sociological science

Supplement to:

Karlson, Kristian Bernt. 2021. “Is Denmark a Much More Educationally Mobile Society than the United States? Comment on Andrade and Thomsen, ‘Inter-generational Educational Mobility in Denmark and the United States’ (2018).” Sociological Science 8: 346-358.

A. Consequences of Using a Crude 3-level Educational Classification

I provide a detailed analysis of the consequences of collapsing educational degrees in different ways for cross-country comparisons of overall educational degree mobility. I draw on the education mobility tables published in Pfeffer (2007) and evaluate the consequences for the unidiff model phi-parameters of collapsing Pfeffer's 5-level ISCED schema in different ways. Table A1 shows the results with each column referring to a different coding (the table note provides detailed information on the coding). To enable straightforward comparison between the U.S. and Denmark, I use Denmark as the reference category in the unidiff models, and place the two countries in the top rows of the table. I estimate the unidiff model using Jann and Seiler's (2019) udiff Stata ado, which yields standard errors and thus enables formal comparison.[1]

In Column (1) in Table A1, I report the phi-parameters based on Pfeffer's 5-level ISCED classification. They reproduce Pfeffer's (2008) result that Denmark and the U.S. have the same level of educational mobility (and the difference is not statistically significant at a 5-percent significance level). In Column (2), I report the phi-parameters using A&T's three-level classification on Pfeffer's (2007) data:

ISCED 0 + 1 + 2: Lower secondary schooling or less / less than high school

ISCED 3: Upper secondary education / high school

ISCED 5 + 6 + 7: At least some higher education / at least associate college degree

Using this crude three-level classification yields a very different result: The U.S. is now less educationally mobile than Denmark. According to the phi-parameters, the U.S. is about 50

[1] Jann, Ben, and Simon Seiler. 2019. "udiff: Stata Module to Estimate the Generalized unidiff Model for Individual-Level Data." http://github.com/benjann/udiff.

percent more immobile than Denmark (i.e., on average, odds ratios are about 50 percent larger in the U.S. than in Denmark), and the U.S. drops from a fifth to a 12th place in country rankings.

What explains this surprising discrepancy in results depending on whether one uses a 3-level or 5-level educational classification? To find out, in the remaining four columns (3 through 6) in Table A1, I report phi-parameters using other collapses of the 5-level classification. These additional analyses clearly show that what drives the discrepancy is the collapsing of ISCED 0/1 (primary education) and 2 (lower secondary education), in particular for parents. To see how I arrive at this conclusion, I go through each of the four columns in Table A1. In Column (3), I collapse the education variables into four categories differentiating ISCED 0/1 from ISCED 2 (i.e., the bottom of the schooling distribution) for both parents and children:

ISCED 0 + 1: Primary schooling or less

ISCED 2: Lower secondary schooling

ISCED 3: Upper secondary education

ISCED 5 + 6 + 7: At least some higher education / at least associate college degree

In Column (4), I instead split short-term and long-term higher education for both parents and children:

ISCED 0 + 1 + 2: Lower secondary schooling or less

ISCED 3: Upper secondary education

ISCED 5: Short-cycle higher education

ISCED 6 + 7: Bachelor's or master's degree (or higher)

For Column (3), I find the exact same result as Pfeffer (2008): Denmark and the U.S. have the same level of educational mobility. However, for Column (4), the U.S. is now less mobile than

Denmark (about 40 percent more immobile on average). To me, these results suggests that not differentiating at the bottom of the educational distribution has a large impact on whether we find the U.S. to be less educationally mobile than Denmark. Given that A&T collapse primary and lower secondary schooling (ISCED 0/1 and 2), this coding choice could explain why they find the U.S. to be less educationally fluid than Denmark.

However, because A&T analyze cohorts born in 1980–1984, all offspring in Denmark obtain at least lower secondary schooling (because of a reform in 1972 that increased the years of compulsory schooling from seven to nine). Thus, although it is not possible to differentiate ISCED 0/1 from 2 for the offspring born in these cohort, it is possible for parents. Therefore, in Columns (5) and (6) in Table A1, I report phi-paramaters from unidiff models in which ISCED 0/1 and 2 are separated for parents but not the offspring. In Column (5), I use A&T's three-level classification for children and the four-level classification that distinguishes ISCED 0/1 and 2 for parents. In Column (6), I use the four-level classification that distinguishes ISCED 5 and 6/7 for children and Pfeffer's full five-level classification for parents. In both columns, the U.S. is slightly less mobile than Denmark (i.e., about 10 percent more immobile on average), but this difference is not statistically significant at a five percent level.

In sum, my analysis suggests that using crude education categories for summarizing and comparing educational fluidity across countries may affect country rankings. For the U.S.-Denmark comparison in particular, collapsing the bottom of the education distribution, especially for parents, has a large impact on how similar Denmark and the U.S. appear to be. For very crude education measures, Denmark and the U.S. are somewhat dissimilar; for less crude education measures—which is normally used in the mobility literature (e.g., Pfeffer 2008, Breen et al. 2009)—Denmark and the U.S. are very similar. Thus, while this analysis is

only indicative of why A&T obtain results that differ from those in Pfeffer (2008), measurement granularity could potentially explain this discrepancy in results.

Table A1. Phi-parameters from Unidiff Model Applied to Different Collapses of Education Mobility Tables Published in Pfeffer (2007)

	(1) Pfeffer (2008) 5-level	(2) A&T 3-level	(3) 4-level, bottom	(4) 4-level, top	(5) Parents 4-level, Offspring 3-level	(6) Parents 5-level Offspring 4-level
DNK	1.00	1.00	1.00	1.00	1.00	1.00
USA	1.03	1.47	0.98	1.39	1.10	1.10
p-value	*0.772*	*< 0.001*	*0.852*	*< 0.001*	*0.286*	*0.269*
The other 17 countries						
BEL	1.54	1.54	1.52	1.70	1.42	1.50
CAN	1.04	1.03	1.09	1.26	1.00	1.06
CHL	1.28	1.66	1.26	1.64	1.32	1.32
CZE	1.08	1.31	1.05	1.15	1.08	1.00
FIN	0.87	0.92	0.86	1.02	0.75	0.78
GB	1.04	0.90	1.09	1.03	0.99	1.03
GER	1.71	1.58	1.73	1.47	1.64	1.52
HUN	1.48	1.55	1.47	1.47	1.34	1.31
IRL	1.30	1.49	1.30	1.52	1.29	1.30
ITA	1.46	1.88	1.47	1.73	1.50	1.41
NIR	0.89	0.83	0.96	1.01	0.81	0.83
NL	1.05	1.22	1.04	1.04	1.07	0.97
NOR	1.50	1.38	1.32	1.19	1.48	1.37
NZL	0.95	0.73	1.05	0.96	0.81	0.87
POL	1.11	1.44	1.10	1.43	1.11	1.10
SVN	1.77	1.87	1.68	1.83	1.75	1.78
SWE	1.08	1.30	1.08	1.44	1.06	1.10

Note: Phi-parameters from a unidiff model with Denmark as reference category. p-values based on coefficients and standard errors comparing USA to DNK. Column 1 is based on Pfeffer's (2008) 5-level ISCED classification; Column 2 on A&T's three 3-level classification (ISCED 0/1/2, ISCED 3, ISCED 5/6/7); Column 3 on a 4-level classification differentiating the bottom (ISCED 0/1, ISCED 2, ISCED 3, ISCED 5/6/7); Column 4 on another 4-level classification differentiating the top (ISCED 0/1/2, ISCED 3, ISCED 5, ISCED 6/7); for Column 5, parents' education is based on the 4-level classification in column 3 (ISCED 0/1, ISCED 2, ISCED 3, ISCED 5/6/7), whereas children's education is based on the 3-level classification in column 2 (ISCED 0/1/2, ISCED 3, ISCED 5/6/7); for Column 5, parent's education is based on Pfeffer's (2008) 5-level classification, whereas children's education is based on the 4-level classification in column 4 (ISCED 0/1/2, ISCED 3, ISCED 5, ISCED 6/7).

B. Minor Statistical and Data-Related Flaws

A&T significantly overstate robustness of findings

A&T report results form a range of supplementary analyses in the online supplement, one of them being "marginal effects" derived from the multinomial logit model. These analyses act as robustness checks of their findings. However, from their reported estimates (and code) for Denmark and the U.S., I can see that these are not marginal effects, but marginal predictions. Thus, the reported "marginal effect" estimates are the exact same as the conditional probabilities they report in Table 2, and consequently, A&T's robustness test is simply a perfect reproduction of a table shown earlier in the paper. The same goes for the linear probability model estimates A&T report in the online supplement. These estimates are a just a different, yet equivalent, way of showing those conditional probabilities. Indeed, had A&T computed marginal effects (not marginal predictions) from the multinomial logit model, these would be identical (yet more efficient) than the one based on the linear probability model. In other words, these additional analyses do not add to the robustness of their findings.

No correction for multiple hypothesis testing

A&T provide several tests of estimates for Denmark against the U.S. If we include their estimates in the online supplement, A&T make around 100 comparisons. Even if the test statistics from these analyses are correlated (as we would assume given the nature of the data), not correcting for multiple hypothesis testing will inevitably result in overstating significance and finding country differences even when there are none. Combined with them reporting artificially small standard errors (see the main text), not correcting for multiple hypothesis poses yet another threat to the conclusions that A&T draw.

Confidence intervals of odds ratios are symmetric

For all odds ratios reported in the paper and online supplement, confidence intervals are symmetric, something that cannot be true given that odds ratios operate on a multiplicative scale. It appears that A&T have calculated these confidence intervals themselves using the standard formula (±1.96×SE) instead of using the standard error reported by the statistical program.