sociological science

How Robust Are Country Rankings in Educational Mobility?

Ely Strömberg,[a] Per Engzell[b]

a) University of Amsterdam; b) University College London

Abstract: We investigate the impact of analytical choices on country comparisons in intergenerational educational mobility using a multiverse approach. A literature survey gives rise to 2,880 plausible ways of measuring educational mobility, which we apply to European Social Survey data from 16 countries. Although some countries consistently appear at the top or bottom of the mobility rankings, most show substantial variation. Beyond our methodological contribution, we report two substantive findings. First, some countries often characterized as low-mobility emerge as matching or surpassing the egalitarian Nordic countries, reinforcing the view that wider mobility differences cannot be attributed solely to the education system but must be sought elsewhere, such as the labor market. Second, the choice of parameter—such as regression coefficients, correlations, or categorical measures—is the single most influential factor that shifts country rankings. As different parameters carry distinct theoretical meanings, researchers should treat parameter choice not merely as a robustness check but as an opportunity to test and refine competing theories.

Keywords: educational inequality; educational mobility; multiverse analysis; social mobility; stratification; welfare states

Reproducibility Package: The microdata underlying our analyses are available to download from the European Social Survey. Code necessary to reproduce the results is available at: https://doi.org/10.17605/OSF.IO/VCDSX

Citation: Strömberg, Ely, Per Engzell. 2025. "How Robust Are Country Rankings in Educational Mobility?" Sociological Science 12: 891-922.

Received: July 9, 2025

Accepted: October 13, 2025

Published: December 11, 2025

Editor(s): Arnout van de Rijt, Kristian B. Karlson

DOI: 10.15195/v12.a36

INTERGENERATIONAL educational mobility is a measure of the degree to which children's educational attainment correlates with that of their parents. In a society with perfect educational mobility, children's level of education is completely independent of that of their parents. The opposite of mobility is persistence, where the attainment of parents plays a significant role in the level of education reached by their children. Educational mobility is widely viewed as a desirable goal, being an indicator of equality of opportunity (Breen and Jonsson 2005). Hence, a large literature attempts to understand its variation across place and time in an effort to identify institutional factors that promote or hinder that goal.

Like in all social science, a challenge in this field is that small and seemingly inconsequential choices of data and specification can exert a large influence on results. This problem of researcher degrees of freedom is not new. In a now classic review on how parents influence their children's attainments, Haveman and Wolfe (1995, p. 1855) noted how comparability was hampered by

> substantial variation in estimation methods ... distressingly small overlap in the explanatory variables included in the models ... large variation in specification of the variables designed to indicate the same

> phenomenon (determinant) across the studies; … inconsistency among the studies in reporting effects for entire samples as opposed to specific subsamples of the population … and substantial variation across the studies in the specification of the outcome variable of interest.

To achieve comparability, researchers often pare down the model to the simplest possible comparison: a single parameter of association between the same educational outcome for parent and child. Even so, choices abound. Should we measure education as years of schooling, credentials attained, or something else? If years of schooling, are we asking about the actual number of years spent studying or the minimum needed to reach a given level? If credentials, how do we reconcile the often large differences between education systems in different countries? Are we interested in the father's education, mother's education, or both? If both, how do we best combine information about the two? What parameter do we use to estimate the association? Even if we manage to fully harmonize data in a given comparison, there might always be a different set of choices that would have yielded a different ranking.

Against this background, it becomes necessary to investigate to what degree analytical choices influence country comparisons in educational mobility. We approach this question through the method of multiverse analysis (Engzell and Mood 2023; Young and Cumberworth 2025). This method starts from the premise that many of the alternatives a researcher faces are equally plausible and can all be justified post hoc. Pressured by theoretical expectations, there is a temptation to veer toward combinations of choices that yield a clean or easily interpretable result. A multiverse analysis removes this pressure by laying bare the full range of possible conclusions that emerge from a given data set and question. Instead of trying to reach a single preferred specification, it records all alternative decisions and performs the analysis of interest under all possible combinations thereof.

Much debate on cross-national differences in educational mobility has centered on the idea of a "Nordic exceptionalism." Early comparative studies helped establish the view that the Nordic countries stand out for their egalitarian school systems and relatively high levels of educational mobility (Erikson and Jonsson 1996; Shavit and Blossfeld, 1993). However, more recent work has complicated this picture. Although the Nordic countries consistently place high in studies of income mobility (Blanden 2013; Bratberg et al. 2017; Engzell and Mood 2023), evidence of a Nordic advantage in educational or occupational mobility is far less clear (Gregg et al. 2017; Karlson and Birkelund 2024; Landersø and Heckman 2017). This discrepancy, where strong income mobility is combined with middling educational or occupational mobility, has been described as a mobility paradox (Breen, Mood, and Jonsson 2016; Karlson 2021), raising broader questions about how institutions shape intergenerational outcomes.

We focus on the case of Europe, where one high-quality data set, the European Social Survey (ESS), allows us to measure education in the same way across 16 societies. In constructing our multiverse, we carry out a systematic literature search to derive relevant dimensions of variation. This lets us identify 2,880 plausible ways of measuring educational mobility and examine how results vary across them. There is wide variation in the estimated levels of mobility, as well as country rankings, across specifications. Our findings regarding a possible Nordic exceptionalism remain ambiguous: while Nordic countries tend to place in the upper half among the countries examined, so too do several countries more commonly characterized as low mobility, such as the UK or Germany. A more consistent divide in our data is therefore between Western European countries that are more mobile and their less mobile Southern and Central European counterparts.

We also go further and test the stability of rankings in relation to various model components, asking which ones exert the greatest influence on results. It turns out that different parameters of association—regression coefficients, correlations, log odds ratios, and rank correlations—are the most important contributor and can dramatically shift the rank of individual countries. We argue that different parameters imply different constructs and speak to different questions, which suggests the need for theory in interpreting results. Other model components, such as gender differences, do not significantly affect mobility rankings despite being stressed in previous research. Strikingly, individual model components do not contribute much in isolation, implying that much variation stems from second- and higher-order interactions between model components, which makes variation challenging to predict and interpret.

These results call into question a simplistic view that conflates different dimensions of mobility (education, occupation, and income) and treats them as primarily determined by educational systems. Even when focusing on educational mobility alone, model variation emerges as an irreducible source of uncertainty in country rankings. Crucially, this is not merely a statistical nuisance: rankings can shift substantially depending on the choice of association parameter, each of which carries distinct substantive meanings. We should not expect rankings to align across different dimensions of mobility, nor assume that country positions are stable across alternative operationalizations of one dimension. At a minimum, researchers should consider multiple parameters and, if favoring one, justify their choice theoretically. More ambitiously, variation across measures and models should be treated as informative in itself: an opportunity to explain their limited overlap and ultimately challenge or reformulate existing theory.

Previous Literature

The question of robustness in educational mobility research was recently brought to the fore by a heated exchange comparing Denmark and the United States. In both scholarly literature and public debate, Denmark is often portrayed as a model egalitarian welfare state, with substantial public investment aimed at equalizing opportunities from early childhood. In contrast, the United States is typically seen as a society of entrenched inequality, where life outcomes are largely determined at birth. A study by Landersø and Heckman (2017) questioned this conventional wisdom by claiming that the two countries are similar in their level of educational mobility. At stake are important policy implications. If education is shown to drive economic mobility, it can be viewed as a successful policy lever for promoting equality of opportunity. If not, it suggests that achieving economic mobility may require other means, such as labor market reform or redistributive policies.

Soon followed a rebuke and a subsequent exchange involving several other researchers (Andrade and Thomsen 2018, 2021; Karlson 2021; Thomsen et al. 2025), each presenting alternative analyses leading to different conclusions. However, without a principled way of exploring the model universe, exchanges like this risk generating more heat than light. Each side can provide ever new analyses supporting their preferred conclusion and defend the assumptions behind them. A multiverse analysis offers a principled way out: rather than making arbitrary analytical decisions, researchers identify all reasonable choices and evaluate them within a common framework. This approach not only makes explicit the extent of model dependence but also places individual results in the context of the broader distribution of possible outcomes. In doing so, it shifts the debate from defending single specifications to understanding patterns of variation across specifications and what this implies for theories of intergenerational mobility.

Although the United States does not appear in our sample of countries, its closest European counterpart is arguably the UK. Both represent "liberal" welfare regimes (Esping-Andersen 1990), and combine high levels of social stratification with a formally open education system, in which the earliest formal branching occurs relatively late, in England with General Certificate of Secondary Education (GCSE) examinations around age 16 (Schneider 2008a). Most children attend state-funded schools that are designed to accommodate students of all ability levels. Nonetheless, selective grammar schools continue to operate in some areas (Burgess, Crawford, and Macmillan 2018), admitting students based on an entrance exam at age 11, whereas fee-paying independent schools provide an alternative pathway for the wealthy (Henderson et al. 2020). Private schools, despite educating only a small minority, remain overrepresented among social elites (Reeves et al. 2017).

At the same time, the view of the UK as a low-mobility regime stems less from comparative evidence on education than from cultural narratives, findings on income mobility, and portrayals of elite institutions. A more ideal-typical case of a rigid education system may therefore be Germany, which represents a "conservative" welfare regime (Esping-Andersen 1990), with strong differentiation between academic and vocational tracks and where crucial educational decisions are made as early as age 10 (Schneider 2008b; Van de Werfhorst and Mijs 2010). In different ways, then, the UK and Germany provide strategic comparisons for testing the notion of Nordic exceptionalism: the UK as a culturally salient "class society" and the closest European analog to the United States, and Germany as a clearer case of low *educational* mobility, epitomized by a strong vocational−academic divide and rigid tracking that channels opportunities at an early age.

Various studies have produced country rankings in educational mobility, covering 17 countries in Chevalier, Denny, and McMahon (2009), 19 in Pfeffer (2008), 20 in Liu and Ding (2020), 42 in Hertz et al. (2008), and 185 in Narayan et al. (2018). A consistent finding is that mobility tends to be higher in more economically equal societies, such as the Nordic welfare states, and lower in less equal contexts, such as parts of Southern and Central Europe and Latin America. However, at the level of individual countries, the results are less stable. Nordic countries usually appear among the more mobile, but Sweden ranks only mid-distribution in Hertz et al. (2008), while Norway is among the least mobile in Pfeffer (2008). Germany also shifts dramatically across studies, classified as highly immobile in Pfeffer (2008) yet among the most mobile in Liu and Ding (2020). The UK is often ranked somewhere around the middle (Chevalier et al. 2009; Liu and Ding, 2020), but its position, too, appears sensitive to methodological choices.[1]

A crucial but understudied question is how to construct the model space. If reproducing a study, one can combine the original model with those of published comments and rejoinders (Muñoz and Young, 2018), or investigate which models the authors have used in their previous research in the same field (Steegen et al. 2016). One can also use own experience to map a set of reasonable analytical choices (Simonsohn, Simmons, and Nelson 2020b; Young and Holsteen 2017). Other multiverse studies have taken a many-analysts approach, letting several research teams devise a way to answer the same question and then mapping the factorial combinations of all choices involved (e.g., Schweinsberg et al. 2021). Most of the examples above replicate one previous finding. Here, our aim is not to replicate a given study but to address a whole research field. To arrive at a plausible set of specifications, we therefore perform a systematic literature search to assess the range of model variation in the field.

Although our analysis focuses on model uncertainty, we do not wish to downplay other sources of variation. Estimation bias (e.g., due to missing data or measurement error) can be important, particularly when different data sources are used (Engzell and Jonsson 2015). Sampling variation also matters, as each estimate is accompanied by a confidence interval that may depend on the chosen model and metric (Mogstad et al. 2024). We abstract from sampling error in our main analysis, as we believe there is value in demonstrating the influence of model variation independent of other forms of uncertainty. Supplementary analyses reported in a separate section show that while sampling variation adds further uncertainty to country rankings, it is small compared to model variation and does not alter the overall patterns or conclusions of our study.

Theoretical Considerations

Then, why do different methods and models produce divergent results? A common view is to treat such variation as analogous to sampling error: mere statistical noise that weakens confidence in results. However, this perspective is both unrealistic and limiting. Variation is not only inevitable but also often theoretically meaningful: models do not simply estimate the same quantity with error but rather capture different facets of the phenomenon. From this perspective, variation becomes a resource rather than a flaw. As Engzell and Mood (2023) argue, a theory-informed multiverse analysis can use this heterogeneity to clarify why results diverge and what those divergences reveal about the mechanisms of stratification.

In studying educational mobility, one model may highlight the returns to human capital (Becker and Tomes 1979), another its signaling value (Arrow 1973; Spence 1973), and a third the positional nature of schooling (Hirsch 1976; Shavit and Park 2016). Looking beyond education, findings of a mobility "paradox" become less paradoxical once we recognize that education, occupation, and income capture distinct phenomena, each shaped by different processes. Occupational attainment reflects not only education but also school-to-work linkages, local industrial structures, the role of informal networks, and many other factors that vary across societies and over time (Bernardi and Ballarino 2016). Income mobility, in turn, is additionally conditioned by wage structures within and across occupations, and features of the tax and transfer system (Breen et al. 2016; Landersø and Heckman 2017). Although we stop short of a full theoretical exposition of our results, we use the remainder of this section to draw some relevant distinctions.

Human capital theory views education as a quantitative asset of the individual, just like other forms of capital such as wealth. Asking respondents about the number of years they spent studying will produce a continuous measure of education that is easy to understand and use. In theory, the measure is comparable across countries, with one year of schooling carrying the same meaning (a full year spent studying) regardless of national education system. However, this assumes an equal increase in skills and knowledge for each additional year of schooling and that all types of skills hold similar value (Braun and Müller 1997; Schneider 2009), assumptions that can be questioned. An alternative way of measuring years of schooling is therefore to assign theoretical years based on the typical time necessary to reach a given qualification.

In signaling theory, the education acquired is not so much a good in itself, as a signal of the underlying skills that allowed a person to reach a given level of education. To employers, the most readily observed signal is usally not years of schooling but an attained degree. In some systems, all students follow a straight path through programmes of increasing complexity, whereas in others,there is a horizontal differentiation with multiple different tracks ranging from theoretical to vocational. Measuring these national levels or programmes can result in categories that are often quite heterogeneous and country specific. To harmonize these on a common scale, international standards exist. One such standard is the ISCED, which we will use below.

Recent research argues that as a result of massive expansion of higher education, schooling increasingly serves as a positional good (Bol 2015; Bukodi and Goldthorpe 2016). Employers may judge applicants not on whether they have met some minimum qualification, but on their relative standing among applicants. The main gain from educational investment is then neither skills nor signals, but the positioning of oneself above one's peers. This points to percentile ranking, which situates people within the education distribution of their own age and cohort. Each of these constructs maps onto different ways of measuring education, which in turn imply different parameters, as we turn to next.

Measuring Educational Mobility

Grounded in the above theoretical motivations, we estimate five different parameters of association: regression and correlation coefficients in years of schooling, log odds ratios between educational levels, unidiff coefficients that aggregate log odds ratios across several categories, and correlations in educational ranks.

Starting with human capital theory, years of schooling are typically used to estimate a regression or correlation coefficient between parent and child. The regression coefficient is the slope β in a linear equation with child education on the left-hand side and the parent's education on the right. Its interpretation is the additional years of child schooling associated with a one-year increment in parent schooling:

$$Y_t = \alpha + \beta Y_{t-1} + \varepsilon. \tag{1}$$

In Equation 1, Y_t represents the child's years of schooling, Y_{t-1} represents the parent's years of schooling, α is a constant, and ε is a well-behaved error term. The regression coefficient β is given by

$$\beta = \frac{Cov(Y_t, Y_{t-1})}{Var(Y_{t-1})} = Corr(Y_t, Y_{t-1}) \frac{\sqrt{Var(Y_t)}}{\sqrt{Var(Y_{t-1})}}. \tag{2}$$

As Equation 2 makes clear, the regression coefficient depends on the relative dispersion of years of schooling in both generations. For example, if the variance in children's years of schooling is greater than that among parents, a given correlation will imply a higher regression coefficient β. The process of educational expansion typically increases dispersion at first, as new opportunities open up, and then decreases it, as educational attainment reaches a new plateau at higher levels. To avoid conflating changes of composition with the dependence structure between parent and child, one may want to factor out the relative dispersion in both generations (Black and Devereux 2011, p. 1504). One way to do so is by the intergenerational correlation in years of schooling:

$$r = Corr(Y_t, Y_{t-1}) = \frac{Cov(Y_t, Y_{t-1})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t-1})}}. \tag{3}$$

Treating years of schooling as continuous assumes that the relationship between parent's and child's education is monotonic and linear (Blanden 2013, p. 44). This can be questioned, especially in complex education systems with multiple paths. Sociologists have long advocated for conceptualizing education as a series of discrete transitions (Breen and Jonsson 2000; Mare 1980). An alternative way is to estimate probabilities to reach a certain educational level given a level of parental education and without assuming linearity of effects between the different levels. One such measure is the log odds ratio, which is a function of conditional associations between levels of parent and child education. The log odds can be represented as

$$\ln OR_{ij,i'j'} = \ln\left(\frac{p_{ij}/p_{i'j}}{p_{ij'}/p_{i'j'}}\right) = \ln\left(\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}}\right), \tag{4}$$

contrasting, for example, the probability p for a university educated parent's child j of completing university education i as opposed to a lower level i', relative to the corresponding probability for a child of parent who has less than university education j'. A categorical model takes degrees as the outcome and therefore corresponds closest to the signaling view of education. However, credentials may also reflect relevant skills or relative standing, as per the human capital or positional good theories.[2]

One issue with categorical measures is that, as soon we deal with a mobility table composed of more than two categories, the number of possible comparisons grows large. This yields several disparate coefficients instead of a single summary measure. To address this, researchers often employ log-multiplicative models known as "unidiff" or uniform difference models (Erikson and Goldthorpe 1992; Xie 1992). Such models impose a proportionality constraint on how log odds ratios vary across tables (e.g., countries). Specifically, it assumes that the pattern of relative mobility captured by the log odds ratios is structurally invariant, differing only by a scalar multiplier that varies between contexts and reflects the overall level of mobility. In the simplest form, this includes a term for excess cases along the diagonal, reflecting immobility between parent and child categories, while allowing the relative mobility level to vary across countries.

A positional view of education instead implies that the rank-order correlation should be the preferred measure. This parameter is equivalent to the correlation coefficient r above, with the distinction that each variable, instead of measured as years of schooling, is transformed to percentile ranks:

$$\rho = Corr(R_t, R_{t-1}) = \frac{Cov(R_t, R_{t-1})}{\sqrt{Var(R_t)}\sqrt{Var(R_{t-1})}}, \quad (5)$$

where $R(Y) \in \{1, 2, \ldots, 100\}$ represents the rank transform. The distribution of percentiles is uniform, so this transformation also equalizes the dispersion of education in both generations. Hence, the distinction between regression and correlation coefficients that arises with years of schooling does not matter here.

Mapping the Literature

In addition to parameter selection, many other choices need to be made. To map this space of variation, we conduct a literature search. We searched Scopus and Web of Science, looking for peer-reviewed articles published in the past 20 years. Specifically, we used the following search terms for Scopus:

(TITLE ("EDUCATIONAL FLUIDITY") OR TITLE ("EDUCATIONAL MOBILITY") OR TITLE ("EDUCATIONAL TRANSMISSION") OR TITLE ("INTERGENERATIONAL TRANSMISSION OF EDUCATION") AND PUBYEAR > 2002) AND (LIMIT TO (DOCTYPE, "AR"))

And for Web of Science:

(TI= ("EDUCATIONAL MOBILITY") OR TI=("EDUCATIONAL FLUIDITY") OR TI=("EDUCATIONAL TRANSMISSION") OR TI=("INTERGENERATIONAL TRANSMISSION OF EDUCATION") AND DT=(ARTICLE)) AND PY=(2002-2022)

The search in Scopus resulted in 141 articles and Web of Science gave 143 results totaling 284, of which 122 were duplicates, resulting in 162 unique hits. Screening of abstracts resulted in 69 articles that appeared to measure relative educational mobility. Reading of articles removed 29 articles, mainly because mobility was not

Table 1: Variation in analytical choices from article search.

Dimension	Alternatives
Sample selection	Age ranges and respective countries in which they were used: 24–65 (GR), 25–35 (2x Europe & CA), 25–65 (Europe), 25–69 (IS), 25–70 (2x CA), 25–74 (CA), 25+ (Europe), 27–46 (AT) 27–47 (SE), 28–55 (FR), 29–33 (DK & US), 30+ (DK). Four articles stated that they excluded foreign-born respondents. Three articles looked only at sons, four looked at only daughters, 18 looked at both separately, and eight looked at both separately and pooled.
Child education	Nine articles used actual years of schooling, seven used years of schooling calculated from highest degree attained, three used dichotomous measures of a yes/no nature, 15 used a categorical measure of highest degree attained, and two used highest degree calculated from years of schooling.
Parent education	Six articles used actual years, six recoded level to theoretical years, two recoded ISCED to theoretical years, three used dichotomous measures, 11 used levels of education, five recoded actual years to levels, and one used ISCED.
Which parent?	10 articles used dominant parental education, seven used only fathers' education, three used only mother's education, four used an average measure, six used both in the same analysis, and eight used both in separate analyses.
Parameter	11 articles presented regression coefficients, seven presented regression coefficients and correlations, four presented odds ratios, two presented risk ratios, nine presented transitions matrices, five presented predicted probabilities, and five articles used other types of measures not common in the literature.

measured directly but only used as a covariate in models but also due to advanced multigenerational models or limited information about methods. The final 40 articles were classified according to six dimensions of variation: sample exclusion criteria such as age restrictions and treatment of foreign-born, child education measure, parent education measure, method of choosing which parent to measure, and parameter of association. Table 1 shows the results of this search.

Defining the Multiverse

Informed by this review of the literature, we specify the dimensions of variation that we study. Table 2 shows the model components we vary, following from the results of the literature review in Table 1.

Sample Selection

As shown in Table 1, previous studies used different age ranges. We vary the age restrictions according to the 12 ranges shown in Table 2. Ranges vary from very limited (29−33) to very broad (25+). Age restrictions are commonly discussed as a way to exclude young people who have not had the opportunity to finish their studies. We also run all models both including and excluding foreign-born respondents. As shown in Table 1, this issue was only discussed in four articles, the argument for exclusion being that inclusion leads to bias as one is partly measuring the effects of education systems of other countries. One may also choose to study sons or daughters, separately or pooled. Men and women have seen different trends in education, with women's attainment outpacing and surpassing that of men in

Table 2: Variation in analytical choices considered in this study.

Dimension	Alternatives
Sample selection	Studied countries were: Belgium, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, the Netherlands, Norway, Poland, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom. Age ranges were: 24–65, 25–35, 25–65, 25–69, 25–70, 25–74, 25+, 27–46, 27–47, 28–55, 29–33, 30+. Exclusion or inclusion of foreign-born respondents was varied. Analyses run on men and women separately.
Child education	Measured with: Actual years of schooling top coded at 25 or 30 years. Theoretical years of schooling calculated from ISCED levels. Relative actual years of schooling based on top-coded absolute measures. Relative theoretical years of schooling calculated from ISCED levels. Primary/secondary/tertiary calculated from ISCED levels and ES-ISCED levels.
Parent education	Measured with: Theoretical years of schooling calculated from ISCED levels. Primary/ secondary/tertiary calculated from ISCED levels and ES-ISCED levels. Relative theoretical years calculated from ISCED levels.
Which parent?	For continuous measures only father, only mother, average, and dominant parental attainment. For categorical measures: only father, only mother, and dominant parental attainment.
Parameter	Regression coefficients, correlations, log odds, unidiff, and rank correlations.

European countries. It is therefore likely that patterns may differ by gender. We consider men and women respondents separately.

Child Education

Multiple measures of respondent education were found in the article search. Most common was a categorical measure of highest degree attained, with varying numbers of categories. Sometimes dichotomous measures were used, such as whether the respondent attained tertiary education. Actual years of schooling were also common. In addition, the theoretical years calculated from the highest degree were used and also the highest degree attained calculated from the actual years of schooling. For continuous measures, we code child education as actual years of schooling, top coded at 25 or 30 years, and as theoretical years of schooling, described in greater detail below. These measures are in turn standardized to calculate correlations and recoded into positional variables to calculate relative educational attainment. We also construct categorical indicators encompassing either two or three educational levels, as described below.

Parent Education

Concerning the choice of measure for parental attainment, the operationalizations are similar as for child education and display a similar range. Categorical measures of education level were the most common. Five studies recoded actual years of schooling to levels, and one study used ISCED levels. Three studies used dichotomous measures. Continuous measures were also used, both actual and theoretical years of schooling, two articles stated that they recoded ISCED levels to theoretical years. For continuous measures, we code parental attainment as theoretical years of schooling, which are standardized and recoded into a positional version as for the child measures. Like for the child's education, we also construct alternative categorical measures of two or three levels of education.

Which Parent?

Another question is how to treat information on mothers and fathers. Historically, parental education has been measured with only fathers' education. As societies have become more gender equal and women participate in education and the labor market on equal terms, it becomes questionable to use only the father. Erikson (1984) proposed a "dominance principle," where the highest value of any parent is chosen, regardless of if it is held by a father or mother. Averaging parental education can be a way to better capture the contribution of both parents to children's educational outcomes but combining their education in interaction variables has also been proposed (Jerrim and Micklewright 2011). Still, researchers often focus on fathers due to convention or data availability (Beller 2009; Thaning and Hällsten 2020). In our article search, dominant parental education was most common, but averages or the education of either parent were also used. We run all models with mother's and father's education separately and dominance coded. To keep one estimate of interest per model, we refrain from using both parents in the same model.[3] For continuous measures such as years of schooling, we include the average of both parents.

Association Parameter

The most common parameter in previous studies was regression coefficients, often together with correlations. In addition, odds and risk ratios, predicted probabilities, and transition matrices were used. For absolute continuous variables, we calculate regression coefficients and correlations. For positional continuous variables, we only calculate regression coefficients, which are, in principle, equivalent to correlations as the variance of a uniform variable is constant. For categorical measures, we calculate log odds. Some of the analytical choices found in the search and presented in Table 1 such as transition matrices are not included in this study, even though they occur in the literature search. This is for the same reason that we exclude two-parent analyses: we want to retain a single parameter per specification. When transition matrices occurred in the literature, this was typically as a complement to simpler summary measures.

Data

We use data from the ESS rounds 1−10. ESS is a cross-country survey performed using face-to-face interviews biennially since 2002, now comprising 10 rounds. Thirty-eight countries have participated over the years, with 14 of these participating in all 10 rounds. Samples are representative of the national population aged 15 or above, and response rates per country range between 40 percent and 70 percent. For more detailed information, see Table A1 in the online supplement.

Table 3: Descriptive statistics of recoded educational variables.

New Variable	Based on	Range	Mean	N	Rounds
Actual years of schooling (max 25)	eduyrs	0–25	12.8	294,945	1–10
Actual years of schooling (max 30)	eduyrs	0–30	12.9	295,722	1–10
Theoretical years of schooling	edulvla and edulvlb	0–24.25	12.7	286,418	1–10
Father's theoretical years of schooling	edulvlfa and edulvlfb	0–24.25	10.5	243,235	1–10
Mother's theoretical years of schooling	edulvlma and edulvlmb	0–24.25	9.9	252,697	1–10
Average parental relative theoretical years of schooling	Theoretical years father and mother	0–24.25	10.1	262,925	1–10
Dominant parental theoretical years of schooling	Theoretical years father and mother	0–24.25	11.1	262,925	1–10
Educational level (ISCED)	edulvla and edulvlb	Prim/sec/tert		295,152	1–10
Educational level (ES-ISCED)	eisced	Prim/sec/tert		298,445	1–10
Father's educational level (ISCED)	edulvlfa and edulvlfb	Prim/sec/tert		239,717	1–10
Mother's educational level (ISCED)	edulvlma and edulvlmb	Prim/sec/tert		246,734	1–10
Dominant parental educational level (ISCED)	edulvl3f and edulvl3m	Prim/sec/tert		256,628	1–10
Father's educational level (ES-ISCED)	eiscedf	Prim/sec/tert		165,070	4–10
Mother's educational level (ES-ISCED)	eiscedm	Prim/sec/tert		170,988	4–10
Dominant parental educational level (ES-ISCED)	eisced3f and eisced3m	Prim/sec/tert		175,221	4–10
Relative actual years of schooling (max 25)	Actual years of schooling (max 25)	1–100	46.0	294,945	1–10
Relative actual years of schooling (max 30)	Actual years of schooling (max 30)	1–100	46.0	295,722	1–10
Relative theoretical years of schooling	Theoretical years of schooling	1–100	44.2	286,418	1–10
Father's relative theoretical years of schooling	Father's theoretical years of schooling	1–100	40.8	243,235	1–10
Mother's relative theoretical years of schooling	Mother's theoretical years of schooling	1–100	40.1	252,697	1–10
Average parental relative theoretical years of schooling	Average parental theoretical years of schooling	1–100	42.2	262,925	1–10
Dominant parental relative theoretical years of schooling	Dominant parental theoretical years of schooling	1–100	41.8	262,925	1–10

To increase the number of observations per country, we only include countries that have been part of eight or more rounds. These are Denmark that have participated in eight rounds; Austria, Czech Republic, and Estonia that have participated in nine rounds; and Belgium, Finland, France, Germany, Hungary, Ireland, the Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden, Switzerland, and the UK that have participated in 10 rounds. However, Austria, the Czech Republic, and Portugal are excluded due to missing or inconsistent educational information.[4] This leaves 16 countries of which all but two have taken part in all rounds. To increase cell counts, we pool all available rounds, meaning that we will estimate 20-year averages instead of specific-year estimates as is common in the previous literature.

Coding of Variables

The fact that ESS contains many different educational variables—and that these have changed over time—adds an additional layer of complexity to the dimensions considered above. Table 3 shows how we handle the different educational variables available in ESS.

The ESS variable of self-reported actual years of schooling *eduyrs* can in theory be used as is. However, the variable contains values up to 60 years, therefore we create two variables by top coding the *eduyrs* variable to max 25 years, and max 30 years to remove extreme values while leaving some space for individuals who re-educate themselves. Higher values are coded as missing, but this only removes 246 respondents with values above 30 or 1023 (246 + 777) above 25 from a total of 299 724 respondents and should therefore in itself have a limited effect on results.

All other educational variables that we use depart from some form of the ISCED. The earliest version in ESS is ISCED-97, eventually supplanted by ISCED-11 from round 5. In addition, the data contain a variant of the ISCED framework developed specifically for the survey, called ES-ISCED (Schneider 2020), to address issues of comparability with actual years of education or the regular ISCED (Schneider 2010). The respective variables in the data set are *edulvla*, which is a five-category measure based on ISCED-97, *edulvlb*, which is a 31-category measure based on ISCED-11, and *eisced*, which contains the ESS adaption ES-ISCED. For parental education, there are similar variables based on ISCED-97, ISCED-11, and *eisced* for both mothers and fathers, but no measure of actual years of schooling.

To create theoretical years of schooling from the various categorical variables, we first use a coding schema by Schneider (2013) that harmonizes ESS codes into ISCED codes. Then, to assign a duration, we rely on country-specific data sets on duration by level that are made available by UNESCO.[5] Multiple programmes of different duration typically sort under the same ISCED code, and in these cases we take the average duration of all programmes classified under the same code (which is why the upper bound appears as a non-integer). When a programme has a range of possible durations we take the minimum. We repeat this procedure for the respondent, their mother, and father. We also create a dominance-coded measure for parents and one containing the average of both parents.

For categorical measures, we distinguish between primary, secondary, and tertiary education.[6] Here, we separate the standard ISCED and the ESS adaptation ES-ISCED (Schneider 2020), as the two sometimes differ. The ES-ISCED framework created for the ESS incorporates hierarchical dimensions and has the advantage of a more even distribution than the ISCED variables (Schneider 2020, p. 4–5). Our mapping from ISCED and ES-ISCED codes into levels is described in Table A2 in the online supplement. When displaying log odds we distinguish tertiary education against all lower levels. Information on ES-ISCED for parents is only available from round four onwards. Parental variables are divided into one variable for fathers, one for mothers, and one following the dominance principle. Due to the ordinal nature of the data, there is no average categorical attainment variable.

Finally, positional variables are created by converting the educational variables into percentile ranks by country. For respondents, we recode actual years of schooling (capped at 25 or 30 years) and the theoretical years from ISCED mappings into positional versions. As there is no information on actual years of schooling for parents, we only recode theoretical years. We create four variables: one for fathers, one for mothers, one with average parental attainment, and one with dominant parental attainment.

Results

The above coding gives rise to five different measures of education for the respondent (three continuous and two categorical) and 10 different measures for the parents (four continuous and six categorical).[7] In addition, we have three sample selection criteria, consisting of 12 age ranges, two options of dealing with foreign-born respondents, and 16 countries. We also run male and female respondents separately totaling in $12 \times 2 \times 2 = 48$ samples. When combined with the five different parameters of association, we have $3 \times 4 \times 48 = 576$ regression coefficients, $3 \times 4 \times 48 = 576$ correlation coefficients, $2 \times 6 \times 48 = 576$ odds ratios,

$3 \times 4 \times 48 = 576$ rank correlations, and $2 \times 6 \times 48 = 576$ unidiff coefficients. For each country, we thus estimate 2,880 alternative models in total and display the results separately by type of coefficient and country. However, to make certain that rankings are not biased by small cell counts, we exclude models where at least one country has less than 500 observations, meaning that the presented number of models often is lower than the full 576 models per country.[8]

We present results graphically using ridge and box plots. The ridge plots visualize the distribution of point estimates by country and make it easy to see whether estimates follow a normal curve and to what extent distributions for the different countries overlap. Even if the size of estimates varies, if a given model choice has the same influence across all countries, the relative ranking of countries would not be affected. Box plots provide complementary information on the distribution of rankings per country. They show whether different model choices affect not only the absolute size of estimates but also the relative rank of countries. We focus on model uncertainty first and explore the more familiar problem of sampling uncertainty in a separate section below. To reduce the influence of sampling uncertainty, we exclude models with small cell counts as described in endnote 8 and also pool all 10 ESS rounds in the same analysis. Descriptive statistics on the distribution of point estimates are shown in Tables A3−A7 and separate results by gender are shown in Figures A1−A5, all in the online supplement.

Regression Coefficients

Figure 1, left panel, shows the distributions of country regression coefficients ordered by median ranking. Distributions overlap to a high degree. Hungary has the highest value at 0.64, and Sweden has the lowest value at 0.13. To see how much we can manipulate regression coefficients by changing analytical choices, the interquartile range (IQR) can be a restrictive measure and the range a more generous one (Table A3 in the online supplement). The average IQR is 0.09, which would not necessarily change our perception about a country's level of mobility, but certainly how it compares to other countries. The average range is 0.29, which would be enough to change whether a country is seen as mobile or not. Taking Denmark as an example, by choosing model, we can present mobile values around 0.15 as well as less mobile values around 0.50, whereas the mean lies at 0.32. Certain countries, such as Ireland, also have many high outlier values. To sum up, regression coefficient models show variation in point estimates, exhibiting sensitivity to analytical choices.

Figure 1, right panel, shows a box plot of the ranking of countries over models, ordered by median rank. The countries seem to be roughly divided into three groups. The Nordic countries together with Ireland are ranked in the top, implying that they are the most mobile. The middle group consists of France, the UK, Germany, Switzerland, and the Netherlands, and in the bottom we find Hungary, Belgium, Slovenia, Spain, and Poland. However, within these groups, there is much variation and overlap. Eight out of 16 countries are ranked first in at least one model, whereas six countries are ranked last. Ireland has the biggest range, being ranked both 1 and 16 with a median of three. Overall, we can see that through

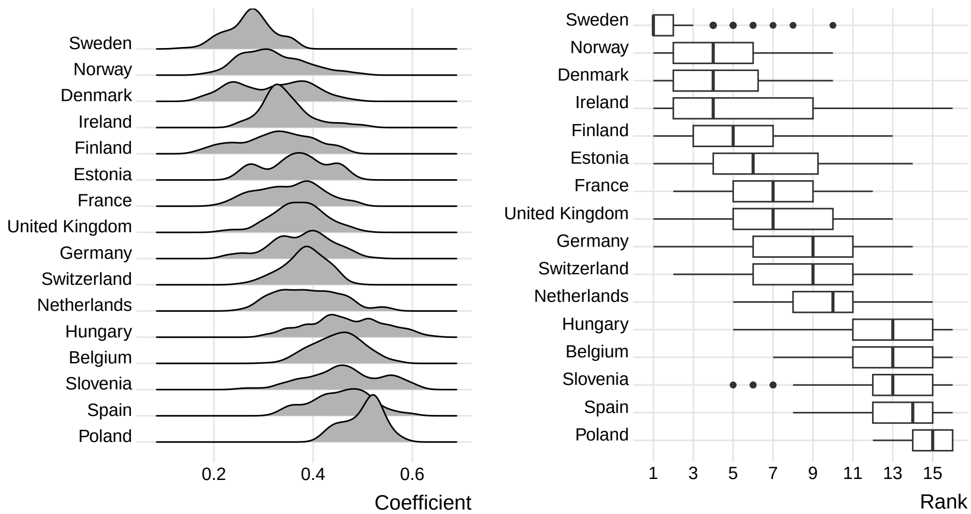

Figure 1: Regression slopes, distribution of coefficients, and country rankings. *Note:* Countries are ordered by the median country rank. 528 models per country.

choosing a specific model, it would be possible to get any of the top 10 countries ranked in the top 3. Although the groups differ, countries in the top and bottom overlap with the middle, and countries in the middle group overlap with both top and bottom, showing substantial influence of analytical choices.

Correlations

Figure 2, left panel, shows the dispersion of correlations by country. The correlation models work the same way as those for regression coefficients, but both respondent and parental education are standardized to factor in educational expansion between parental and respondent generations. Distributions still overlap to a high degree, and values range from 0.16 for Denmark to 0.67 for Hungary. Compared with its low minimum estimate, Denmark has a mean estimate at the higher 0.31 (Table A4 in the online supplement), which is the smallest mean value followed by Germany and the UK, a clear change from the regression models. If we consider the distribution of coefficients as a measure of the importance of model choice, correlation results are comparable to regression coefficient models with an average IQR at 0.08 and an average range at 0.28, meaning model choice can change comparisons and perception of country mobility levels. Finland has a remarkably high range of 0.37, with values quite evenly spread. Some countries such as Sweden have a greater range than in the regression coefficient models, whereas others such as Denmark have a narrower one.

Figure 2, right panel, shows a box plot of the ranking of countries by correlation. Compared with the regression models, there is less grouping in the more mobile half of the distribution. Denmark now seems to be the most mobile with a median ranking of two, and Sweden has dropped from first to the middle of the ranking distribution. Germany has moved up to second highest median rank, from rank 9 to 10 in regression models, appearing more mobile when using correlations. Belgium,

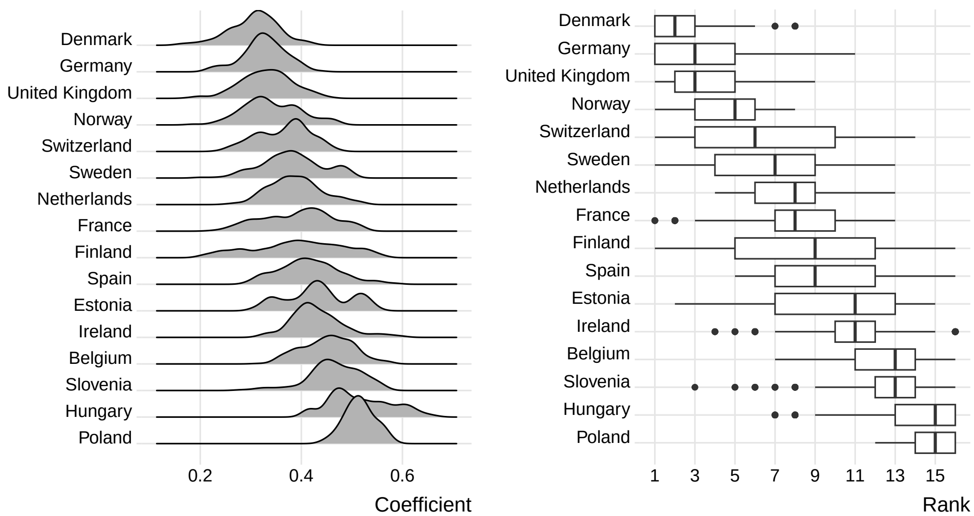

Figure 2: Correlations, distribution of coefficients, and country rankings. *Note:* Countries are ordered by the median country rank. 528 models per country.

Poland, Hungary, and Slovenia are still in the bottom for median ranks, whereas Spain appears more mobile. Only Finland is ranked both highest and lowest. 8 out of 16 countries can be ranked as first by choosing specific models. Sweden and Finland, which were grouped in the top with regression models, are now in the middle of the ranking distribution, which suggests that the distribution is more dispersed in the parent generation and has subsequently contracted with mass higher education in the child generation.[9]

Log Odds Ratios

Figure 3, left panel, shows the distributions of estimates from categorical models with log odds for tertiary education by parental tertiary education. As with the previous models, estimates overlap to a high degree, especially in the top half. However, ranges are visibly smaller for many countries, with highly dense distributions for Belgium and Poland. For Germany, the Netherlands, and Switzerland, distributions are instead non-normally distributed, something which might imply a sensitivity to certain analytical choices. Estimates range from 0.72 (Sweden) to 3.79 (Switzerland) with much overlap in the top half (Table A5 in the online supplement). Estonia has the smallest range at 0.68 and Switzerland has the greatest at 2.69, with the average at 1.30. The average IQR is at 0.29 so the impression of a country's mobility level might not change much over models, but seeing the high overlap, it can still change country comparisons.

Figure 3, right panel, shows a box plot of the ranking of countries over categorical models. The overlap in the top half of the left panel is here visible as clear overlap in rankings of the top seven countries. Sweden now appears the most mobile, trailed by Germany, with a median rank of 2 and 3, respectively. We can note that Sweden's position when calculating log odds ratios is similar to that from the regression coefficient models, whereas Germany's position instead is similar

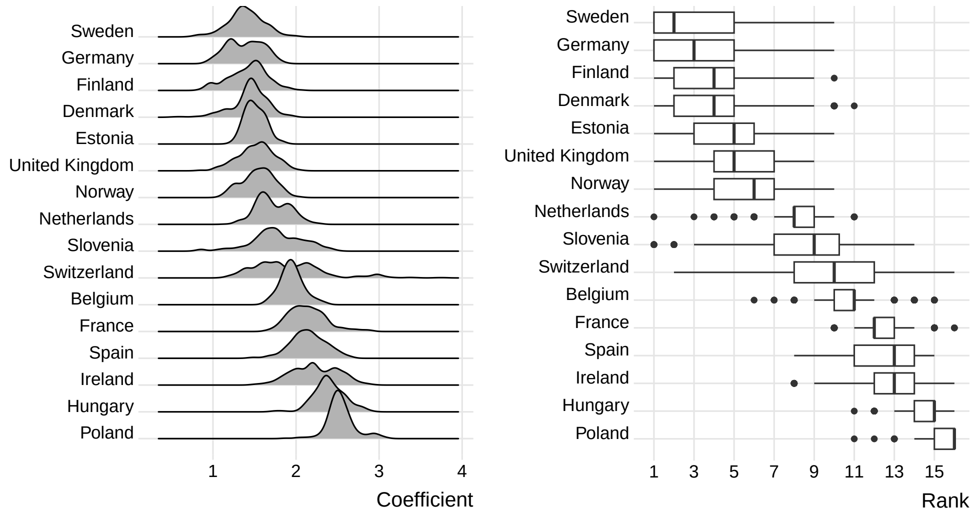

Figure 3: Log odds ratios, distribution of coefficients, and country rankings. *Note:* Countries are ordered by the median country rank. 508 models per country.

to that of the correlation models. In the bottom, we find Poland and Hungary. However, Slovenia has risen toward the middle of the ranking distribution, with a median rank of nine and outliers at 1 and 2. As shown in the left figure, Slovenia and Switzerland show much dispersion in log odds ratios, whereas other countries such as Belgium and Poland show very small dispersion. Switzerland is ranked as both 2 and 16 in different models, but otherwise ranges of ranks are smaller than in the previous models, and there is a visible order to the country rankings of the bottom half, even though overlap is still high.

Unidiff Coefficients

Figure 4, left panel, shows the distributions of estimates from unidiff models, standardized to have mean 0 and standard deviation 1 across each specification.[10] As with the previous models, estimates overlap, especially in the middle. However, there is a clear order of countries, with some having very tight distributions, especially Belgium and Ireland. The UK, Sweden, the Netherlands, and Germany show highly nonnormal distributions, implying sensitivity to analytical choices. Estimates range from -2.38 for Sweden to 2.32 for Hungary with much overlap in the top half (Table A6 in the online supplement). Hungary has the smallest range at 1.26 and Sweden has the greatest at 3.03, with the average at 2.05. The average IQR is at 0.47 with the highest being Sweden at 0.83 and the lowest France and Hungary at 0.31.

Figure 4, right panel, shows a box plot of the ranking of countries over categorical models. The UK now appears the most mobile, trailed by Finland and Sweden, both with a median rank of three. Hungary and Poland are clearly placed in the bottom, followed by Switzerland and Slovenia that however both have long tails with many outliers. Germany, who was ranked second in the log odds models, is now ranked 11, but with a range from 1 to 14. Sweden, ranked third, and Slovenia,

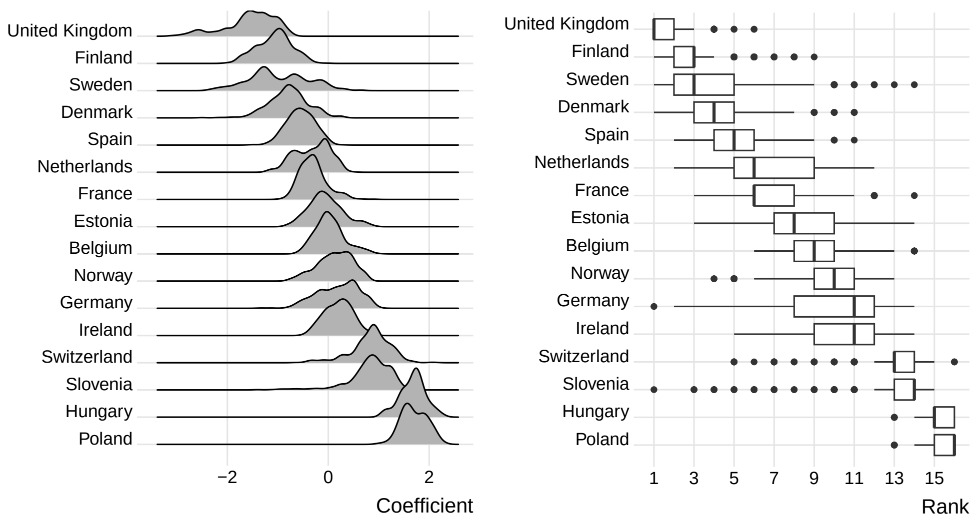

Figure 4: Unidiff, distribution of coefficients, and country rankings. *Note:* Countries are ordered by the median country rank. 482 models per country.

ranked 14, have the same range, pointing to all three countries being sensitive to analytical choices in unidiff analyses. Overall, many countries show a wide range when taking outliers into account, but with otherwise quite dense distributions. Therefore, there appears to be a clearer ranking than for other parameters, implying that unidiff analyses might be less sensitive to the analytical choices in our multiverse. Although we do not examine how unidiff results depend on the granularity of categories, supplementary analyses in Karlson (2021) find that country rankings are not overly sensitive to this analytical choice either, consistent with our conclusions.

Rank Correlations

The last model type is positional educational attainment, where regression coefficients are calculated based on relative educational attainment instead of absolute. Figure 5, left panel, shows a ridge plot of regression coefficient values from the positional models. Looking at the top, we find the UK with a mean of 0.28, coming out as the most mobile (Table A7 in the online supplement). The other countries in the top half overlap to such a degree that it is hard to distinguish, which is more mobile; however, the distribution of Finland stands out with an exceptionally flat distribution with a range of 0.34 and the lowest estimate at 0.17. Hungary has the highest value at 0.64 and is again found in the bottom together with Poland and Slovenia. The average range is 0.26 and the average IQR is 0.08, meaning that using relative education does not remove the effect of analytical choices.

Figure 5, right panel, shows country rankings under positional models. The pattern diverges sharply from absolute measures: the UK holds the highest median rank, followed by France, an unexpected result given that France has consistently ranked seventh or lower using absolute education. Sweden and Estonia drop to the bottom half (9th and 12th, respectively), whereas Slovenia, Poland, and Hungary

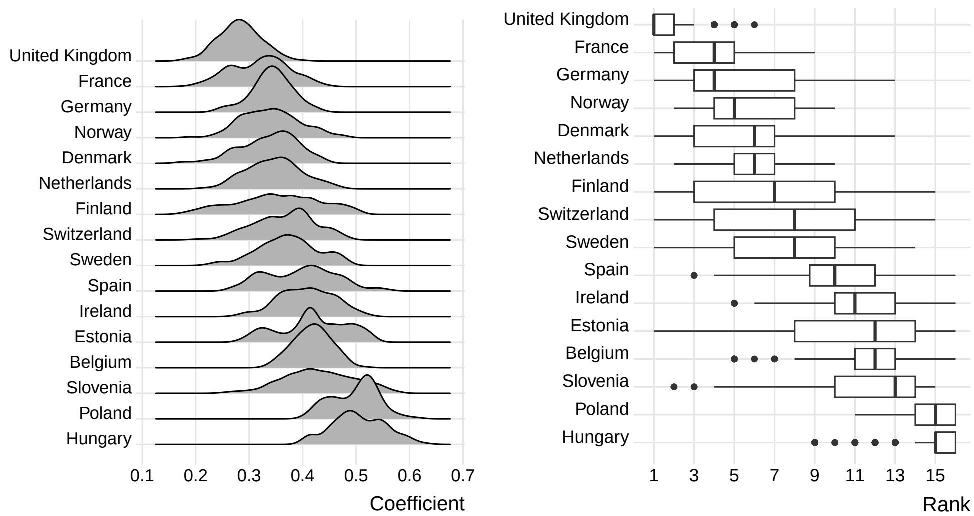

Figure 5: Rank correlations, distribution of coefficients, and country rankings. *Note:* Countries are ordered by the median country rank. 528 models per country.

remain at the bottom. Spain also ranks low (11th), Switzerland sits in the middle (eighth), and three countries span the full range from first to 15th or 16th. These shifts highlight how positional measures can drastically reshape country rankings, supporting Bol's (2015) argument that relative education differs from absolute education and Bukodi and Goldthorpe's (2016) finding that the effects of expansion depend on whether education is treated as positional.

Stability of Rankings

So far, we have treated the five parameters separately under the assumption that they tap different constructs and should not be compared on the same scale. Conflating the different parameters leads to even greater uncertainty in rankings, as shown in Figure A6 in the online supplement. In this section, we formalize this notion and also test how variation in rankings due to parameter choice compares to that stemming from other model components.

To quantify stability, we use the coefficient of determination (R^2) in a least-squares regression with rank $(1, 2, \ldots K)$ as the outcome and K country indicators as regressors. Intuitively, this yields a measure that equals 0 if rank distributions for countries are completely overlapping, and 1 if they are completely separated, that is, if the rank of a country does not vary across specifications. We refer to this measure as *rank stability*. To further understand the proportion of variation in rankings due to different model components, we calculate the same measure using country indicators interacted with a given component, such as the parameter of association or the respondent's gender. Intuitively, this tells us what the aggregate rank stability is within different values of each component, for example, within a given parameter or within the groups of men and women.

We first confirm that the parameter of association is indeed the most important contributor to variation in rank across countries, as shown in Figure 6. Pooling all

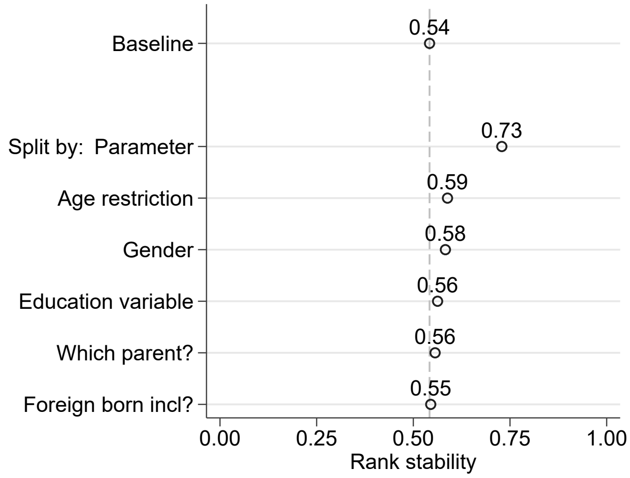

Figure 6: Rank stability. *Note:* The figure shows rank stability at baseline and split by model components.

parameters of association together (regression coefficients, correlations, log odds, unidiff, and rank correlations), the rank stability coefficient amounts to 0.54. In other words, 54 percent of variation in ranks occurs between countries, and the rest across different specifications within countries. Adding the parameter of association to the rank regression, thereby moving this factor from the within to the between variation, pushes the rank stability up to 0.73. In other words, agreement about ranks is markedly higher within individual parameters than it is taken across the multiverse as a whole. No other single model component makes a comparable contribution to rank stability. Accounting for any other component in the model pushes the rank stability coefficient no higher than 0.60 (Fig. 6).

In supplemental materials, we compare rankings using different parameters: how consistent they are with each other, and which parameters yield rankings that deviate more or less from the overall pattern. Table A8 in the online supplement shows the correlation matrix of rankings between parameters and Table A10 in the online supplement shows the factor loadings for different parameters on a shared, underlying dimension. Correlations and rank correlations yield rankings closest to the common underlying factor, with factor loadings of 0.88 and 0.85, respectively. For remaining parameters (regression coefficients, log odds ratios, and unidiff coefficients) the factor loadings are weaker, and all in the range of $0.65-0.73$. These measures produce rankings that are more distinct from the overall pattern across all rankings.

Rank stability varies considerably between parameters, as shown by the values labeled "baseline" in Figure A7 in the online supplement. It is lowest for regression coefficients, correlations, and rank correlations, all at $0.67-0.68$. It is higher for categorical measures such as the log odds ratio (0.80) and unidiff coefficient (0.84). In other words, countries display more similar levels of mobility when using linear measures of association and appear more distinct when using categorical measures.[11] Figure A7 in the online supplement also shows how rank stability improves further when splitting the analysis by any of the other model components. None of the components contribute to rank stability by more than 16 percent, the largest improvement coming from age restriction in Figure A7 in the online supplement, top right panel ($0.78/0.67 = 1.16$). One implication of this is that most model variation depends on higher-order interactions and is therefore hard to predict.

Notably, gender differences are not a major contributor to country variation in mobility rankings. This implies that rankings are similar whether we study men, women, or both. At the face of it, this runs counter to the findings of Engzell and Mood (2023) where gender emerges as the main organizing frame for differences in income mobility trends. They argue that gender dynamics are "a force powerful enough to be the main driver of trends and differences across countries, and cannot be ignored" (p. 619). However, this discrepancy is unsurprising in light of the fact that we study education. As shown in Engzell and Mood (2023) their findings are driven by gender differences in the economic returns to education, rather than differences in the importance of family background for education. Gender may still matter for mobility in other outcomes such as occupation or earnings, or in education for older cohorts where women had yet to reach parity with men.

Incorporating Sampling Error

So far, we have abstracted from the issue of sampling error. Model and sampling variability can be addressed jointly, for example, through bootstrapping or similar approaches that generate empirical distributions incorporating both sources of uncertainty (Ibarra et al. 2025; Simonsohn, Simmons, and Nelson 2020a). We believe there are three issues at stake. First, is model variation large or small compared to the sampling variation covered in most extant research? Second, are our estimates of model variation confounded by sampling variation? This is a subtler point: because the underlying data change from one specification to the next, we would expect some degree of uncertainty simply from this aspect of the analysis. Third, how much do results shift when we account for both sources of variability together? We take each question in turn.

To assess the size of the two sources of variation, we compute two statistics at the parameter−country level: the mean standard error within a country (sampling variation) and the standard deviation of point estimates across specifications (its empirical analog for model variation). This comparison is limited to the four parameter types that produce parametric standard errors by default: regression coefficients, correlations, log odds ratios, and rank correlations. Results are summarized in Figure A8 in the online supplement. Across the board, sampling uncertainty is smaller than model uncertainty, often by a wide margin. For regression coefficients, correlations, and rank correlations, the standard deviation of point estimates is about four times the mean standard error; for log odds ratios, the ratio is roughly two-to-one. These findings reinforce our view that model uncertainty, often overlooked, is a dominant source of variation that researchers ignore at their peril.

To probe whether our results are confounded by sampling error, we illustrate how estimates and ranks vary purely as a function of sampling. We arbitrarily select one specification for each parameter (regression, correlation, odds ratio, unidiff, and rank correlation).[12] For each specification within that parameter, we then draw 250 bootstrap samples per country using ESS weights and report the results in the same format as our main figures. As shown in Figures A9−A13 in the online supplement, parameter estimates and rankings fluctuate much less when variation is induced solely by sampling uncertainty, demonstrating that sampling error alone does not explain model variation.

Finally, we assess how results change when sampling and model variation are considered together. Because this involves a larger number of models, instead of computationally intensive bootstrapping, we use parametric standard errors to simulate plausible values. Specifically, for each specification, we draw 250 values from a normal distribution centered on the point estimate, with a standard deviation equal to its reported standard error. Figures A14−A17 in the online supplement present the results. The distributions resemble those obtained when ignoring sampling variability, because the additional noise from sampling uncertainty largely overlaps with the variation already introduced by model choice. For country rankings, we observe an increase in outliers, but the overall ordering and broad conclusions remain the same. Thus, while sampling error adds further uncertainty, it does not alter our substantive findings.

Discussion: Nordic Exceptionalism in a New Light?

How well do our results align with stylized facts in the field? There are at least two ways to look at this. First, we can ask whether the variation we document is artificially suppressed in published studies. Conventional publications often present a single model specification—chosen for theoretical or pragmatic reasons—as if it were definitive. This practice may obscure the extent of disagreement across plausible alternatives, giving the impression of greater certainty than actually exists. Second, we can ask whether, despite this variation, the broad ordering of countries in our results is nevertheless consistent with what other studies have found. If our results diverge sharply from published findings, this may in turn suggest that what is perceived as common knowledge may partly reflect selective reporting of specification choices.

To first examine the range of variation, we turn to the case of Denmark that we have used to motivate our analysis. Figure 7 shows a density curve of our regression coefficient estimates for Denmark (containing the same information as the ridge plot for Denmark in Figure 1). Overlaid are point estimates from Andrade and Thomsen (2018) who find regression coefficients of 0.37 looking at fathers and 0.42 looking at both parents, Hertz et al. (2008) who puts the regression coefficient at 0.49, and from Liu and Ding (2020) who found estimates of 0.20 for mothers only and 0.23 for fathers only. Our distribution has two clusters around 0.25 and 0.37, which corresponds quite well to paternal estimates by Liu and Ding (2020) and Andrade and Thomsen (2018). The average point estimate in our analyses lies around 0.32, and while that does not correspond to any previous estimate, they all fall within the multiverse distribution, which would point to the analysis capturing the analytical choices of previous estimates quite well. The estimate by Hertz et al. (2008) falls at the right end of the distribution.[13]

Should this variation be seen as large or small? Previously, Karlson (2021, p. 349) has argued that a difference between U.S. and Danish regression coefficients

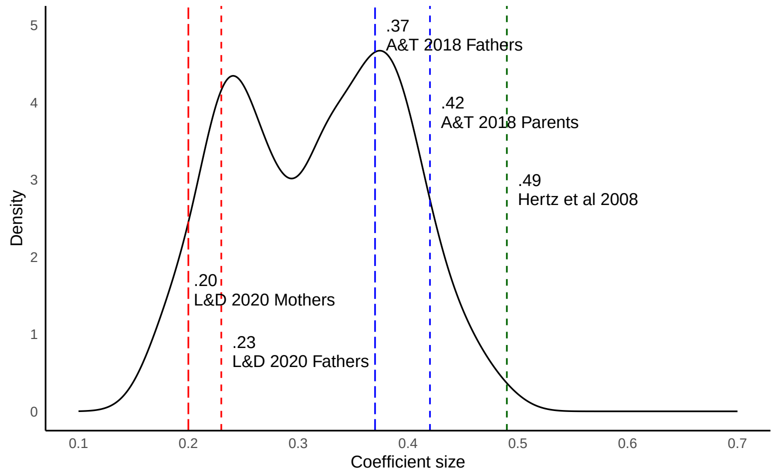

Figure 7: Kernel density plot of the distribution of regression coefficient values for Denmark across all models. Estimates of regression coefficients for Denmark from previous research estimates overlaid as dashed lines.

of 0.04 (as found by Andrade and Thomsen 2018) is not to be considered substantial, comparing it with differences in the magnitude of 0.27 in income elasticity. We can compare these differences with the overall dispersion of estimates for one country in our study. For example, the IQR of regression coefficients for Denmark amounts to 0.13 (the average IQR across all countries is 0.09). This is certainly greater than the difference between Denmark and the United States of 0.04, and points to limited robustness of educational mobility point estimates. One could therefore argue that researchers should not draw too big conclusions from educational mobility comparisons when differences are of this magnitude, not because they are smaller than income elasticity differences, but because the difference could easily stem from model construction.

What does the ranking of countries in our study say about a possible Nordic exceptionalism? If we combine all rankings, as shown in Figure A6 in the online supplement, the Nordic countries are ranked as second to sixth most mobile, with Denmark, Sweden, and Finland having a median ranking of four across all models. On average, then, our results confirm the stylized fact that the Nordics tend to cluster in the upper half of the mobility distribution.[14] Whether occupying the top half of the distribution qualifies as "exceptional" is a separate matter that depends on whether exceptionalism is understood as being consistently at the top, or simply as performing better than most other countries.

At the same time, we present robust results that qualify some common assumptions about how institutions shape intergenerational mobility. The UK and Germany are often classified as low-mobility regimes, due to a selective private sector in the UK and early educational tracking in Germany. However, in our analyses, both countries appear among the more mobile, matching or even surpassing the levels observed in the Nordic countries.

The most striking case is the UK, which across all specifications emerges as the single most mobile country, with a median ranking of three. This is unexpected: no previous ranking has placed the UK at the very top, and it has alternated between the top and bottom halves depending on the study. Although our results place it first overall, the wide range of 13 indicates that very high and very low rankings are both possible depending on analytical choices. Germany also performs better than often assumed, consistently appearing in the top half of the distribution, with a median rank of five across all models (Figure A6 in the online supplement), despite having been placed near the bottom in some earlier work (Chevalier et al. 2009; Pfeffer 2008).

Have the UK and Germany been mischaracterized in earlier literature? A closer reading suggests that some of these patterns have been hiding in plain sight. For example, Gregg et al. (2017) show that differences in income mobility across Sweden, the UK, and the United States cannot be explained by disparities in educational attainment, which are strikingly similar across the three. Likewise, Pensiero and Barone (2024) demonstrate that the intergenerational persistence of education has declined in the UK, placing it on par with the Nordics. In the German case, the apparent contradiction between relatively high measured mobility and the expectation of rigidity under early tracking has long drawn attention, with scholars such as Schneider (2009, 2010) highlighting the mismatch and proposing revisions of the ISCED classification to better capture theoretical intuitions.

These examples suggest that the anomalies are not entirely absent from previous research; rather, they are acknowledged but often fail to shift the prevailing consensus. The issue is not so much selective reporting as the subjective updating of what is taken to be common knowledge in the field. Here lies one of the advantages of a multiverse approach: by presenting the full range of plausible results, it becomes harder to dismiss unexpected findings as statistical outliers. Instead, we are confronted with genuine theoretical puzzles that demand explanation.

Conclusion

This study has shown that estimates of intergenerational educational mobility, and the country rankings derived from them, are highly sensitive to analytical choices. Using a multiverse of 2,880 specifications applied to ESS data from 16 European countries, we find that while some patterns are consistent—Nordic, UK, and German cases often placing in the upper half, and Southern and Central European countries in the lower half—rankings for many countries vary widely, sometimes spanning both top and bottom positions. Most importantly, the choice of association parameter is the single factor with the greatest impact on results, underscoring that different measures capture distinct theoretical constructs rather than interchangeable versions of the same phenomenon.

These findings have two implications. Substantively, they challenge the idea of a stable "Nordic exceptionalism" in educational mobility and point instead to a broader west−east/south divide in Europe, with economic mobility shaped by institutions beyond schooling alone. Methodologically, they caution against presenting single estimates as definitive. Analytical variation is not only noise but also reflects theoretical disagreements about how education operates: as human capital, positional good, or signal. Future research should therefore (1) treat parameter choice as a substantive theoretical decision, (2) consider multiple specifications rather than relying on a single preferred model, and (3) use variation across models as an opportunity to refine theory rather than dismissing it as uncertainty. In this way, multiverse analysis moves us beyond robustness checks toward a more transparent and theoretically grounded understanding of mobility processes.

We have emphasized the need for theory in adjudicating between rankings. Does this imply that if choices of measurement are systematically guided by a single theoretical orientation, results will become more robust? We are skeptical. Theory can help explain why rankings diverge, but it cannot resolve the fact that different frameworks capture distinct mechanisms, all of which are relevant. The obvious answer to whether education functions as human capital, signal, or positional good is: all of the above. Thus, while theory-guided choices may reduce "within-framework" variation, they cannot eliminate disagreement across frameworks. We therefore see the value of multiverse analysis not in collapsing toward a single "true" ranking, but in making explicit how results hinge on theory, clarifying which conclusions are robust across orientations and where genuine theoretical debate remains.

This also highlights the limitations of a common practice in sociological research: selecting one mobility measure, often justified on theoretical grounds, and assessing robustness only by varying sample definitions or control variables. Our results suggest that this strategy risks obscuring how different measures embody distinct assumptions and capture different facets of mobility. Future work should treat the choice of measure not as a purely technical decision but as a substantive one, explicitly tied to the dimension of mobility under study. Multiverse analysis offers a way forward: rather than asking whether results are robust to small perturbations of a preferred model, researchers can examine how findings vary across theoretically informed alternatives. This shift encourages more cautious interpretation of single estimates while enriching theoretical dialog by showing which findings hold consistently and which depend on particular modeling choices.

Some variation can be understood as reflecting real theoretical puzzles, pointing to opportunities for deeper investigation. Nevertheless, considerable variation in rankings is likely to remain unexplained. Should these results then lead us to rethink the literature? On the one hand, the sheer range of variation is hard to dismiss. For countries that place in the middle range, it becomes possible to pick specifications to support almost any conclusion. Even for countries that place closer to one extreme of the distribution, there is room for considerable disagreement. On the other hand, the bigger picture that emerges from our analysis is by and large consistent with existing literature. A key lesson is therefore that this variation is an irreducible feature of intergenerational mobility analyses. Searching for single "best" estimates or unambiguous rankings is likely to be futile. A more fruitful approach is to embrace model uncertainty, focus on broader comparative patterns, and use discrepancies as a basis for theory refinement.

Notes

1 When Karlson (2021) tests different International Standard Classification of Education (ISCED) codings based on mobility tables from Pfeffer (2008), the UK spans positions

from the middle to the third most mobile country. In Hertz et al. (2008), its position ranges from the top 10 percent to the bottom 30 percent, depending on the parameter used.

2 Similar to a correlation coefficient, an odds ratio takes a fixed value (i.e., unity) if parent and child education are independent. Unlike a correlation coefficient, it has no clear upper bound. Odds ratios increase exponentially with small denominators: if some levels of parental education are rare in the population, this will result in small cell counts and extremely high odds ratios, which can make it hard to compare results on the same scale. Therefore, it is common to impose the log transformation, as we do here.

3 Previous research has investigated the relative importance of fathers and mothers, with mixed results: some studies report mother's education being more important for daughters (Heineck and Riphahn 2009; Johnston et al. 2005), father's being more important for sons than daughters (Björklund et al. 2007), and others finding no clear patterns (Ermisch and Francesconi 2001; Jerrim and Micklewright 2011).

4 For Portugal, there are no mappings available that make it possible to recode ISCED-97 and ISCED-11. Austria has no data for the ISCED educational variables from rounds four to six, and in addition there are no mappings available for ISCED-11. The Czech Republic has very skewed distributions toward primary parental education, resulting in small cell counts and extreme values in categorical models.

5 The United Nations Educational, Scientific and Cultural Organization; available at http://uis.unesco.org/en/isced-mappings

6 Tertiary education includes university and also tertiary education below the bachelor level (ISCED 5A) and advanced vocational qualifications (ISCED 5B).

7 Differences between respondents and parents stem from the availability of actual years of education for respondents, and for the recoding to dominant and average values for parents. We do not count standardized or relative education as measures here but rather as recodings.

8 In some educational codings, the sample size (n) drops as low as 259 observations. For regression coefficients, correlations, and rank correlations, we drop 48 out of 576 models with $n < 500$, leaving 528 models for analysis. In binary models, 68 out of 576 are dropped, resulting in 508 models for analysis. In the unidiff analysis, 52 models cannot be estimated due to empty cells, and we exclude another 42 due to low n, resulting in 94 dropped models and 482 remaining for analysis. We have also tried putting the cap at 1,000 observations per model, but the results were substantively similar.

9 When the standard deviation of education is larger in the parent than in the child generation, the regression coefficient will always be lower than the correlation, and vice versa (Black and Devereux 2011, p. 1504).

10 Unidiff models require us to set a country as reference category. We opt for Belgium, which tends to be situated in the middle of the distribution. That we subsequently z-standardize unidiff parameters across each specification explains why Belgium, despite being the reference category, shows a dispersion around zero.

11 The same information is apparent from the narrower bands around the median ranks for categorical measures shown in Figures 1–5. However, these categorical measures do not agree very strongly with each other (Spearman ρ = 0.57, Table A8 in the online supplement) or with the underlying factor (0.73 and 0.65 for log odds ratios and unidiff coefficients, respectively, Table A10 in the online supplement). In other words, they yield a ranking that is consistent but largely idiosyncratic. Rankings that accord closer with the common pattern come with more uncertainty.

12 For continuous measures, we use theoretically coded years of education with average parental attainment, and for categorical measures, ES-ISCED with dominant parental attainment. We include respondents between 25 and 74 and do not exclude foreign born respondents. Both genders are run separately.

13 A potentially important difference is that our density curve represents estimates of the average association over an 20-year period, whereas previous research has looked at specific time points, with Hertz et al. (2008) using Danish data from 1998. In addition, type of data will surely play a role in these differences. Andrade and Thomsen (2018) used register data, which is of higher quality than for a survey such as the ESS, whereas Hertz et al. (2008) and Liu and Ding (2020) also use survey data.

14 Previous articles, with few exceptions, have placed the Nordic countries in the top half (Chevalier et al. 2009; Hertz et al. 2008), but examples to the contrary are Norway in Pfeffer (2008), and Sweden in one out of four rankings in Liu and Ding (2020). When comparing our rankings with those of previous studies, one also has to keep in mind that the countries of comparison differ. Chevalier et al. (2009) also included Czech Republic, Italy, Northern Ireland, and the United States. Pfeffer (2008) compared with Canada, Chile, Italy, New Zealand, Northern Ireland, and the United States. Neither included France, Estonia, and Spain as done in our study.

References

Andrade, Stefan and Jens-Peter Thomsen. 2018. "Intergenerational Educational Mobility in Denmark and the United States." *Sociological Science* 5:93–113. https://doi.org/10.15195/v5.a5

Andrade, Stefan and Jens-Peter Thomsen. 2021. "Yes, Denmark Is a More Educationally Mobile Society than the United States: Rejoinder to Kristian Karlson." *Sociological Science* 8:359–70. https://doi.org/10.15195/v8.a18

Arrow, Kenneth J. 1973. "Higher Education as a Filter." *Journal of Public Economics* 2:193–216. https://doi.org/10.1016/0047-2727(73)90013-3

Becker, Gary S. and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy* 87:1153–89. https://doi.org/10.1086/260831

Beller, Emily. 2009. "Bringing Intergenerational Social Mobility Research into the Twenty-First Century: Why Mothers Matter." *American Sociological Review* 74:507–28. https://doi.org/10.1177/000312240907400401

Bernardi, Fabrizio and Gabriele Ballarino. 2016. *Education, Occupation and Social Origin: A Comparative Analysis of the Transmission of Socio-Economic Inequalities*. Edward Elgar Publishing. https://doi.org/10.4337/9781785360459

Björklund, Anders, Markus Jäntti, and Gary Solon. 2007. "Nature and Nurture in the Intergenerational Transmission of Socioeconomic Status: Evidence from Swedish Children and Their Biological and Rearing Parents." *The B.E. Journal of Economic Analysis & Policy* 7:4. https://doi.org/10.2202/1935-1682.1753

Black, Sandra and Paul Devereux. 2011. "Recent Developments in Intergenerational Mobility." Pp. 1487–541 in *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card, vol. 4. North Holland Press, Elsevier. https://doi.org/10.1016/S0169-7218(11)02414-2

Blanden, Jo. 2013. "Cross-Country Rankings in Intergenerational Mobility: A Comparison of Approaches from Economics and Sociology." *Journal of Economic Surveys* 27:38–73. https://doi.org/10.1111/j.1467-6419.2011.00690.x

Bol, Thijs. 2015. "Has Education Become More Positional? Educational Expansion and Labour Market Outcomes, 1985–2007." *Acta Sociologica* 58:105–20. https://doi.org/10.1177/0001699315570918

Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel D. Schnitzlein, and Kjell Vaage. 2017. "A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US." *Scandinavian Journal of Economics* 119:72–101. https://doi.org/10.1111/sjoe.12197

Braun, Michael and Walter Müller. 1997. "Measurement of Education in Comparative Research." *Comparative Social Research* 16:163–201.

Breen, Richard and Jan O. Jonsson. 2000. "Analyzing Educational Careers: A Multinomial Transition Model." *American Sociological Review* 65:754–772. https://doi.org/10.1177/000312240006500507

Breen, Richard and Jan O. Jonsson. 2005. "Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility." *Annual Review of Sociology* 31:223–43. https://doi.org/10.1146/annurev.soc.31.041304.122232

Breen, Richard, Carina Mood, and Jan O. Jonsson. 2016. "How Much Scope for a Mobility Paradox? The Relationship between Social and Income Mobility in Sweden." *Sociological Science* 3:39–60. https://doi.org/10.15195/v3.a3

Bukodi, Erzsébet and John H. Goldthorpe. 2016. "Educational Attainment - Relative or Absolute - As a Mediator of Intergenerational Class Mobility in Britain." *Research in Social Stratification and Mobility* 43:5–15. https://doi.org/10.1016/j.rssm.2015.01.003

Burgess, Simon, Claire Crawford, and Lindsey Macmillan. 2018. "Access to Grammar Schools by Socio-Economic Status." *Environment and Planning A: Economy and Space* 50:1381–85. https://doi.org/10.1177/0308518X18787820

Chevalier, Arnaud, Kevin Denny, and Dorren McMahon. 2009. "A Multi-Country Study of Intergenerational Educational Mobility." Pp. 260-81 in *Education and Inequality Across Europe*, edited by P. Dolton and R. Asplund, and E. Barth. Edward Elgar Publishing. https://doi.org/10.4337/9781035305582.00020

Engzell, Per and Jan O. Jonsson. 2015. "Estimating Social and Ethnic Inequality in School Surveys: Biases from Child Misreporting and Parent Nonresponse." *European Sociological Review* 31:312–25. https://doi.org/10.1093/esr/jcv005

Engzell, Per and Carina Mood. 2023. "Understanding Patterns and Trends in Income Mobility through Multiverse Analysis." *American Sociological Review*, 88:600–26. https://doi.org/10.1177/00031224231180607

Erikson, Robert. 1984. "Social Class of Men, Women and Families." *Sociology* 18:500–14. https://doi.org/10.1177/0038038584018004003

Erikson, Robert and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford University Press.

Erikson, Robert and Jan O. Jonsson (eds.). 1996. *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Boulder, CO: Westview Press.

Ermisch, John and Marco Francesconi. 2001. "Family Matters: Impacts of Family Background on Educational Attainments." *Economica* 68:137–56. https://doi.org/10.1111/1468-0335.00239

Esping-Andersen, Gosta. 1990. *The Three Worlds of Welfare Capitalism*. Princeton University Press. https://doi.org/10.1177/095892879100100108

Gregg, Paul, Jan O. Jonsson, Lindsey Macmillan, and Carina Mood. 2017. "The Role of Education for Intergenerational Income Mobility: A Comparison of the United States, Great Britain, and Sweden." *Social Forces* 96:121–52. https://doi.org/10.1093/sf/sox051

Haveman, Robert and Barbara Wolfe. 1995. "The Determinants of Children's Attainments: A Review of Methods and Findings." *Journal of Economic Literature* 33:1829–78.

Heineck, Guido and Regina T. Riphahn. 2009. "Intergenerational Transmission of Educational Attainment in Germany: The Last Five Decades." *Jahrbücher für Nationalökonomie und Statistik* 229:36–60. https://doi.org/10.1515/jbnst-2009-0104

Henderson, Morag, Jake Anders, Francis Green, and Golo Henseke. 2020. "Private Schooling, Subject Choice, Upper Secondary Attainment and Progression to University." *Oxford Review of Education* 46:295–312. https://doi.org/10.1080/03054985.2019.1669551

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina. 2008. "The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends." *The B.E. Journal of Economic Analysis & Policy* 7:10. https://doi.org/10.2202/1935-1682.1775

Hirsch, Fred. 1976. *Social Limits to Growth*. Harvard University Press. https://doi.org/10.4159/harvard.9780674497900

Ibarra, Daniel Valdenegro, Jiani Yan, Duiyi Dai, and Charles Rahal. 2025. "Introducing RobustiPy: An Efficient Next Generation Multiversal Library with Model Selection, Averaging, Resampling, and Explainable Artificial Intelligence." *arXiv preprint arXiv:2506.19958*.

Jerrim, John and John Micklewright. 2011. "Children's Cognitive Ability and Parents' Education: Distinguishing the Impact of Mothers and Fathers." Pp. 261–86 in *Persistence, Privilege, and Parenting: The Comparative Study of Intergenerational Mobility*, edited by T. M. Smeeding, R. Erikson, and M. Jäntti. New York: Russell Sage Foundation.

Johnston, Aaron Douglas, Harry B. G. Ganzeboom, and Donald J. Treiman. 2005. "Mothers' and Fathers' Influences on Educational Attainment." Paper presented at the ISA RC28 Spring Meeting, Oslo, Norway.

Karlson, Kristian Bernt. 2021. "Is Denmark a Much More Educationally Mobile Society than the United States? Comment on Andrade and Thomsen, 'Intergenerational Educational Mobility in Denmark and the United States'." *Sociological Science* 8:346–58. https://doi.org/10.15195/v8.a17

Karlson, Kristian Bernt and Jesper Fels Birkelund. 2024. "Origins of Attainment: Do Brother Correlations in Occupational Status and Income Overlap?" *European Sociological Review* 40:379–89. https://doi.org/10.1093/esr/jcad030

Landersø, Rasmus and James J. Heckman. 2017. "The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US." *The Scandinavian Journal of Economics* 119:178–230. https://doi.org/10.1111/sjoe.12219

Liu, Qiang and Ruichang Ding. 2020. "Does Poverty Matter or Inequality? An International Comparative Analysis on the Intergenerational Education Persistence." Pp. 187–205 in *Education and Mobilities, Perspectives on Rethinking and Reforming Education*, edited by X. Zhu and J. Li, M. Li, Q. Liu, and H. Starkey. Singapore: Springer. https://doi.org/10.1007/978-981-13-9031-9_10

Mare, Robert D. 1980. "Social Background and School Continuation Decisions." *Journal of the American Statistical Association* 75:295–305. https://doi.org/10.1080/01621459.1980.10477466

Mogstad, Magne, Joseph P. Romano, Azeem M. Shaikh, and Daniel Wilhelm. 2024. "Inference for Ranks with Applications to Mobility across Neighbourhoods and Academic Achievement across Countries." *Review of Economic Studies* 91:476–518. https://doi.org/10.1093/restud/rdad006

Muñoz, John and Cristobal Young. 2018. "We Ran 9 Billion Regressions: Eliminating False Positives through Computational Model Robustness." *Sociological Methodology* 48:1–33. https://doi.org/10.1177/0081175018777988

Narayan, Ambar, Roy Van der Weide, Alexandru Cojocaru, Christoph Lakner, Silvia Redaelli, Daniel Gerszon Mahler, Rakesh Gupta N. Ramasubbaiah, and Stefan Thewissen. 2018. *Fair Progress?: Economic Mobility Across Generations Around the World*. World Bank Publications. https://doi.org/10.1596/978-1-4648-1210-1

Pensiero, Nicola and Carlo Barone. 2024. "Parental Schooling, Educational Attainment, Skills, and Earnings: A Trend Analysis across Fifteen Countries." *Social Forces* 102:1288–309. https://doi.org/10.1093/sf/soad144

Pfeffer, Fabian T. 2008. "Persistent Inequality in Educational Attainment and Its Institutional Context." *European Sociological Review* 24:543–65. https://doi.org/10.1093/esr/jcn026

Reeves, Aaron, Sam Friedman, Charles Rahal, and Magne Flemmen. 2017. "The Decline and Persistence of the Old Boy: Private Schools and Elite Recruitment 1897 to 2016." *American Sociological Review* 82:1139–66. https://doi.org/10.1177/0003122417735742

Schneider, Silke. 2008a. "The Application of the ISCED-97 to the UK's Educational Qualifications." Pp. 281–300 in *The International Standard Classification of Education (ISCED-97) An Evaluation of Content and Criterion Validity for 15 European Countries*, edited by Ruud Luijkx, Manon d. Heus, and Silke Schneider. Mannheimer Zentrum für Europäische Sozialforschung (MZES).

Schneider, Silke. 2008b. "Applying the ISCED-97 to the German educational qualifications." Pp. 76–102 in *The International Standard Classification of Education (ISCED-97) An Evaluation of Content and Criterion Validity for 15 European Countries*, edited by Ruud Luijkx, Manon d. Heus, and Silke Schneider. Mannheimer Zentrum für Europäische Sozialforschung (MZES).

Schneider, Silke. 2009. "Confusing Credentials: The Cross-Nationally Comparable Measurement of Educational Attainment." DPhil thesis, University of Oxford, Nuffield College.

Schneider, Silke. 2010. "Nominal Comparability Is Not Enough: (In-)Equivalence of Construct Validity of Cross-National Measures of Educational Attainment in the European Social Survey." *Research in Social Stratification and Mobility* 28:343–57. https://doi.org/10.1016/j.rssm.2010.03.001

Schneider, Silke. 2013. "International Standard Classification of Education 2011 and Its Application in the European Social Survey." In *CSDI Workshop*. March 22, Stockholm. https://doi.org/10.1108/S0195-6310(2013)0000030017

Schneider, Silke. 2020. *Guidelines for the Measurement of Educational Attainment in the European Social Survey*. Mannheim, Germany: GESIS Leibniz Institute for the Social Sciences.

Schweinsberg, Martin, Michael Feldman, Nicola Staub, Olmo R. van den Akker, Robbie C. M. van Aert, Marcel A. L. M. Van Assen, Yang Liu, Tim Althoff, Jeffrey Heer, Alex Kale, et al. 2021. "Same Data, Different Conclusions: Radical Dispersion in Empirical Results When Independent Analysts Operationalize and Test the Same Hypothesis." *Organizational Behavior and Human Decision Processes* 165:228–249.

Shavit, Yossi and Hans-Peter Blossfeld (eds.). 1993. *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Boulder, CO: Westview Press.

Shavit, Yossi and Hyunjoon Park. 2016. "Introduction to the Special Issue: Education as a Positional Good." *Research in Social Stratification and Mobility* 43:1–3. https://doi.org/10.1016/j.rssm.2016.03.003

Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson. 2020a. "Specification Curve Analysis." *Nature Human Behaviour* 4:1208–14. https://doi.org/10.1038/s41562-020-0912-z

Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson. 2020b. "Supplement: Specification Curve Analysis." *Nature Human Behaviour* 4:1208–14. https://doi.org/10.1038/s41562-020-0912-z

Spence, Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87:355–374. https://doi.org/10.2307/1882010

Steegen, Sara, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. 2016. "Increasing Transparency Through a Multiverse Analysis." *Perspectives on Psychological Science* 11:702–12. https://doi.org/10.1177/1745691616658637

Thaning, Max and Martin Hällsten. 2020. "The End of Dominance? Evaluating Measures of Socio-Economic Background in Stratification Research." *European Sociological Review* 36:533–47. https://doi.org/10.1093/esr/jcaa009

Thomsen, Jens-Peter, Stefan Andrade, Florian Hertel, Øyvind Wiborg, and Max Thaning. 2025. "Intergenerational Educational Mobility in Scandinavia and the U.S." Mimeo.

Van de Werfhorst, Herman G. and Jonathan J. B. Mijs. 2010. "Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective." *Annual Review of Sociology* 36:407–28. https://doi.org/10.1146/annurev.soc.012809.102538

Xie, Yu. 1992. "The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables." *American Sociological Review* 57:380–95. https://doi.org/10.2307/2096242

Young, Cristobal and Erin Cumberworth. 2025. *Multiverse Analysis: Computational Methods for Robust Results*. Cambridge University Press. https://doi.org/10.1017/9781009003391

Young, Cristobal and Katherine Holsteen. 2017. "Model Uncertainty and Robustness: A Computational Framework for Multimodel Analysis." *Sociological Methods & Research* 46:3–40. https://doi.org/10.1177/0049124115610347

Acknowledgments: Per Engzell acknowledges funding from the European Research Council, grant no. 101165962 (MaMo). Earlier versions of this work were presented at the 2023 Spring Meeting of the ISA Research Committee 28 on Social Stratification and Mobility (RC28) in Paris, the 2024 Conference of the European Consortium for Sociological Research (ECSR) in Barcelona, and in seminars at the Swedish Institute for Social Research (SOFI) and the Amsterdam Institute for Social Science Research (AISSR). For comments that improved the manuscript, we thank Editor-in-Chief Arnout van de Rijt, Deputy Editor Kristian Karlson, two external reviewers, as well as Adam Altmejd, Krzysztof Czarnecki, Harry Ganzeboom, Jan Helmdag, Mike Hout, Linda Kridahl, Liliya Leopold, Silke Schneider, Edvin Syk, Max Thaning, Jens-Peter Thomsen, Andreas Videbæk Jensen, Kim Weeden, Herman van de Werfhorst, and Daniel Wilhelm. Any errors remain our own.

Ely Strömberg: Department of Sociology, University of Amsterdam.
E-mail: e.o.stromberg@uva.nl.
Per Engzell: UCL Social Research Institute, University College London; Swedish Institute for Social Research, Stockholm University. E-mail: p.engzell@ucl.ac.uk.