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” (2018)

Kristian Bernt Karlson
University of Copenhagen

Abstract: I evaluate Andrade and Thomsen (A&T)’s (2018) study, which concludes that Denmark is significantly more educationally mobile than the United States. I make three observations. First, A&T overstate the difference in educational mobility between Denmark and the United States. Both in international comparison and compared with differences in intergenerational income mobility, A&T’s reported country differences in educational mobility are negligible. For example, whereas income mobility estimates reported in the literature differ by 300 to 600 percent between the two countries, the corresponding educational mobility estimates that A&T report differ by 10 to 20 percent. Second, I provide evidence suggesting that A&T’s use of crude categorical education measures leads them to overstate these negligible differences. Third, A&T’s empirical analyses of the U.S. data contain several statistical and data-related flaws, some so severe that they potentially undermine the credibility of their analyses. In sum, A&T’s results are perfectly consistent with the existence of a mobility paradox very similar to what Sweden–United States comparisons show: although Denmark and the United States are dissimilar with respect to income mobility, they are similar with respect to educational mobility. Understanding the nature of this paradox should be a key concern for future mobility research.

Keywords: educational mobility; social mobility; stratification; welfare states; educational inequality; education

Comparing levels and patterns of intergenerational mobility among countries is a hallmark of sociological research. Such comparisons speak to important questions of who gets ahead and why, and they have consequences for the way in which we conceptualize how welfare state institutions shape opportunities for children. In 2018, Andrade and Thomsen (hereinafter A&T) published an article in Sociological Science comparing levels of intergenerational educational mobility in Denmark and the United States for cohorts born from 1980 to 1984. A&T frame their study in terms of a comparison between representatives of social-democratic welfare states (Denmark) and liberal welfare states (the United States), and they offer an empirical analysis based on both years of schooling and educational degrees. They conclude, “Educational mobility in Denmark and the United States is not similar; mobility in Denmark is significantly higher than in the United States” (P. 106). This conclusion runs counter to well-known findings in the comparative mobility literature: it contradicts Pfeffer’s (2008) 19-country comparative study, in

Received: June 11, 2021
Accepted: July 11, 2021
Published: November 17, 2021
Editor(s): Jesper Sørensen, Filiz Garip
DOI: 10.15195/v8.a17
Copyright: © 2021 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited.
which Denmark and the United States have the exact same level of educational mobility, and the finding that occupational and educational mobility is very similar in the United States and Sweden, the latter being another representative of social-democratic welfare states (Björklund and Jännti 2000; Breen and Jonsson 2005; Beller and Hout 2006).

In this comment, I revisit A&T’s conclusion and examine three aspects of their comparative study. First, I offer a reinterpretation of A&T’s results by comparing their reported mobility estimates with what we know from other comparative studies, including the intergenerational income mobility literature. In this perspective, A&T’s estimates point to widespread similarity between Denmark and the United States and support the well-known finding of a “mobility paradox” in which educational and occupational mobility in the United States and the Scandinavian countries are very similar, whereas income mobility is dramatically different (Breen, Mood, and Jonsson 2016).

Second, drawing on the education mobility tables on which Pfeffer (2008) bases his analysis, I provide indirect evidence that A&T’s findings for educational mobility measured in terms of odds ratios likely stem from A&T using a very crude education measure. Because A&T do not differentiate between primary and lower secondary schooling among parents, they likely understate the degree of educational mobility in the United States. Third, in A&T’s analyses of the U.S. data, I identify several statistical and data-related flaws, three of which I would classify as so substantial that they potentially undermine the conclusions that A&T are able to draw from their comparisons: (1) all standard errors are downwardly biased; (2) A&T report statistically significant differences even when there are none; and (3) A&T exclude from key analyses nonresidential parents, who constitute a nontrivial and likely selected portion of the U.S. data. I conclude my comment with a discussion of some potential, future avenues for comparative mobility research.

Putting A&T’s Empirical Estimates in an International Perspective

A&T frame their study by setting up a contrast between Scandinavian and liberal welfare states. The first sentence in their abstract states, “An overall finding in comparative mobility studies is that intergenerational mobility is greater in Scandinavia than in liberal welfare state countries like the United States and United Kingdom” (P. 93), a statement that they repeat in the Introduction (p. 93) and other places in the article (e.g., pp. 95–96). Although A&T back up this statement by citing, among others, the review by Breen and Jonsson (2005), they do not provide a balanced treatment of this literature. Here is what Breen and Jonsson (2005) conclude with regard to country differences in different types of mobility:

An interesting issue is the ranking of the United States [in terms of social class mobility]. In an attempt to make a comparison with European societies, Erikson & Goldthorpe (1992) concluded that the United States is fairly similar to them; the somewhat higher degree of fluidity they found was attributed to problems of comparability, stemming from
lack of precision in the American occupational codings. In a direct comparison between educational inequality in the United States and Sweden (one of the most equal countries in the existing literature), Hout & Dohan (1996) found the two to be very similar. It is interesting to contrast these results with those found when inequality of opportunity is measured in terms of income. Studies of father-to-son (and sometimes -daughter) income mobility as well as sibling correlations of income show the United States to be noticeably more rigid than the countries with which it has been compared (mostly the Nordic countries). (Pp. 232–33)

Thus the existing literature shows that the relationship between type of welfare state and levels of intergenerational mobility is far from as clear-cut as A&T argue, particularly when it comes to a comparison between Scandinavian countries and the United States (see also Beller and Hout 2006). What holds for income mobility does not hold for educational and occupational mobility. Moreover, although A&T review the comparative educational mobility study by Hertz et al. (2008), they neither review nor cite the study by Pfeffer (2008), which provides a direct comparison of overall educational degree mobility levels in 19 countries, including Denmark and the United States. Pfeffer (2008) finds that Denmark and the United States have the exact same level of relative educational degree mobility (or fluidity) as measured by the unidiff model. Pfeffer’s (2008) study is cited more than 550 times in Google Scholar and is a well-known and innovative study of the persistence of educational inequalities.2

Apart from not reviewing key studies pointing to similarity between Denmark and the United States, A&T also do not compare their estimates with what we know from the existing literature. Indeed, A&T base their conclusion on the estimates for the United States being statistically significantly higher than those for Denmark, not on the social or substantive significance of these differences. Thus, in what follows I put A&T’s estimates in an international perspective by making two comparisons. First, I compare their education mobility estimates with what we know from the earnings mobility literature. Second, using Duncan’s dissimilarity index, I compare how much the conditional distribution of offspring’s education given parents’ education differs between Denmark and the United States when we compare this difference with variation across the 19 countries on which Pfeffer (2008) bases his analysis. Both comparisons show that the differences in educational mobility between Denmark and the United States are negligible.

Before I make these two comparisons, I briefly outline the setup of A&T’s study. Their main focus is on comparing educational mobility, measured in different ways, between Denmark and the United States. They measure educational mobility in terms of the association between parents’ and children’s years of schooling (regression coefficients, Pearson correlations) on the one hand and educational degrees (conditional probabilities, odds ratios) on the other, with the strength of the association determining the level of persistence across generations. They compare cohorts born from 1980 to 1984 and use administrative register data for Denmark (roughly 250,000 observations) and the National Longitudinal Survey of Youth 1997 (NLSY97) for the United States (ranging from 2,755 to 4,983 observations).
Comparison with Earnings Mobility Estimates

To compare A&T’s educational mobility estimates with estimates of intergenerational earnings mobility, I draw on the estimates in Black and Devereux (2011), a standard reference in the field with more than 1,200 citations in Google Scholar: the earnings elasticity among men is 0.071 and 0.517 in Denmark and the United States, respectively, suggesting that the earnings elasticity is about seven times larger in the United States. In substantive terms, these coefficients mean that, on average, about seven and 52 percent of earnings inequality in the parent generation are transmitted to (or reappear in) the child generation in Denmark and the United States. I consider this a substantial difference, and it is line with the mobility paradox finding described earlier. Compare these earnings elasticities with the intergenerational “education elasticities” (i.e., regression coefficients) that A&T report in their Table 4: 0.42 and 0.46 for Denmark and the United States, respectively, suggesting that the education elasticity is 1.1 times larger in the United States. By this comparison, Denmark and the United States are extremely similar in terms of educational mobility: earnings mobility differs by about 600 percent, whereas educational mobility differs by about 10 percent.

Black and Devereux (2011) also report estimates of the intergenerational earnings correlation, another widely used measure of intergenerational persistence, which controls for changes in the dispersion of earnings across generations. The intergenerational earnings correlations are 0.09 and 0.36 among men in Denmark and the United States, respectively, amounting to the United States being about four times less mobile by this measure. A&T report in their Table 4 intergenerational education correlations of 0.39 and 0.47, suggesting that intergenerational educational persistence is about 1.2 times higher in the United States than in Denmark. Thus, by this measure, earnings mobility differs by about 300 percent, whereas educational mobility differs by about 20 percent, again pointing to widespread similarity in educational mobility.

Considered together, the mobility estimates that A&T report are very much in line with the mobility paradox that Sweden–United States comparisons show: although Denmark and the United States differ with respects to earnings mobility, they are similar when it comes to intergenerational educational mobility.

Dissimilarity Index–Based Comparison Using Mobility Tables

A&T also report mobility estimates based on degrees attained. One example is the final panel in their Table 2 in which A&T report the conditional distribution of children’s degree given parents’ (highest) degree, where degrees are measured in three overall categories: no high school (including less than a high school degree and GEDs), high school degree, and a college degree (including an associate two-year degree, a four-year or bachelor’s degree, and advanced degrees such as a master’s degree, PhD, etc.). I reproduce in Table 1 the final panel in A&T’s Table 2.

The two countries’ conditional distributions in Table 1 are very similar. The conditional distribution in row (c) is virtually identical in the two countries. For row (a), it appears that the bottom of the schooling distribution is slightly stickier
Table 1: Reproduction of final panel in Table 2 in Andrade and Thomsen (2018)

<table>
<thead>
<tr>
<th>Parents</th>
<th>Offspring</th>
<th>College</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) No high school</td>
<td>32</td>
<td>48</td>
<td>20</td>
</tr>
<tr>
<td>(b) High school</td>
<td>16</td>
<td>48</td>
<td>36</td>
</tr>
<tr>
<td>(c) College</td>
<td>8</td>
<td>28</td>
<td>64</td>
</tr>
</tbody>
</table>

United States

<table>
<thead>
<tr>
<th>Parents</th>
<th>Offspring</th>
<th>College</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) No high school</td>
<td>38</td>
<td>48</td>
<td>14</td>
</tr>
<tr>
<td>(b) High school</td>
<td>18</td>
<td>51</td>
<td>31</td>
</tr>
<tr>
<td>(c) College</td>
<td>6</td>
<td>29</td>
<td>65</td>
</tr>
</tbody>
</table>

Notes: Values are percentages. Parents’ education is the highest of the two parents. For the United States, I report the unweighted conditional probabilities. No high school includes GED certificates for the United States; college includes associate (short-cycle) degrees, regular four-year college degrees, and advanced degrees (master’s, PhD, etc.). In terms of ISCED, the classification corresponds to ISCED 0/1/2, ISCED 3, and ISCED 5/6/7.
Figure 1: Cumulative distribution functions of Duncan’s dissimilarity index in 171 pairwise country comparisons based on data published in Pfeffer (2007). Note: Estimates based on collapsing the mobility tables published in Pfeffer (2007) to the three-level educational degree classification that A&T use (i.e., ISCED 0/1/2, ISCED 3, and ISCED 5/6/7).

considerably and, in some cases, up to about 0.50. Medians are in the range of 0.12 to 0.16. To facilitate comparison with A&T’s study, I add dashed vertical lines that correspond to the dissimilarity indices based on their reported conditional probabilities. These lines show that the Denmark–United States differences that A&T report always are among the 10 percent least dissimilar out of the 171 comparisons. If we consider the average across rows (a), (b), and (c), shown in Figure 1(d), A&T’s estimates are in fact the least dissimilar of any of the 171 pairwise comparisons. Thus, considered in an international perspective, the conditional distributions that A&T report unequivocally point to similarity, not dissimilarity, between Denmark and the United States.
Consequences of Using a Crude Three-Level Educational Classification

A&T base part of their analyses on educational degree variables for which they compare the magnitudes of odds ratios between Denmark and the United States (their Table 5). If we disregard the methods- and data-related issues in this part of their analysis, on which I report in a later section, their estimates point to Denmark being somewhat more mobile than the United States. Considering the log of the odds ratios reported in their Table 5, the United States is between 25 and 51 percent more immobile than Denmark. Taking a simple average across the four log odds ratios, the United States is 34 percent more immobile than Denmark. Although these averages are far from the 300 and 600 percent differences reported for earnings mobility, they point to some difference between the two countries.

However, considering Pfeffer’s (2008) finding of no difference in “average odds ratios” between the two countries using the unidiff model, the question arises why A&T reach a different conclusion than Pfeffer (2008). Although one explanation is that Pfeffer’s (2008) study covers birth cohorts other than those A&T analyze, in this section I focus on a different and perhaps less obvious explanation: the coarseness of the educational degree measure that A&T use. As I described earlier, A&T group parents and offspring into three overall groups (no high school, high school degree, or college degree). Although A&T, like Pfeffer (2008), rightly point out that comparing educational mobility between the United States and Denmark or European countries more generally is challenging given the different educational systems in place, such crude grouping could potentially ignore pertinent differences at the bottom and/or top of the educational distribution.

To study the potential consequences of such crude grouping, I draw on the education mobility tables published in Pfeffer (2007) and examine the consequences for the unidiff model phi-parameters of collapsing Pfeffer’s five-level ISCED schema in different ways. I briefly present the main results and refer interested readers to part A of the online supplement for details. I find that using crude education categories for summarizing and comparing educational degree mobility across countries affects country rankings. For the United States–Denmark comparison in particular, collapsing the bottom of the education distribution, especially for parents, has a large impact on how similar Denmark and the United States appear to be. The more granular coding, the more similar Denmark and the United States are. Thus using crude education conceals heterogeneity that may profoundly impact country differences.

Methods- and Data-Related Flaws

I have identified six methods- and data-related flaws in A&T’s empirical analyses of the U.S. data. Three of these flaws are so severe that they potentially undermine the credibility of the country comparison that they make. The remaining three flaws are minor and are described in part B of the online supplement.
All Standard Errors for the United States Are Downwardly Biased

A&T make two mistakes with respect to getting the standard errors for the United States right: (1) they implement survey weights (on which virtually all estimates in the main text are based) in an incorrect way, and (2) they do not correct the standard errors for the stratified two-stage sampling design of the NLSY97. Because these two errors lead to artificially small standard errors, they will impact the cross-country inferences on which A&T’s conclusions rest.

1. With respect to implementing survey weights, their Table 5—which compares odds ratios between the two countries—contains standard errors for the U.S. estimates calculated on a sample weighted up to the population size, that is, as if the sample were about 14,000,000 observations. This substantial error makes it impossible to test whether the (log) odds ratios that A&T report differ between Denmark and the United States. The error lies in A&T using frequency weights when estimating the odds ratios, bringing the “sample size” from about 4,500 to a total of about 14,000,000 observations. Moreover, for the regression and correlation coefficients that A&T report, a similar point holds, as A&T use sampling weights without specifying the standard errors as robust (robust standard errors give unbiased standard errors under heteroskedasticity). As using sampling weights introduces heteroskedasticity, not correcting for it will produce artificially small standard errors, which means that A&T are more likely to reject their null hypothesis, even if there are no significant differences between estimates for the two countries.

2. With respect to the stratified two-sample design of the NLSY97, A&T do not correct their standard errors for this property of their sample. As the NLSY97 samples siblings in the same household, not correcting for clustering leads to biased standard errors, especially because parental education is the key predictor in all models, meaning that the inference is effectively operating at the household level. How much this error impacts standard errors is difficult to predict upfront, as it depends on the intraclass (or sibling) correlation in schooling and the number of household units (Snijders and Bosker 2012). However, the U.S. Bureau of Labor Statistics provides some idea, as it reports the design effect’s multiplier on standard errors (DEFT) for selected analyses. The Bureau reports DEFTs ranging from 1.30 to 1.62. Thus, standard errors could be up to about 60 percent larger than the ones A&T report, suggesting that the standard errors A&T report are very optimistic.

A&T Report Significant Differences When There Are None

In almost all estimation tables in their article and online supplement, A&T note that “All tests (both with weighted and not weighted U.S. data) show that country differences are significant ($p < 0.000$).” A&T also note that they use one-tailed $t$-tests (Denmark < the United States), although they provide no justification for this unconventional choice. Nonetheless, some of their reported estimates are not statistically significantly different at conventional significance levels.
Table 2: Reproduction of panel 3, Table 4, in Andrade and Thomsen (2018)

<table>
<thead>
<tr>
<th></th>
<th>Denmark</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression coefficient</td>
<td>0.370</td>
<td>0.399</td>
</tr>
<tr>
<td>Standard error</td>
<td>(0.002)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>[0.366; 0.374]</td>
<td>[0.367; 0.430]</td>
</tr>
</tbody>
</table>

Note: I thank A&T for providing the estimates with several decimals.

To give an example, consider panel 3 in Table 4 in A&T’s main text in which they provide regression coefficients from a regression of children’s schooling on father’s schooling. I reproduce this panel in Table 2. The U.S. confidence interval overlaps with the point estimate for Denmark, and a two-sample *t*-test yields a *t* value of 1.799. Although this *t* value is larger than the 1.645 critical value for a one-tailed test (assuming Denmark < the United States), it is only significant at a five percent level, and surely not at a 0.1 percent level. Had a two-tailed test been used (and thus a 1.96 critical value), the difference would not be statistically significant different at a conventional five percent level.

Another example is the marginal effects A&T report in their Table A5 in the online supplement. Virtually none of those estimates are significantly different between the two countries, and many point estimates are exactly the same. A similar pattern also appears for some of the odds ratios based on the unweighted U.S. data in their Table A5. For example, in panel B, the U.S. confidence intervals for the two odds ratios in the high school row include the point estimates for Denmark. Had A&T been using two-tailed tests, this difference would not be statistically significant at a five percent level. Moreover, on a more general note, using one-tailed tests is an unusual choice given that it makes it easier for A&T to corroborate their guiding hypothesis that Denmark is more educationally mobile than the United States.

**A&T Exclude Nonresidential Biological Parents**

As A&T have kindly shared their code, I can see that A&T only use some of the information available for parental education degrees. This degree information comes from household rosters. A&T use the 1998 household roster to identify biological mothers and fathers and then identify each of their reported educational degrees, resulting in a sample with valid information on about 2,750 fathers and about 4,500 parents. However, A&T do not include biological parents not living with their children, who comprise about 1,800 fathers and about 600 mothers (with valid parental degree information). Information on these parents is available in the 1997 “non-household roster.” Although it is difficult to predict the consequences of omitting a large and selected portion of the data, it could have a significant impact on the results A&T report, particularly for the analyses based on fathers only.
Discussion

A&T’s comparative analysis of educational mobility in the United States and Denmark does not support the conclusion that A&T draw, namely, that Denmark is significantly more educationally mobile than the United States. Even if we ignore the statistical and data-related flaws in their empirical analyses, their own estimates show widespread similarity (not dissimilarity) in educational mobility between the two countries. My evaluation of A&T’s study points to at least three areas that future research could engage in. First, A&T’s estimates confirm the existence of a mobility paradox: although Denmark and United States are worlds apart when it comes to intergenerational income mobility, they are very similar when it comes to educational mobility, much like what United States–Sweden comparisons show (e.g., Hout and Dohan 1996; Beller and Hout 2006). This result is open to several interpretations (cf. Breen et al. 2016), one of them being that the transmission of underlying human and cultural capital within families is much more resistant to welfare state and labor market institutions than the transmission of earnings (potential) in which wage-setting institutions likely play an important role because they affect the economic returns to schooling (Corak 2013). Evaluating and testing such a substantive interpretation appears key to future research on social mobility.

Second, the coarseness with which we examine educational attainment in comparative (mobility) research might be more important than we think (Schneider 2021). As my analyses based on Pfeffer’s (2008) data show, using very crude education classifications can alter conclusions regarding country rankings, or at least change the rankings of certain countries (such as the United States). My analyses show that differentiating at the bottom of the schooling distribution is critical, especially among parents, and future research could examine whether this result holds generally.

Third, as Bernardi, Chakhaia, and Leopold (2016) point out, statistical significance is an insufficient criterion for evaluating the substantive meaning of coefficients (cf. Firebaugh 2008). Thus in two-country comparisons like the one A&T provide, similar or related estimates from existing research ought to be factored in as “informed benchmarks” before conclusions based solely on statistical significance are made (see Bernardi et al. 2016). This is particularly the case when we base comparisons on total population data such as the Danish administrative registers for which standard errors will be tiny. Although evaluating magnitudes of coefficients is inherently difficult, the wealth of comparative evidence on intergenerational mobility should constitute the primary benchmark when evaluating whether countries differ in their levels and patterns of mobility.

Notes

1 My evaluation is based purely on the estimates that A&T report and what we know from the existing literature. It does not involve replicating or reproducing their results.

2 A&T’s review is also affected by other omissions. They do not review the study by Chevalier, Denny, and McMahon (2009), which is very similar to study by Hertz et al. (2008). As Blanden (2013:73) shows, Chevalier et al. (2009) find similarity in educational
mobility between Denmark and the United States with Spearman rank-order correlations of 0.24 and 0.28, respectively, whereas Hertz et al. (2008) find dissimilarity with Pearson’s correlations (0.30 and 0.46, respectively) and similarity with regression slopes (0.49 and 0.46, respectively). Moreover, according to the correlations reported in Hertz et al. (2008), Denmark (whose correlation is 0.30) is also more mobile than the other Scandinavian countries, in particular Sweden (whose correlation is 0.40), but very similar to Great Britain (whose correlation is 0.31), another representative of the liberal welfare states. Although these results are far from conclusive, they do not support A&T’s sweeping statement that intergenerational mobility is greater in Scandinavia than in liberal welfare state countries.

3 Black and Devereux (2011) report the estimates in Jännti et al. (2006).

4 In Denmark, the equivalent of an associate degree is a short-cycle program in higher education with most programs having a duration of two years.


6 I thank Richard Breen for suggesting using the dissimilarity index as a way of approaching this question.

7 Moreover, apart from Pfeffer (2008), influential studies on educational opportunity also use more detailed versions (see, e.g., Breen et al. 2009), suggesting that A&T deviate from the existing literature in this regard.

8 The impact on the standard errors is so big that the standard errors for the odds ratios are smaller for the United States than for Denmark, despite the sample size being 55 times greater in Denmark (250,000 against 4,500). A&T’s online supplement directly shows this error. In Tables B2 to B9, sample sizes for the United States range from about 9,000,000 to 16,000,000 observations. According to publicly available vital statistics for the United States, roughly 18,000,000 persons were born from 1980 to 1984.

9 Because A&T have been so kind as to share their Stata code, I can see that A&T consistently implement weights in an erroneous way. They consistently use the weight option [w=SAMPLING_WEIGHT2013], where SAMPLING_WEIGHT2013 is an NLSY97 sampling weight. Only specifying w in Stata means that Stata uses the default standard error. For linear regressions and correlations, Stata assumes analytical weights (aweight); for multinomial logits (by which the odds ratios are calculated), Stata assumes frequency weights (fweight). Correct implementation would be to use probability weights (pweights) throughout.

10 Not all parents live in the same household.

11 I assume that “p < 0.000” should be “p < 0.001.”

12 A similar result holds for the regression coefficients A&T report in panel 4 in their Table 4 (regression coefficients, parents–child). The t value for this comparison is 3.037, implying a p value of 0.0012 for a one-tailed test and a p value of 0.0024 for a two-tailed test, none of which are below the 0.1 percent significance level that A&T are operating with.

13 A&T also ignore the retrospective reports on parents’ educational degrees that are available from round 6 in the NLSY97. Including both the non-household roster information and the retrospective information would lead to a small fraction of records with missing parental degree information.
References


Kristian Bernt Karlson: Department of Sociology, University of Copenhagen. E-mail: kbk@soc.ku.dk.