.11 | 0.12 |
| | Other | 0.13 | 0.12 | 0.13 | 0.21 |
| Age at time of birth grandfather | Mean | 30.57 | 31.47 | 29.76 | 29.06 |
| Age at time of birth grandmother | Mean | 27.99 | 28.81 | 27.36 | 26.42 |
| Observations | Total number | 41,774 | 22,417 | 12,930 | 6,427 |

*Notes*: Table reports means, SD, and shares for the pooled sample and by survey cohort. If neither mean nor SD is specified, the reported statistic refers to the share of the respective variable. Child skills are measured as the percentile rank of test scores of linked children in full sample of children taking the test in a given year based on the administrative data. Parent skills are measured as the percentile rank of test scores of linked parents in full sample of parents and nonparents in an education cohort. Comparative skill advantage is measured as the difference between the percentile ranks in math and language. Classroom skills are measured as the percentile rank of average test score of parents' classroom peers (leave-out mean) in full sample of parents and nonparents in an education cohort. Classroom comparative skill advantage is measured as the difference between the percentile ranks in math and language of parents' classroom peers. Children's gender, course profile, and field of study are taken from administrative data. Students are designated as following a STEM course profile if they take the Technical or Agriculture profile (low academic track) or the Nature & Technical or Nature & Health profile (middle/high academic track). STEM study choice is determined based on the 1-digit ISCED97 fields of education classification. Study programs in the Science, Mathematics and Computing, Engineering, Manufacturing and Construction, Agriculture, and Medicine and Nursery were classified as a STEM choice of study. Students who chose a 'combination' course profile, where its STEM-component is unknown, have been coded as non-STEM. Only students progressed far enough in the education system can be assigned a STEM/non-STEM profile/field of study. Parent household income is measured as the percentile in the Dutch distribution in terms of yearly spendable income (sources: labor income, owned companies, and social security benefits). Parent personal income is measured as the percentile in the Dutch income distribution. Household wealth is measured as the percentile in the Dutch distribution in terms of the household's total wealth. Income and wealth data are taken from the administrative data in the child's test-taking year. Parent education is measured as the highest educational degree obtained by the parent observed in the survey data. In parent education, "low" denotes maximum lower secondary education (ISCED 1 or 2); "medium" denotes higher secondary or upper secondary vocational education (ISCED 3 or 4); "high" denotes tertiary education, consisting of higher vocational education and university (ISCED 5 and above). Grandparent education is the highest level of education of both grandparents. Social background is based on the occupation type of the main breadwinner in the parent household at the time of the parent's skill assessment. The "other" category includes, among others, grandparents who are unemployed, pensioned, disabled, or work in their own household. For expositional reasons, mean age of grandparents at the time of the parent's birth is shown; in the regressions, we control for the following age groups: below 21, 21-25, 26-30, 31-35, 36-40, 41 and above. Apart from income and wealth, all (grand-)parent characteristics stem from the survey data. (Grand-)parent characteristics are reported at the child level. *Data sources:* Administrative data; pooled ITS survey database.

## Table 2: Parent cognitive skills and long-term outcomes

| | Higher education | STEM field of study | Log hourly wage | Personal income | Household income | Household wealth |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | | | **Panel A: Math** | | | |
| Math skill rank | 0.0049 | 0.0007 | 0.0039 | 0.187 | 0.140 | 0.179 |
| | (0.0001) | (0.0001) | (0.0001) | (0.004) | (0.004) | (0.004) |
| Further controls | yes | yes | yes | yes | yes | yes |
| R-squared | 0.236 | 0.111 | 0.276 | 0.316 | 0.079 | 0.133 |
| Observations | 61,756 | 28,264 | 41,928 | 53,099 | 55,320 | 53,963 |
| | | | **Panel B: Language** | | | |
| Language skill rank | 0.0047 | 0.00004 | 0.0035 | 0.160 | 0.110 | 0.140 |
| | (0.0001) | (0.0001) | (0.0001) | (0.004) | (0.004) | (0.004) |
| Further controls | yes | yes | yes | yes | yes | yes |
| R-squared | 0.229 | 0.109 | 0.262 | 0.307 | 0.070 | 0.119 |
| Observations | 61,756 | 28,264 | 41,928 | 53,099 | 55,230 | 53,963 |
| | | | **Panel C: Math and language** | | | |
| Math skill rank | 0.0033 | 0.0010 | 0.0029 | 0.143 | 0.115 | 0.147 |
| | (0.0001) | (0.0001) | (0.0001) | (0.005) | 0.005 | (0.005) |
| Language skill rank | 0.0028 | -0.0005 | 0.0018 | 0.077 | 0.044 | 0.055 |
| | (0.0001) | (0.0001) | (0.0001) | (0.005) | (0.005) | (0.005) |
| Further controls | yes | yes | yes | yes | yes | yes |
| R-squared | 0.255 | 0.112 | 0.286 | 0.320 | 0.080 | 0.135 |
| Observations | 61,756 | 28,264 | 41,928 | 53,099 | 55,230 | 53,963 |

*Notes:* Least squares regressions. Sample: Pooled sample of all individuals (parents and nonparents) in the three survey cohorts in columns (1)–(6). All wage, income, and wealth variables are measured 30 years after the skill assessment took place (i.e., 2007 for 1977 cohort; 2012 for 1982 cohort; 2019 for 1989 cohort); higher education degree completion is based on the highest educational degree obtained by the individual observed in the survey data. Dependent variables: Binary variable taking a value of 1 if surveyed individuals obtained a degree in higher vocational education or university education; 0 otherwise (column 1); Binary variable taking a value of 1 if surveyed individuals' highest obtained degree 30 years after the skill assessment took is in a STEM field (column 2); log gross hourly wage, trimmed at the 1st and 99th percentile (column 3); personal income including income from labor, income from owned companies, unemployment and social security, measured as the percentile of the individual in the Dutch personal income distribution (column 4); sum of the personal incomes of all household members measured as the percentile of the household in the Dutch household distribution in terms of yearly spendable income (column 5); household wealth, measured as the percentile of the household in the Dutch household distribution in terms of the household's total wealth, determined by assets minus debts (column 6). Individuals' cognitive skills are measured as the percentile ranks of test scores in the full sample in each survey cohort. All regressions control for individual's gender, migration background, number of siblings, survey indicators, and municipality-of-residence fixed effects (measured at the time of test-taking). Regressions also control for education; social status, and age of individuals' parents at the time of the skill assessment (age refers to individuals' birth). Standard errors (in parentheses) are clustered at the individual level. *Data sources*: Administrative data; pooled ITS survey database.

**Table 3: Intergenerational transmission of comparative skill advantages (OLS)**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Parent comparative skill advantage | 0.098 | 0.097 | 0.096 | 0.094 | 0.098 |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.010) |
| Parent-child gender match | | | | | |
| × Male parent & female child | | | | | -0.001 |
| | | | | | (0.013) |
| × Female parent & male child | | | | | -0.004 |
| | | | | | (0.013) |
| × Female parent & female child | | | | | -0.005 |
| | | | | | (0.013) |
| Grandparent education | | yes | yes | yes | yes |
| Grandparent social background | | | yes | yes | yes |
| Municipality fixed effects | | | | yes | yes |
| R-squared | 0.013 | 0.013 | 0.014 | 0.018 | 0.123 |
| Observations | 41,774 | 41,774 | 41,774 | 41,774 | 41,774 |

*Notes*: Least squares regressions. Sample: All matched parent-children observations in the three education cohorts. Dependent variable: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. Grandparent education is measured by four categories of the highest level of education of both grandparents. Grandparent social background is measured by seven categories of occupational status of the main breadwinner in the parent household. Grandparent education, grandparent social background, and municipality fixed effects refer to time when parents took the skill test. All regressions control for individual's gender, migration background, number of siblings, age of parents at the time of individual's birth, parent survey indicators and children test year fixed effects. Standard errors clustered at the parent level in parentheses. Data sources: Administrative data; pooled ITS survey database.

**Table 4: Addressing measurement error**

| | Ranks in percentiles (baseline) | Ranks in deciles | CSA indicator (all) | CSA indicator (w/o 5) | CSA indicator (w/o 10) | CSA indicator (w/o 15) |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Parent comparative skill advantage | 0.094 | 0.092 | 0.065 | 0.079 | 0.091 | 0.103 |
| | (0.005) | (0.005) | (0.005) | (0.006) | (0.007) | (0.008) |
| Further controls | yes | yes | yes | yes | yes | yes |
| R-squared | 0.016 | 0.017 | 0.008 | 0.011 | 0.014 | 0.016 |
| Observations | 41,774 | 41,774 | 41,774 | 33,478 | 27,099 | 21,521 |

*Notes:* Least squares regressions. Sample: Pooled sample of all matched parent-children observations in the three education cohorts. In column (1), comparative skill advantages of children and parents are measured in percentile ranks, replicating the results from column (4) of Table 3; in column (2), comparative skill advantages of children and parents are measured in decile ranks; in column (3), comparative skill advantages of children and parents are measured as binary variables taking a value of one if the percentile rank in math skills is equal or larger than the percentile rank in language skills, and zero otherwise. In columns (4), (5), and (6), we use the indicator of comparative skill advantage for children and parents from column (3) when dropping parents who are in the range of 5, 10, or 15 percentile positions in the difference between math and language skills, respectively. For children, ranks are calculated in full sample of children taking the test in each test year; for parents, ranks are calculated in full sample of parents and nonparents in an education cohort. Further controls include grandparent education, grandparent social background based on the occupation type of the main breadwinner in the parent household, and municipality fixed effects (all referring to the time when parents took the skill test). All regressions additionally control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the parent level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset.

**Table 5: Intergenerational transmission of comparative skill advantages (IV)**

| | OLS model | First stage IV | Reduced form | Second stage IV | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Parent comparative skill advantage | 0.083 | | | 0.106 | 0.110 |
| | (0.009) | | | (0.046) | (0.047) |
| Classroom comparative skill advantage | | 0.290 | 0.031 | | |
| | | (0.019) | (0.013) | | |
| Further controls | | | | | yes |
| F-statistic excluded instrument | | | | 225.01 | 212.58 |
| R-squared | 0.01 | 0.09 | 0.002 | 0.01 | 0.02 |
| Observations | 12,268 | 12,268 | 12,268 | 12,268 | 12,268 |

*Notes:* Least squares and two-stage least squares regressions. Sample: All matched parent-children observations in the education cohorts of 1982 and 1989. Dependent variables: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data in columns (1), (3), (4), (5), and (6); difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort in column (2). Column (1) replicates baseline least squares model (see column 1 of Table 3) in the IV sample. Classroom comparative skill advantage is measured as the difference between the percentile ranks in math and language of parents' classroom peers within a parent's education cohort. Further controls include grandparent education and grandparent social background based on the occupation type of the main breadwinner in the parent household (all referring to the time when parents took the skill test). All regressions additionally control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the classroom level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset (1982 and 1989 cohort).

**Table 6: Intergenerational transmission of comparative skill advantages (IV): Controlling for children's school quality**

| | OLS model | First stage IV | Reduced form | Second stage IV | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Parent comparative skill advantage | 0.082 | | | 0.092 | 0.097 | 0.096 |
| | (0.009) | | | (0.044) | (0.045) | (0.046) |
| Classroom comparative skill advantage | | 0.289 | 0.027 | | | |
| | | (0.019) | (0.013) | | | |
| Children's school quality | 0.147 | 0.017 | 0.148 | 0.147 | 0.143 | |
| (ranks) | (0.012) | (0.014) | (0.012) | (0.012) | (0.012) | |
| Children's school quality | | | | | | 13.218 |
| (absolute) | | | | | | (1.022) |
| Further controls | | | | | yes | yes |
| F-statistic excluded instrument | | | | 224.91 | 211.67 | 211.73 |
| R-squared | 0.02 | 0.04 | 0.02 | 0.02 | 0.03 | 0.03 |
| Observations | 12,241 | 12,241 | 12,241 | 12,241 | 12,241 | 12,241 |

*Notes:* Least squares and two-stage least squares regressions. Sample: All matched parent-children observations in the education cohorts of 1982 and 1989; children with missing school information are excluded. Dependent variables: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data in columns (1), (3), (4), (5), and (6); difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort in column (2). Column (1) replicates baseline least squares model (see column 1 of Table 3) in the IV sample. Classroom comparative skill advantage is measured as the difference between the percentile ranks in math and language of parents' classroom peers within a parent's education cohort. Children's school quality (ranks) is measured as the difference between the percentile ranks in math and language of children's school peers in the national test score distribution in a given year. Children's school quality (absolute) is measured as the test-year-standardized test score difference between math and language of children's school peers. Further controls include grandparent education and grandparent social background based on the occupation type of the main breadwinner in the parent household (all referring to the time when parents took the skill test). All regressions additionally control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the classroom level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset (1982 and 1989 cohort).

## Table 7: Parents’ comparative skill advantage and STEM choices of parents and children

| | Parent STEM field of study | Child (all) STEM profile | Child (all) STEM field of study | Child (survey) STEM profile | Child (survey) STEM field of study |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | **Panel A: Full sample** | | | | |
| Parent comparative skill advantage (/10) | 0.0078 | | | 0.0090 | 0.0054 |
| | (0.0011) | | | (0.0015) | (0.0013) |
| Child comparative skill advantage (/10) | | 0.0193 | 0.0146 | | |
| | | (0.0002) | (0.0002) | | |
| Further controls | yes | yes | yes | yes | yes |
| Baseline outcome | 0.252 | 0.410 | 0.331 | 0.439 | 0.338 |
| R-squared | 0.085 | 0.065 | 0.065 | 0.016 | 0.011 |
| Observations | 28,264 | 1,161,303 | 1,161,303 | 28,665 | 28,665 |
| | **Panel B: Male sample** | | | | |
| Parent comparative skill advantage (/10) | 0.0092 | | | 0.0090 | 0.0070 |
| | (0.0017) | | | (0.0020) | (0.0020) |
| Child comparative skill advantage (/10) | | 0.0140 | 0.0151 | | |
| | | (0.0003) | (0.0003) | | |
| Further controls | yes | yes | yes | yes | yes |
| Baseline outcome | 0.368 | 0.488 | 0.436 | 0.527 | 0.438 |
| R-squared | 0.032 | 0.043 | 0.028 | 0.019 | 0.026 |
| Observations | 14,236 | 576,031 | 576,031 | 14,358 | 14,358 |
| | **Panel C: Female sample** | | | | |
| Parent comparative skill advantage (/10) | 0.0074 | | | 0.0092 | 0.0041 |
| | (0.0013) | | | (0.0019) | (0.0017) |
| Child comparative skill advantage (/10) | | 0.0246 | 0.0140 | | |
| | | (0.0003) | (0.0002) | | |
| Further controls | yes | yes | yes | yes | yes |
| Baseline outcome | 0.135 | 0.334 | 0.229 | 0.351 | 0.238 |
| R-squared | 0.025 | 0.059 | 0.022 | 0.025 | 0.014 |
| Observations | 14,028 | 585,272 | 585,272 | 14,307 | 14,307 |

*Notes:* Least squares regressions. Sample: Pooled sample of all individuals (parents and nonparents) in the three survey cohorts in column (1); pooled sample of all students that took the CITO test at the end of primary education between 2006–2019 for who we observe both their course- and study profile choice in columns (2) and (3). Children of individuals in the first survey cohort (1977) for who we observe both their course- and study profile choice in columns (4) and (5). Dependent variables: Binary variable taking a value of 1 if surveyed individuals’ highest obtained degree 30 years after participating in the survey is in a STEM (Science, Technology, Engineering, and Mathematics) field (column 1); Binary variable indicating the choice of a STEM course profile at secondary school in columns (2) and (4); binary variable indicating the choice of a STEM field of study after secondary school in columns (3) and (5). Students are designated as following a STEM-course profile if they take the Technical or Agriculture course profile (low academic track) or the Nature & Technical or Nature & Health course profile (middle/high academic track). STEM study choice is determined based on the 1-digit ISCED97 fields of education classification (UNESCO, 2003), where study programs categorized as Science, Mathematics and Computing, Engineering, Manufacturing and Construction, Agriculture, as well as Medicine and Nursery were classified as a STEM choice of study. Baseline values are calculated based on observations with non-missing information on STEM choices. Regressions in column (1) control for individual's gender, migration background, number of siblings, age of parents at the time of individual's birth, survey indicators, education and social background of grandparents, as well as municipality fixed effects. Regressions in columns (2) and (3) control for student gender, migration background, student test year, and school fixed effects. Standard errors (in parentheses) are clustered at the individual level in column (1), at the school level in columns (2) and (3), and at the parent level in columns (4) and (5). *Data sources*: Administrative data; pooled ITS survey database.

# Where Do STEM Graduates Stem From? The Intergenerational Transmission of Comparative Skill Advantages

Eric A. Hanushek, Babs Jacobs, Guido Schwerdt,
Rolf van der Velden, Stan Vermeulen, and Simon Wiederhold[†]

## Online Appendices

[†] Hanushek: Hoover Institution, Stanford University, CESifo, and NBER, hanushek@stanford.edu; Jacobs: Research Centre for Education and the Labour Market (ROA) at Maastricht University, bpja.jacobs@maastrichtuniversity.nl; Schwerdt: University of Konstanz, ifo Institute, CESifo, and IZA, guido.schwerdt@uni-konstanz.de; van der Velden: Research Centre for Education and the Labour Market (ROA) at Maastricht University, r.vandervelden@maastrichtuniversity.nl; Vermeulen: Research Centre for Education and the Labour Market (ROA) at Maastricht University, c.vermeulen@maastrichtuniversity.nl; Wiederhold: Halle Institute for Economic Research (IWH), University of Halle, ifo Institute, and CESifo, simon.wiederhold@iwh-halle.de

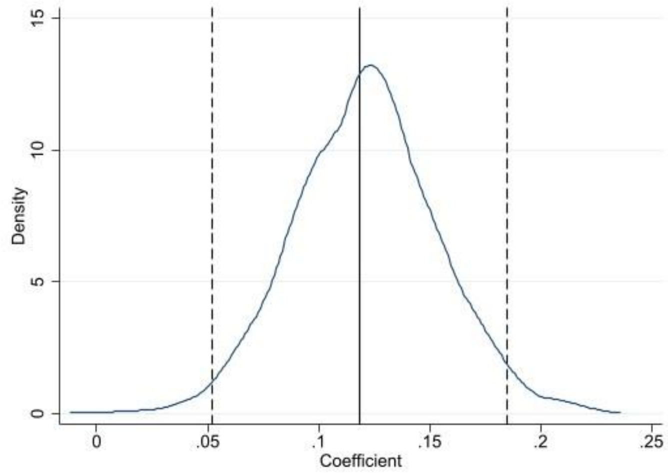
## A.1 Appendix for Measurement error

**Potential measurement error due to observing only one parent**

We usually observe the cognitive skills of only one of the parents in our linked data, and this could potentially induce measurement error in the parent skill variables. To address this, we make use of the subsample of 365 students in the ITS dataset where we observe both parents. We perform the following analysis: In the two-parent sample, we randomly drop one of the parents and estimate the relationship between child and parent comparative skill advantages. Figure A1 shows the distribution of the coefficients on parents' comparative skill advantage when redrawing samples 1,200 times. The resulting estimates are close to the coefficient obtained in the two-parent sample (indicated by the solid vertical line). In fact, 96 percent of the bootstrapped coefficients are within the 95 percent confidence interval of the two-parent-sample coefficient (indicated by the dashed vertical lines). This exercise provides direct evidence that observing only one of the parents in the majority of our data is unlikely to affect our results.[57]

[57] In the two-parent sample, the cognitive skills of mothers and fathers are significantly positively correlated (correlation coefficients of 0.25 for math, 0.32 for language, and 0.14 for the difference between math and language). This corroborates previous evidence on positive assortative mating on educational attainment (e.g., Eika, Mogstad, and Zafar (2019), Educational Assortative Mating and Household Income Inequality, *Journal of Political Economy* 127, no. 6: 2795-2835).

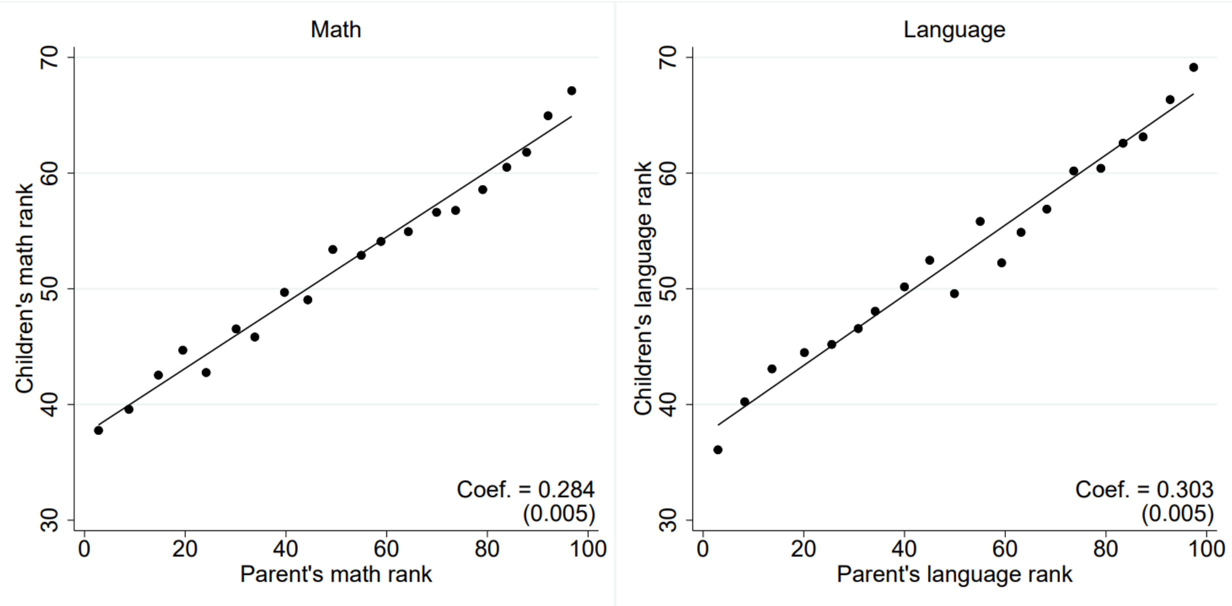
**Figure A1: Randomly dropping one parent in two-parent sample**




*Notes:* The figure depicts estimated coefficients on parents' comparative skill advantage in the least squares model (see eq. 10) when redrawing samples 1,200 times. Estimations are conducted based on 365 children for whom we observe both parents in the survey data. In each of the 1,200 iterations we randomly drop one of the parents for each child and estimate the relationship between child and parent comparative skill advantages. Solid vertical line indicates coefficient in the two-parent estimation, dashed lines indicate 95 percent confidence interval. *Data sources:* Administrative data; pooled ITS survey dataset.

## A.2 Appendix for Section 6.1: OLS Models

**Figure A2: Binned scatterplots of child cognitive skills and parent cognitive skills**




*Notes:* The figure displays two binned scatterplots showing the strength of parent-child transmissions in math skills (left) and language skills (right). Child skills are measured as the percentile rank of test scores of linked children in full sample of children taking the test in a given year based on the administrative data. Parent skills are measured as the percentile rank of test scores of linked parents in full sample of parents and nonparents in an education cohort. To construct the figure, we divided the parent skill rank into 20 ranked equal-sized groups and plotted the mean of the children skill rank against the mean of the parent skill rank in each bin. The best-fit line, the coefficient, and the standard error (clustered at the parent level) are calculated from bivariate regressions on the micro data. *Data sources:* ITS dataset (linked administrative and pooled survey data).

## Table A1: Intergenerational transmission of subject-specific skills

| | Child math skill rank | Child language skill rank |
|---|---|---|
| | (1) | (2) |
| | **Panel A: Math** | |
| Math skill rank | 0.260 | 0.234 |
| | (0.006) | (0.006) |
| R-squared | 0.121 | 0.124 |
| Observations | 41,774 | 41,774 |
| | **Panel B: Language** | |
| Language skill rank | 0.208 | 0.264 |
| | (0.006) | (0.006) |
| R-squared | 0.101 | 0.136 |
| Observations | 41,774 | 41,774 |
| | **Panel C: Math and language** | |
| Math skill rank | 0.209 | 0.125 |
| | (0.007) | (0.007) |
| Language skill rank | 0.089 | 0.193 |
| | (0.007) | (0.007) |
| R-squared | 0.125 | 0.144 |
| Observations | 41,774 | 41,774 |
| | **Control variables in all panels** | |
| Grandparent education | yes | yes |
| Grandparent social background | yes | yes |
| Municipality fixed effects | yes | yes |

*Notes:* Least squares regressions. Sample: Pooled sample of all matched parent-children observations in the three education cohorts. Dependent variables: Math skills of children in column (1); language skills of children in column (2). Children's cognitive skills are measured as the percentile rank of test score of children in full sample of children taking the test in a given year based on the administrative data. Parents' cognitive skills are measured as the percentile rank of test score of parents in full sample of parents in an education cohort. Grandparent education is measured by four categories of the highest level of education of both grandparents. Grandparent social background is measured by seven categories of occupational status of the main breadwinner in the parent household. Grandparent education, grandparent social background, and municipality fixed effects refer to time when parents took the skill test. All regressions control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, and children test year fixed effects. Standard errors clustered at the parent level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset.

**Table A2: Estimates of intergenerational skill transmission for each cohort separately**

| | Panel A: Math | | | |
|---|---|---|---|---|
| | Pooled | Cohort | | |
| | | 1977 | 1982 | 1989 |
| Parent skill rank | 0.260 | 0.268 | 0.250 | 0.242 |
| | (0.006) | (0.008) | (0.011) | (0.016) |
| R-squared | 0.121 | 0.130 | 0.134 | 0.146 |
| Observations (students) | 41,774 | 22,417 | 12,930 | 6,427 |
| | **Panel B: Language** | | | |
| Parent skill rank | 0.264 | 0.288 | 0.224 | 0.251 |
| | (0.006) | (0.008) | (0.010) | (0.016) |
| R-squared | 0.136 | 0.149 | 0.141 | 0.164 |
| Observations (students) | 41,774 | 22,417 | 12,930 | 6,427 |
| | **Panel C: Math and language** | | | |
| Parent comparative skill advantage | 0.094 | 0.122 | 0.068 | 0.081 |
| | (0.005) | (0.008) | (0.009) | (0.013) |
| R-squared | 0.067 | 0.025 | 0.015 | 0.022 |
| Observations | 41,774 | 22,417 | 12,930 | 6,427 |
| | **Control variables in all panels** | | | |
| Grandparent education | yes | yes | yes | yes |
| Grandparent social background | yes | yes | yes | yes |
| Municipality fixed effects | yes | yes | yes | yes |

*Notes:* Least squares regressions. Sample: Pooled sample of all matched parent-children observations in the three education cohorts. Dependent variables: Math skill rank of children in Panel A; language skill rank of children in Panel B; skill rank difference between math and language in Panel C; rank is the percentile rank of test scores of linked children in full sample of children taking the test in a given year based on the administrative data. Parent skill rank is the percentile rank of test scores of linked parents in full sample of parents and nonparents in an education cohort; parent comparative skill advantage is the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. Grandparent education is measured by four categories of the highest level of education of both grandparents. Grandparent social background is measured by seven categories of occupational status of the main breadwinner in the parent household. Grandparent education, grandparent social background, and municipality fixed effects refer to time when parents took the skill test. All regressions control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, and children test year fixed effects. In Panel C: Standard errors clustered at the parent level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset.

## Table A3: Coefficients on control variables in the least squares model (Table 3, Col. 4)

| Variables | (1) | Variables | (2) |
|---|---|---|---|
| Parent comparative skill advantage | 0.094 | Other | -1.771 |
| | (0.005) | | (0.606) |
| | | No answer | 0.374 |
| | | | (1.068) |
| *Parent characteristics* | | *Grandparent characteristics* | |
| Female | 0.936 | Age grandfather at time of parent birth: 21-25 | 0.682 |
| | (0.258) | | (1.176) |
| Migrant | -0.208 | Age grandfather at time of parent birth: 26-30 | 0.310 |
| | (0.444) | | (1.200) |
| Number of siblings: 1 | -0.090 | Age grandfather at time of parent birth: 31-35 | 0.544 |
| | (0.533) | | (1.232) |
| Number of siblings: 2 | -0.328 | Age grandfather at time of parent birth: 36-40 | 0.204 |
| | (0.547) | | (1.289) |
| Number of siblings: 3 or more | 0.885 | Age grandfather at time of parent birth: 41 and above | 0.102 |
| | (0.566) | | (1.376) |
| Number of siblings: missing | -1.074 | Age grandfather at time of parent birth: missing | -0.112 |
| | (0.902) | | (2.207) |
| *Grandparent education* | | Age grandmother at time of parent birth: 21-25 | -0.851 |
| Grandparent education: lower secondary | -0.655 | | (0.635) |
| | (0.372) | Age grandmother at time of parent birth: 26-30 | -0.840 |
| Grandparent education: upper secondary | -0.762 | | (0.684) |
| | (0.399) | Age grandmother at time of parent birth: 31-35 | -1.647 |
| Grandparent education: tertiary | -1.520 | | (0.764) |
| | (0.503) | Age grandmother at time of parent birth: 36-40 | -0.589 |
| Grandparent education: missing | -1.097 | | (0.891) |
| | (0.988) | Age grandmother at time of parent birth: 41 and above | -1.346 |
| *Grandparent social background* | | | (1.241) |
| Blue-collar worker | -1.721 | Age grandmother at time of parent birth: missing | 9.805 |
| | (0.535) | | (6.948) |
| Employer with staff | -1.618 | | |
| | (0.728) | | |
| Lower white-collar worker | -2.318 | | |
| | (0.611) | | |
| Middle white-collar worker | -2.287 | | |
| | (0.576) | | |
| Professionals | -2.067 | | |
| | (0.633) | | |
| Municipality fixed effects | | yes | |
| R-squared | 0.018 | Observations | 41,774 |

*Notes:* Least squares regressions. Sample: All matched parent-children observations in the three education cohorts. Dependent variable: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. Omitted categories: Gender: male; migration background: native; number of siblings: none; grandparent education: primary; grandparent social background: employer without staff; age grandfather at time of parent birth: 20 years or lower; age grandmother at time of parent birth: 20 years or lower. Grandparent education, grandparent social background, and municipality fixed effects refer to time when parents took the skill test. All regressions control for parent survey indicators and children test year fixed effects. Standard errors clustered at the parent level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset.

**Table A4: Intergenerational transmission of comparative skill advantage (cardinal skill measures)**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Parent comparative skill advantage | 0.098 | 0.098 | 0.097 | 0.096 |
| | (0.005) | (0.005) | (0.005) | (0.005) |
| Grandparent education | | yes | yes | yes |
| Grandparent social background | | | yes | yes |
| Municipality fixed effects | | | | yes |
| R-squared | 0.013 | 0.014 | 0.014 | 0.018 |
| Observations | 41,774 | 41,774 | 41,774 | 41,774 |

*Notes:* Least squares regressions. Sample: All matched parent-children observations in the three education cohorts. Dependent variable: Difference between math and language test scores of linked children; test scores are standardized with mean zero and SD one in full sample of children taking the test in each test year. Parent comparative skill advantage is measured as the difference between math and language test scores of linked parents; test scores are standardized with mean zero and SD one in full sample of parents and nonparents in each education cohort. Grandparent education is measured by four categories of the highest level of education of both grandparents. Grandparent social background is measured by seven categories of occupational status of the main breadwinner in the parent household. Grandparent education, grandparent social background, and municipality fixed effects refer to time when parents took the skill test. All regressions control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the parent level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset.

**Table A5: Effect heterogeneity**

| | (1) | (2) |
|---|---|---|
| Parent comparative skill advantage | 0.062 | 0.097 |
| | (0.012) | (0.009) |
| Grandparent education | | |
| × Lower secondary | 0.035 | |
| | (0.015) | |
| × Upper secondary | 0.046 | |
| | (0.015) | |
| × Tertiary | 0.048 | |
| | (0.017) | |
| × Missing education information | 0.019 | |
| | (0.026) | |
| Grandparent social background | | |
| × Independent contractor | | -0.016 |
| | | (0.020) |
| × Employer with staff | | 0.031 |
| | | (0.023) |
| × Lower white-collar worker | | -0.011 |
| | | (0.017) |
| × Middle white-collar worker | | 0.018 |
| | | (0.015) |
| × Professionals | | -0.013 |
| | | (0.017) |
| × Other | | -0.028 |
| | | (0.016) |
| × No answer | | 0.007 |
| | | (0.028) |
| Municipality fixed effects | yes | yes |
| R-squared | 0.019 | 0.018 |
| Observations | 41,774 | 41,774 |

*Notes:* Least squares regressions. Sample: All matched parent-children observations in the three education cohorts. Dependent variable: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. The coarser definition of grandparent education used in this table combines primary and lower secondary education to the lower education category, while upper secondary and tertiary education are referred to as medium and tertiary education, respectively. The coarser definition of parent social status lumps together "employer without staff" and "employer with staff" in the "employer" category, and the "other" and "unknown" in the "other" category. Omitted category in column (1) is low education (at most lower secondary); omitted category in column (2) is blue collar worker. Grandparent education, grandparent social background, and municipality fixed effects refer to time when parents took the skill test. All regressions further control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the parent level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset.

### Potential Mechanisms

Our estimates of the intergenerational transmission of comparative skill advantages still leave several open questions. In particular, it would be valuable to understand why parents with different cognitive skill mixes when they finished primary education produce offspring with similar skill mixes. Linking the ITS data with administrative data on parents' future outcomes, we pursue an exploratory investigation of possible mediators of the skill transmission. Specifically, we observe the highest obtained educational degree and current income of parents, as well as household income and wealth – each of which is a plausible contributor to child skills.

We observe that parents who performed relatively better in math than in language at school advance farther in the education system, earn more, and accumulate more wealth (Table A6). However, the role of these economic factors in explaining the extent to which comparative skill advantages are transmitted from one generation to the next is very limited. Adding the parental economic variables to the baseline transmission model leaves the parent skill coefficient virtually unchanged (Table A7). This reflects the fact that the considered measures of parent economic success are only weakly, if at all, correlated with child comparative skill advantages after conditioning on parent skill advantages.[58]

Our simple analysis of mechanisms has two important caveats. First, interpreting the results in Table A7 as showing the effect of parents' comparative skill advantages net of the mediator hinges on additional conditional independence assumptions with respect to unmeasured mediators and confounders correlated with both the included mediator and the outcome. Second, a straightforward decomposition of the effect of parent skill advantages on child skill advantages into shares attributed to one or several mediators can only be achieved when imposing additional assumptions (see Heckman, Pinto, and Savelyev (2013)).[59]

If parent education, income, and wealth do not drive intergenerational skill transmission, what might? Plausible alternative mechanisms are factors that affect subject-specific informal learning in the family, such as role model effects (leading by example), passion for a subject, or

[58] In an unreported subject-specific mediation analysis, we find that the considered mediators (in particular, the highest obtained educational degree of parents) are relevant in explaining the subject-specific skill transmission from parents to their children. However, the mediators affect math and language skills similarly, so they cannot meaningfully explain the transmission of comparative skill advantages.

[59] More advanced decomposition methods could be contemplated (e.g., Heckman, Pinto, and Savelyev (2013), Heckman and Pinto (2015)). However, because the observed potential mediators explain very little of the intergenerational transmission of comparative skill advantages, we stop at the basic analysis in Table A7.

pedagogical skills. It seems likely that parents with particularly high skills in one subject will also be more willing and more able to transmit these skills to their children. Unfortunately, our data do not allow to test this presumption directly.

**Table A6: Potential mediators of intergenerational transmission of comparative skill advantages**

| | Parent higher education (1) | Parent income (2) | Household income (3) | Household wealth (4) |
|---|---|---|---|---|
| Parent comparative skill advantage | 0.0003 | 0.0199 | 0.0156 | 0.0292 |
| | (0.0001) | (0.006) | (0.005) | (0.007) |
| Grandparent education | yes | yes | yes | yes |
| Grandparent social background | yes | yes | yes | yes |
| Municipality fixed effects | yes | yes | yes | yes |
| R-squared | 0.161 | 0.426 | 0.103 | 0.184 |
| Observations | 41,774 | 38,957 | 41,134 | 36,973 |

*Notes:* Least squares regressions. Sample: Pooled sample of all matched parent-children observations in the three education cohorts. Dependent variables: Binary variable taking a value of 1 if parents obtained a degree in higher vocational education or university education; 0 otherwise (column 1). Parent income including income from labor, income from owned companies, unemployment and social security, measured as the percentile of the parent in the Dutch personal income distribution in the child's test-taking year (column 2). Sum of the personal incomes of all household members measured as the percentile of the household in the Dutch household distribution in terms of yearly spendable income in the child's test-taking year (column 3). Household wealth, measured as the percentile of the household in the Dutch household distribution in terms of the household's total wealth, determined by assets minus debts in the child's test-taking year (column 4). Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. Grandparent education is measured by four categories of the highest level of education of both grandparents. Grandparent social background is measured by seven categories of occupational status of the main breadwinner in the parent household. Grandparent education, grandparent social background, and municipality fixed effects refer to time when parents took the skill test. All regressions further control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the parent level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset.

**Table A7: Analysis of potential mechanisms**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Parent comparative skill advantage | 0.094 | 0.094 | 0.094 | 0.094 | 0.094 |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| Parent education | | | | | |
| Medium | | -0.168 | | | |
| | | (0.327) | | | |
| High | | -1.182 | | | |
| | | (0.377) | | | |
| Missing | | 0.616 | | | |
| | | (0.528) | | | |
| Parent income | | | 0.016 | | |
| | | | (0.054) | | |
| Household income | | | | 0.137 | |
| | | | | (0.058) | |
| Household wealth | | | | | 0.232 |
| | | | | | (0.053) |
| Grandparent education | yes | yes | yes | yes | yes |
| Grandparent social background | yes | yes | yes | yes | yes |
| Municipality fixed effects | yes | yes | yes | yes | yes |
| R-squared | 0.018 | 0.019 | 0.018 | 0.018 | 0.019 |
| Observations | 41,774 | 41,774 | 41,774 | 41,774 | 41,774 |

*Notes:* Least squares regressions. Sample: All matched parent-children observations in the three education cohorts. Dependent variable: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. Parent education is measured as the highest educational degree obtained by the observed parent (omitted category: low education); low education: at most lower secondary; medium education: higher secondary and upper secondary vocational education; high education: tertiary education, consisting of higher vocational education and university. Household income is based on the percentile of the household in the Dutch household distribution in terms of yearly spendable income in the child's test-taking year. Parent personal income is based on the percentile of the parent in the Dutch personal income distribution (including income from labor, income from owned companies, unemployment and social security) in the child's test-taking year. Household wealth is based on the percentile of the household in the Dutch household distribution in terms of the household's total wealth, determined by assets minus debts in the child's test-taking year. Missing values for parent education (3.5 percent), parent income (6.7 percent), household income (1.5 percent), and household wealth (11.5 percent) are imputed (imputation dummies added to the regression models). Grandparent education is measured by four categories of the highest level of education of both grandparents. Grandparent social background is measured by seven categories of occupational status of the main breadwinner in the parent household. Grandparent education, grandparent social background, and municipality fixed effects refer to time when parents took the skill test. All regressions further control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the parent level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset.

## A.3 Appendix for Section 6.2: Instrumental Variable Approach

**Identification of classrooms**

Sampling was done at the classroom level in all three parent cohorts. However, for the 1977 cohort school and class identifiers were removed by Statistics Netherlands and could not be retrieved. In the 1989 cohort, classroom identifiers are directly available. For the 1982 cohort, which is sampled in the last year of primary school, a classroom identifier was collected but the identifier is no longer available. In this cohort, however, we can approximate students' classmates by combining available information at the school and municipality level that is consistently available for all students. At the school level, we have religious denomination and number of grade 6 classrooms. Together with the municipality code of students' place of residence, this provides an indication of which students were potentially classmates. For example, if 20 students resided in the same municipality and attended the same protestant primary school with one grade 6 classroom, they can reasonably be assumed to have been classmates. However, for larger municipalities and more common denominations, this combined information is not sufficient to uniquely identify classrooms. Hence, we put a lower- and an upper-bound on class size to include only those students in the sample for whom we can be reasonably certain that they were indeed classmates.

In the main IV analyses for the 1982 cohort, minimum class size has been restricted to 15 students, and maximum class size to 30 students. We used these values because a class size of 15 students corresponds to the $10^{th}$ percentile and a class size of 29 students to the $90^{th}$ percentile of the class-size distribution in the 1989 cohort.[60] The minimum class size restriction is introduced because classmates are partly identified based on municipality code of residence, not on municipality code of school attendance. An unreasonably small number of students from a certain municipality likely implies that they attend a school in a different municipality. While they still may attend the same school as their peers from the same municipality, they will also share a classroom with other students whom we are not able to identify. The reason for a maximum class size is that in large municipalities, the combination of number of grade 6

[60] For comparison, the first percentile of the class-size distribution in the 1989 cohort corresponds to a class with 9 students, while the $99^{th}$ percentile corresponds to class with 32 students.

classrooms and denomination does not uniquely identify schools.[61] There are likely to be more schools with the same profile from the same municipality that participate in the survey, and assigning all these students to the same 'classroom' would not be appropriate.

Our class size restrictions could introduce selectivity in the type of schools and students for whom we can implement our IV approach in the 1982 cohort. This might affect our estimated average effect if effect heterogeneity is large. We address this concern in two ways. First, we extend our class size restrictions to include a range of class sizes from 10 to 35 in the 1982 cohort. The IV estimate on parent comparative skill advantage in the full IV sample drops from 0.110 in the baseline to 0.071 when we use the extended class-size range for the 1982 cohort but remains significant at the 10 percent level. The decrease in coefficient magnitude is not surprising when considering that the broader range of included class sizes introduces some measurement error. Second, we impose a class size restriction of 15 to 30 students also in the sample of the 1989 cohort, for which we have perfectly reliable class identifiers. We find that this restriction has virtually no effect on our IV estimate.

Furthermore, to benchmark the quality of our classroom assignment procedure in the 1982 cohort, we apply the same procedure to the data of the 1989 cohort. The correlation coefficient between the comparative skill advantages of the actual classroom and the predicted classroom (based on our procedure) is 0.72. The correlation coefficient between the class ranks in math (language) of the actual and predicted classroom are 0.86 (0.88). The corresponding IV estimates of the intergenerational transmission of comparative skill advantages based on the actual classroom and the predicted classroom are not statistically significantly different from each other.

**Robustness to other definitions of the comparative skill advantage of classroom peers**

The core idea behind the IV approach is that differences in parent classroom environments affect parents' comparative skill advantage, but do not have an independent impact on children's skill advantage. In operationalizing this idea, we have some leeway of how to construct the instrument. In our baseline specification, we use the difference between the percentile ranks in math and language tests of parents' classroom peers. That is, we calculate for every parent the

[61] Note that we identify 'schoolmates' in cases where we can uniquely identify a school, but know that the number of surveyed classrooms in this school is larger than one. However, the vast majority of schools have only one classroom.

average performance of classmates, while excluding the parent's test score in the calculation of the average (i.e., leave-out mean). This is a straightforward and intuitive way to measure the quality of the classroom environment, but there are also other plausible approaches.

In Table A8 we show that the IV results are robust to various other ways of constructing the instrument. All estimates of parents' comparative skill advantage in columns (1) to (6) are not statistically significantly different from each other. In column (1), we report our baseline estimate. In column (2), we construct differences in performance ranks between math and language of the entire classroom (i.e., including the parents). However, with this specification of the instrument, the strong first-stage relationship is partly mechanical because the class rank instrument also includes parent cognitive skills. Column (3) presents a non-parametrical version of the leave-out mean class rank instrument, which relaxes the functional form assumption of linearity. This instrument simply indicates whether the leave-out mean class rank is higher in math or language. In column (4), we construct the dummy instrument using absolute (i.e., level) differences in leave-out means instead of differences in ranks. Column (5) directly uses the absolute differences in leave-out means as an instrument, which again implies making a linearity assumption. Finally, column (6) takes into account that children in the 1989 cohort were tested in their first year in secondary school, that is, after tracking. Thus, we construct our baseline class rank instrument for the 1989 cohort separately by track, which addresses the potential concern that differences in the rank of math and language skills may be track-specific.

**Table A8: Different definitions of classroom's comparative skill advantage**

| | Rank Class Leave-Out (Main) | Rank Class | Rank Class Dummy Leave-Out | Level Class Dummy Leave-Out | Level Class Absolute Leave-Out | Rank Class Track-Specific |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Parent comparative skill advantage | 0.110 | 0.096 | 0.094 | 0.099 | 0.082 | 0.122 |
| | (0.047) | (0.029) | (0.057) | (0.051) | (0.044) | (0.054) |
| Further controls | yes | yes | yes | yes | yes | yes |
| F-statistic excluded instrument | 212.58 | 612.56 | 93.24 | 122.53 | 217.96 | 144.55 |
| R-squared | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.015 |
| Observations | 12,268 | 12,268 | 12,268 | 12,268 | 12,268 | 12,268 |

*Notes:* Two-stage least squares regressions. Sample: All matched parent-children observations in the education cohorts of 1982 and 1989. Dependent variable: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. Instruments: Column (1): Difference between the percentile ranks of classroom peers in math and language within a parent's education cohort; column (2): difference between the percentile ranks of full classroom in math and language within a parent's education cohort; column (3): Binary indicator for higher ranked classroom peers (math vs. language) within the parent's education cohort; column (4): Binary indicator for better performing classroom peers (math vs. language); column (5): Test scores in math and language of classroom peers; column (6): Like column (1), but rank of math and language classrooms in the 1989 cohort (where children were sampled in the first year of secondary school) calculated by track, distinguishing between 11 different tracks. Further controls include grandparent education and grandparent social background (all referring to the time when parents took the skill test). All regressions additionally control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the classroom level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset (1982 and 1989 cohort).

**Subsample analysis: Addressing potential violations of the exclusion restriction**

In this section, we address various concerns about potential violations of the exclusion restriction of our IV approach by estimating the IV model based on child-parent matches in subsamples that are arguably less prone to such concerns.

**Addressing correlated intergenerational peer composition**

We start by addressing the concern that peer quality may be correlated across the parent and child generation because of endogenous sorting of children within schools. To this end, we replicate the analysis from Table 6 in one-classroom schools, controlling for skill differences between math and language of children's *classroom* peers (Table A9). While skill differences of children's classroom peers are strongly related to the skill differences of children, they hardly affect the estimated strength of the intergenerational transmission of comparative skill advantages. However, the transmission is less precisely estimated due to the reduction in sample size.

In Table A10, we account in various ways for potential effects of parents' classroom peers on the formation of children's skills that are not running through parent skills. In column (1), we exclude parents who have been classmates in early formal education and whose children are schoolmates today. For children who attend the same school as children of their parents' former classmates, parents' peers could directly affect children's skill development. Reassuringly, the IV estimate in this sample is very similar to our baseline IV estimate in column (5) of Table 5.[62]

In column (2) of Table A10, we further restrict the sample to children whose school is located in a municipality different from the parents' municipality of school attendance. In the further specifications of Table A10, we restrict the sample even further to child-parent matches where children attend a school that is at least 50 (column 3) or 100 (column 4) kilometers away from their parent's former school, or where children attend a school in a different province than the parent's school. Throughout all subsamples, the IV estimates remain sizeable, but fail to capture statistical significance in column (2) ($p$=0.214) and column (5) ($p$=0.282).

[62] A related concern might be that in our full sample we have 365 children for which we observe both parents in our data. In most of these cases, both parents attended the same school or even class. We can address this concern by excluding these 365 children from our sample and estimate the IV model based on a sample of children for which only one parent got sampled in any class of the survey. Our IV results are not affected by this sample restriction.

**Table A9: Controlling for children's school quality (one-classroom schools)**

| | OLS model | First stage IV | Reduced form | Second stage IV | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Parent comparative skill advantage | 0.074 | | | 0.086 | 0.099 | 0.097 |
| | (0.013) | | | (0.071) | (0.075) | (0.075) |
| Classroom comparative skill advantage | | 0.263 | 0.023 | | | |
| | | (0.027) | (0.018) | | | |
| Children's school quality | 0.118 | 0.013 | 0.119 | 0.118 | 0.113 | |
| | (0.015) | (0.018) | (0.015) | (0.015) | (0.015) | |
| Children's school quality | | | | | | 10.427 |
| (absolute) | | | | | | (1.214) |
| Further controls | | | | | yes | yes |
| F-statistic excluded instrument | | | | 97.76 | 86.27 | 86.36 |
| R-squared | 0.02 | 0.04 | 0.01 | 0.02 | 0.03 | 0.03 |
| Observations | 5,620 | 5,620 | 5,620 | 5,620 | 5,620 | 5,620 |

*Notes:* Table replicates Table 6 for children whom we observe in a school with at most 30 grade-six students in a given year; this is our proxy for one-classroom schools, as classroom identifiers are not available in the administrative CITO data. Least squares and two-stage least squares regressions. Sample: All matched parent-children observations in the education cohorts of 1982 and 1989 in school-year combinations with 30 or less total observations; children with missing school information are excluded. Dependent variables: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data in columns (1), (3), (4), (5), and (6); difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort in column (2). Column (1) replicates baseline least squares model (see column (1) of Table 3) in the IV sample. Classroom comparative skill advantage is measured as the difference between the percentile ranks in math and language of parents' classroom peers within a parent's education cohort. Children's school quality (ranks) is measured as the difference between the percentile ranks in math and language of children's school peers in the national test score distribution in a given year. Children's school quality (absolute) is measured as the test-year-standardized test score difference between math and language of children's school peers. Further controls include grandparent education and grandparent social background based on the occupation type of the main breadwinner in the parent household (all referring to the time when parents took the skill test). All regressions additionally control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the classroom level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset (1982 and 1989 cohort).

## Table A10: Regional movers

| | Without children of parent's classmates | Child & parent school not in same municipality | Child & parent school not in same municipality (distance >50 km) | Child & parent school not in same municipality (distance >100 km) | Child & parent school not in same province |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Parent comparative skill advantage | 0.092 | 0.080 | 0.209 | 0.255 | 0.119 |
| | (0.050) | (0.065) | (0.110) | (0.147) | (0.111) |
| Further controls | yes | yes | yes | yes | yes |
| F-statistic excluded instrument | 176.63 | 134.69 | 25.91 | 20.65 | 34.71 |
| R-squared | 0.017 | 0.017 | 0.042 | 0.056 | 0.030 |
| Observations | 10,970 | 6,414 | 1,360 | 585 | 2,311 |

*Notes:* Two-stage least squares regressions in the sample of matched parent-children observations in the education cohorts of 1982 and 1989. Samples: Column (1): Excluding children who attend the same school and whose parents have been classmates in the education cohorts of 1982 and 1989; column (2): as in column (1), while keeping only children whose school is located in a different municipality than the parent's school in the education cohorts of 1982 and 1989; column (3) (column 4): as in column (2), while keeping only children whose school is located in a municipality that is more than 50 km (100 km) away from the municipality of the parent's school in the education cohorts of 1982 and 1989 (using the municipality centroid); column (5): as in column (1), while keeping only children whose school is located in a different province than the parent's school in the education cohorts of 1982 and 1989. Results in columns (2) and (5) contain only children with a valid municipality or province identifier (92.06 percent of the total IV sample). Results in columns (3) and (4) contain only children and parents with available municipality longitude and latitude coordinates (88.52 percent of the total IV sample). Dependent variable: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. The instrument is classroom comparative skill advantage, measured as the difference between the percentile ranks in math and language of parents' classroom peers within a parent's education cohort. Further controls include grandparent education and grandparent social background (all referring to the time when parents took the skill test). All regressions additionally control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the classroom level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset (1982 and 1989 cohort).

**Addressing potential between- or within-school sorting of parents**

Our estimation already accounts for potential sorting of parents to schools or teachers based on factors that similarly affect the formation of math and language skills. However, the estimates might be biased if sorting is based on factors that affect subject-specific skill production over generations within families. Our IV estimation results could be biased upward if, for instance, parents belonging to mathematically gifted families systematically attended schools with more knowledgeable math teachers, or if principals tended to assign parents from mathematically gifted families to teachers with high math knowledge.

Table A11 suggests that subject-specific sorting when parents attended school is unlikely to drive our results. We first address between-school sorting by restricting the sample to students living in rural areas (column 2). In this case, students likely have little choice between different schools, because there is usually only one relevant school in rural areas. The estimated IV effect for students in rural areas is very similar to our baseline effect, reported in column (1). To address the concern of within-school sorting, we focus on a subsample of schools with only one classroom, implying that principals cannot assign students to teachers based on their subject-specific ability or preferences. As shown in column (3), the IV estimate on parent comparative skill advantage in this subsample even tends to be somewhat larger than the baseline estimate. Column (4) shows that our results hold even when we restrict the sample to one-classroom schools in rural areas, simultaneously addressing across-school and within-school sorting. This is remarkable because this restricted sample is only one-third the size of the full sample.

In columns (5) and (6) of Table A11, we show the IV results separately for students in the 1982 cohort, who were tested at the end of primary school, and for students in the 1989 cohort, where testing took place at the beginning of secondary school. While still positive and sizable, the IV estimate in the 1989 cohort is not statistically significant. One plausible explanation is that parents in this cohort took the test in the first year of secondary school (i.e., after tracking), so they had considerable less exposure to peers or teachers than parents in the 1982 cohort. This is also reflected in the weaker first stage in the 1989 cohort.

## Table A11: School sorting in the parent generation

| | Main | Rural schools | One-classroom schools | Rural & one-classroom schools | Cohort 1982 | Cohort 1989 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Parent comparative skill advantage | 0.110 | 0.121 | 0.157 | 0.142 | 0.140 | 0.052 |
| | (0.047) | (0.054) | (0.063) | (0.069) | (0.060) | (0.078) |
| Further controls | yes | yes | yes | yes | yes | yes |
| F-statistic excluded instrument | 212.58 | 139.52 | 158.86 | 116.83 | 163.83 | 45.56 |
| R-squared | 0.016 | 0.020 | 0.010 | 0.021 | 0.015 | 0.019 |
| Observations | 12,268 | 5,525 | 6,648 | 3,670 | 5,841 | 6,427 |

*Notes:* Two-stage least squares regressions. Samples: Column (1): All matched parent-children observations in the education cohorts of 1982 and 1989; column (2): Matched parent-children observations form rural schools in the education cohorts of 1982 and 1989; column (3): Matched parent-children observations from schools with exactly one classroom in the education cohorts of 1982 and 1989; column (4): Matched parent-children observations from rural schools with exactly one classroom in the education cohorts of 1982 and 1989; column (5): All matched parent-children observations in the education cohort of 1982; column (6): All matched parent-children observations in the education cohort of 1989. Dependent variable: Difference between the percentile ranks of linked children's math and language test scores in full sample of children taking the test in a given year based on the administrative data. Parent comparative skill advantage is measured as the difference between the percentile ranks of linked parents' math and language test scores in full sample of parents and nonparents in an education cohort. The instrument is classroom comparative skill advantage, measured as the difference between the percentile ranks in math and language of parents' classroom peers within a parent's education cohort. Further controls include grandparent education and grandparent social background (all referring to the time when parents took the skill test). All regressions further control for parent gender, parent migration background, number of siblings of parents, age of grandparents at the time of parent birth, parent survey indicators, and children test year fixed effects. Standard errors clustered at the classroom level in parentheses. *Data sources:* Administrative data; pooled ITS survey dataset (1982 and 1989 cohort).

## A.4 Appendix for Section 7: Parents' Comparative Skill Advantage on Children's STEM Choices

### Table A12: Parents' comparative skill advantage and STEM choices of parents and children – Narrow STEM definition

| | Parent STEM field of study | Child (all) STEM profile | Child (all) STEM field of study | Child (survey) STEM profile | Child (survey) STEM field of study |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | **Panel A: Full sample** | | | | |
| Parent comparative skill advantage (/10) | 0.0064 | | | 0.0064 | 0.0046 |
| | (0.0009) | | | (0.0012) | (0.0011) |
| Child comparative skill advantage (/10) | | 0.0166 | 0.0115 | | |
| | | (0.0002) | (0.0002) | | |
| Further controls | yes | yes | yes | yes | yes |
| Baseline outcome | 0.183 | 0.232 | 0.220 | 0.250 | 0.221 |
| R-squared | 0.129 | 0.101 | 0.128 | 0.014 | 0.010 |
| Observations | 28,264 | 1,161,303 | 1,161,303 | 28,665 | 28,665 |
| | **Panel B: Male sample** | | | | |
| Parent comparative skill advantage (/10) | 0.0097 | | | 0.0095 | 0.0079 |
| | (0.0016) | | | (0.0020) | (0.0019) |
| Child comparative skill advantage (/10) | | 0.0168 | 0.0159 | | |
| | | (0.0003) | (0.0003) | | |
| Further controls | yes | yes | yes | yes | yes |
| Baseline outcome | 0.311 | 0.349 | 0.364 | 0.379 | 0.364 |
| R-squared | 0.039 | 0.036 | 0.022 | 0.021 | 0.023 |
| Observations | 14,236 | 576,031 | 576,031 | 14,358 | 14,358 |
| | **Panel C: Female sample** | | | | |
| Parent comparative skill advantage (/10) | 0.0038 | | | 0.0032 | 0.0014 |
| | (0.0008) | | | (0.0013) | (0.0010) |
| Child comparative skill advantage (/10) | | 0.0162 | 0.0069 | | |
| | | (0.0002) | (0.0002) | | |
| Further controls | yes | yes | yes | yes | yes |
| Baseline outcome | 0.053 | 0.117 | 0.078 | 0.120 | 0.078 |
| R-squared | 0.012 | 0.042 | 0.010 | 0.016 | 0.018 |
| Observations | 14,028 | 585,272 | 585,272 | 14,307 | 14,307 |

*Notes:* Least squares regressions. Sample: Pooled sample of all individuals (parents and nonparents) in the three survey cohorts in column (1); pooled sample of all students that took the CITO test at the end of primary education between 2006-2019 for who we observe both their course- and study profile choice in columns (2) and (3). Children of individuals in the first survey cohort (1977) for whom we observe both their course- and study profile choice in columns (4) and (5). Dependent variables: Binary variable taking a value of 1 if surveyed individuals' highest obtained degree 30 years after participating in the survey is in a STEM (Science, Technology, Engineering, and Mathematics) field (column 1); Binary variable indicating the choice of a STEM course profile at secondary school in columns (2) and (4); binary variable indicating the choice of a STEM field of study after secondary school in columns (3) and (5). Students are designated as following a STEM-course profile if they take the Technical course profile (low academic track) or the Nature & Technical course profile (middle/high academic track). STEM study choice is determined based on the 1-digit ISCED97 fields of education classification (UNESCO, 2003), where study programs categorized as Science, Mathematics and Computing, Engineering, Manufacturing and Construction, were classified as a STEM choice of study. Baseline values are calculated based on observations with non-missing information on STEM choices. Regressions in column (1) control for individual's gender, migration background, number of siblings, age of parents at the time of individual's birth, survey indicators, education and social background of grandparents, as well as municipality fixed effects. Regressions in columns (4) and (5) additionally include child test year fixed effects. Regressions in columns (2) and (3) control for student gender, migration background, student test year and school fixed effects. Standard errors (in parentheses) are clustered at the individual level in column (1), at the school level in columns (2) and (3), and at the parent level in columns (4) and (5). *Data sources*: Administrative data; pooled ITS survey database.

Halle Institute for Economic Research –
Member of the Leibniz Association

Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany

Postal Adress: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

Leibniz
Association