

This is a pre-copyedited, author-produced version of an article accepted for publication in *European Sociological Review* on 2024-03-21 following peer review. The version of record is available at: <https://doi.org/10.1093/esr/jcae020>.

Steeper at the Top: Cognitive Ability and Earnings in Finland and Norway*

Bernt Bratsberg^a, Ole Rogeberg^a, and Marko Terviö^{b,c}

^aFrisch Center for Economic Research, Oslo

^bAalto University School of Business

^cHelsinki GSE

March 15, 2024

Abstract

We document a convex relationship between earnings rank and cognitive ability for men in Finland and Norway using administrative data on over 350,000 men in each country: the top earnings percentile score on average 1 standard deviation higher than median earners, while median earners score about 0.5 standard deviation higher than the bottom percentile of earners. Top earners also have substantially less variation in cognitive test scores. While some high-scoring men are observed with very low earnings, the lowest cognitive scores are almost absent among the top earners. Overall, the joint distribution of earnings rank and ability is very similar in Finland and Norway. We find that the slope of the ability curve across earnings ranks is steepest in the upper tail, as is the slope of the earnings curve across cognitive ability. The steep slope of the ability curve across the top earnings percentiles differs markedly from the flat or declining slope recently reported for Sweden (Keuschnigg, van de Rijt and Bol, 2023). JEL codes: D31, J24, J31.

*We thank Ramin Izadi, Oskar Nordström Skans, Mika Päällysaho, Roope Uusitalo, and seminar participants at Uppsala University for useful comments and suggestions. Permission to use the Finnish administrative data was granted by Statistics Finland (license TK-53-228-14) and by the Finnish Defense Forces (AJ23378). The Norwegian data was made available by Statistics Norway. Permission to use the cognitive score data was granted by the Norwegian Armed Forces, who are not responsible for any of the findings and conclusions reported in the paper. Bernt Bratsberg and Ole Rogeberg were financed by Norwegian Research Council grant 300917.

1 Introduction

Cognitive ability is a robust predictor for a broad range of important life outcomes, including educational attainment, earnings, fertility, health and mortality (Calvin et al. 2011; Hegelund et al. 2018; Bratsberg, Rogeberg, and Skirbekk 2023), but its role nonetheless remains contested. A claim popularized by Gladwell holds that cognitive ability “doesn’t seem to translate into any measurable real-world advantage” beyond some threshold value around 120 IQ points (Gladwell 2008). Others have argued that a subset of top earners made it to the top without high cognitive ability with the help of wealthy, well-connected families or pure luck, consistent with data from Sweden showing a plateau and decline of cognitive ability across top earning percentiles, i.e., a concave ability-earnings relationship (Keuschnigg, van de Rijt, and Bol 2023).

In the present paper, our main objective is to describe and compare the empirical relationship between cognitive ability and labor earnings in Finland and Norway. We use comparable high-quality data from both countries. Cognitive ability is measured using scores from standardized cognitive ability tests administered to male military conscripts, while earnings are drawn from administrative registers. For both countries, we are able to match individual-level test scores to later-life earnings during the 35-45 age range for the 1962-1975 birth cohorts. This is the maximal sample available, as 1962 is the first birth cohort for which Finnish test data are available while those born after 1975 were not yet 45 in the last available data year (2020). The resulting datasets are large, about 400 thousand men in each country, allowing for detailed non-parametric descriptions.

Our main analysis is partly motivated by the surprising plateau finding of Keuschnigg et al (2023). The Keuschnigg et al study used Swedish administrative data on earnings and military test results that are largely comparable with our data. They found that average cognitive ability is essentially flat across the top 10 percentiles of wage earners, and even declines over the three highest percentiles. In the present study, we calculate the average cognitive ability score within each percentile of earnings, using earnings across the same 35-45 age range as the Swedish study. However, in sharp contrast with the prior study, we find that the resulting ability-earnings curve not only increases but also steepens near the top. The shape of the relationship is qualitatively similar in Finland and Norway.

These differences across the countries are surprising. The Swedish institutional context is largely equivalent with its Finnish and Norwegian counterparts: comprehensive male conscription, standardized cognitive ability tests, linkage to adminis-

trative data on later-life earnings. All three countries are social democracies with broadly similar welfare policies and labor market institutions that exemplify the “Scandinavian model” (Barth, Moene, and Willumsen 2014), they have similar intergenerational mobility patterns (Jantti et al. 2006), and are geographically close with Sweden sandwiched between Finland and Norway.

In addition to examining how *average* ability differs by earnings, we also assess the *variability* of cognitive ability conditional on earnings. We find substantially less variation in ability within the higher earnings percentiles, and a clear asymmetry across the earnings distribution: while many high-ability men appear in the lower earnings brackets, low-ability men are almost completely absent from the top earnings percentiles.

Our main analysis uses data for the full population of native-born males regardless of their labor market history. By contrast, the Swedish study only analyzed males identified as labor market entrants in a specific calendar year, reducing the coverage to about 20% of the males within the relevant birth cohorts. Implementing the sample selection rules of Keuschnigg et al (2023) on the Finnish and Norwegian data gives a similar sample reduction (down to 25% of the full study populations) but does not appreciably alter the shape of the ability-earnings relationship (it does, of course, reduce the precision).

2 Data and transformations

We analyze the earnings and cognitive test score data for men born between 1962 and 1975. We define our study population as all native-born men in these birth cohorts who are observed throughout the age range 35-45. Hence, those who die or emigrate before age 45 are not included. Basic summary statistics can be found in Table 1.

Cognitive ability data

The cognitive ability scores come from military conscription tests unique to each country, initially developed in the 1950s. Both tests include subtests for vocabulary, mathematical skills, and abstract reasoning (items similar to Raven’s matrices). In Finland, each subtest has 40 items, while Norwegian subtests range from 30 to 54

items. Tests in both countries remained unchanged throughout our data period.¹

While both Finland and Norway historically had a system of comprehensive conscription for male citizens, the timing of their cognitive tests differs. As the Norwegian test was assigned before the service, these data include those opting for civilian service as well as a subset of those exempted for medical and other reasons, resulting in valid scores for 92% of the study population. In Finland, conscripts take the test during the second week of military service, so the scores are missing for those who opted for civilian service or who were exempted due to medical reasons or criminal record. The Finnish data includes test scores for 82% of the study population.

Both the Finnish and Norwegian militaries use a nine-valued integer scale to summarize individual test results. In theory, these “stanines” divide observations into nine bins of predetermined width where the 5th stanine is centered around the sample mean and all stanine borders are 0.5 standard deviation apart. However, in practice, the test administrators have fixed the conversion rules from raw subtest scores to stanines based on some base year or test sample. Subsequent changes in distributions of raw test scores cause the mapping from raw test scores to stanines to deviate from the theoretical definition of stanines. Henceforth we refer to these officially defined values simply as stanines, but it is important to note that their distribution varies by country and over time. Figure 1 shows the distribution of cognitive scores by stanine in each country.

We use test score stanines because they are available in both countries. Finnish data also includes the raw test scores, i.e., the number of correctly answered questions. Using these data, we show that it makes very little difference to results whether we use stanines or raw scores. For each country, we transform the cognitive scores to base-year standard deviations (SD), using 1962, the earliest birth year in our study, as the base year. The base-year distribution of cognitive scores has then, by definition, mean 0 and standard deviation 1 in both countries. The resulting population average is slightly higher in Finland due to a more pronounced trend in test scores (the Flynn effect) during this period.

¹These military tests and their institutional background have been described in detail in earlier research. For Finland, see Jokela et al. (2017), especially the supplementary appendix. For Norway, see Sundet, Barlaug, and Torjussen (2004).

Earnings data

Our earnings measure derives from administrative panel datasets that detail the annual incomes of all residents. The Finnish data originates from the income panel data module (“FOLK-tulo”) of Statistics Finland, while the Norwegian data comes from the earnings register of the national welfare and tax administrations. Neither dataset is top-coded. Since we define our study population as being resident throughout the age range 35-45, we observe the yearly earnings in this age range for everyone. We do not drop zeros. We define earnings as the sum of wage earnings and entrepreneurial income. In terms of calendar years, the earnings data ranges from 1997 (when the earliest cohorts were 35) to 2020 (when the latest cohorts were 45). We deflate nominal earnings to 2020 levels using each country’s Consumer Price Index. Our measure of individual earnings is the average over these 11 years of earnings. We calculate earnings percentiles based on the earnings rank within each birth cohort of native-born men. The ranks are calculated within birth cohorts because otherwise the top percentiles would be disproportionately drawn from later cohorts due to real wage growth across the sample period. Earnings percentiles are defined relative to the entire birth cohort within the study population including men without test scores.

	Mean	SD	Median	N
<i>Finland</i>				
Earnings (€k/year)	41,2	31,5	38,2	445 451
Cognitive score (1-9)	5,43	1,90	5	367 566
(base year SDs)	0,19	0,99	-0,03	367 566
Cognitive score (0-120)	67,0	18,5	68	367 566
(base year SDs)	0,19	0,98	0,24	367 566
<i>Norway</i>				
Earnings (€k/year)	55,9	50,0	51,5	417 227
Cognitive score (1-9)	5,10	1,84	5	384 202
(base year SDs)	0,09	0,98	0,04	384 202

Table 1: Summary statistics.

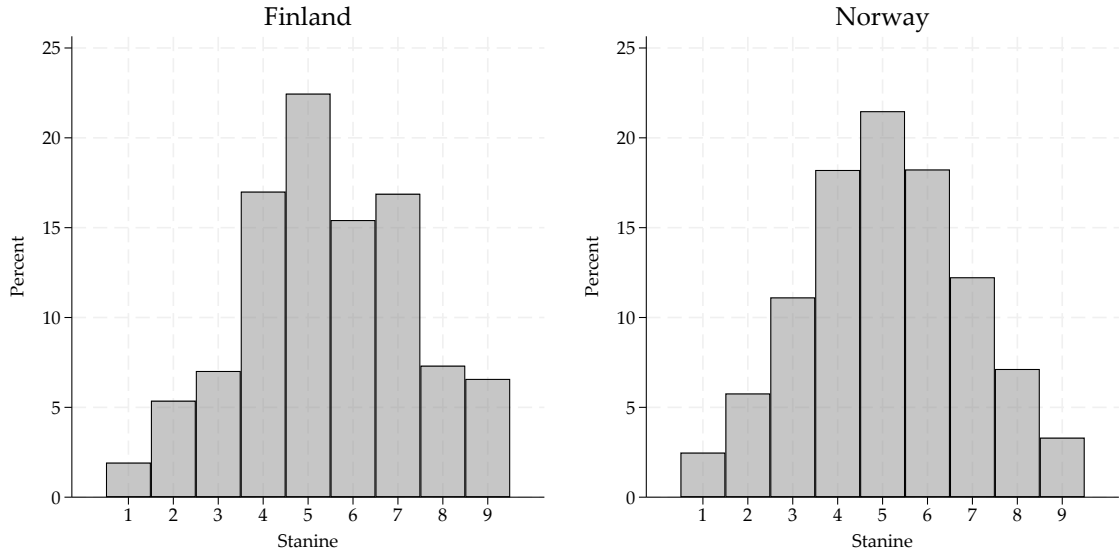


Figure 1: Distribution of cognitive scores by stanine.

3 Results

Main results

We begin by describing the relationship between average cognitive ability and earnings. Figure 2 shows the average cognitive score by earnings percentile in each country. Beyond the unsurprisingly positive overall relationship, two distinct commonalities are visible. The relationship gets steeper near the top, and there is a distinctly low-scoring segment at the very bottom. This steepening near the top is more pronounced in Finland, where there is also a wider almost-flat section below the median. Yet, qualitatively, the relationship is remarkably similar in the two countries.

In both countries, the difference in average cognitive ability between top and median earning brackets is roughly twice as large as the difference between median and bottom brackets. In this sense the relation between ability and earnings is convex. Furthermore, the difference in average cognitive ability between neighboring earnings percentiles gets larger at the top. Within the top decile this “ability gradient” is on average 0.046 and 0.037 standard deviations per earnings percentile in Finland and Norway respectively; below the top decile it is on average 0.013 and 0.012 respectively. This increasing steepness holds even within the top decile, with the highest gradient observed between the top two earnings percentiles. The relationship between cognitive ability and earnings is at its steepest at the top.

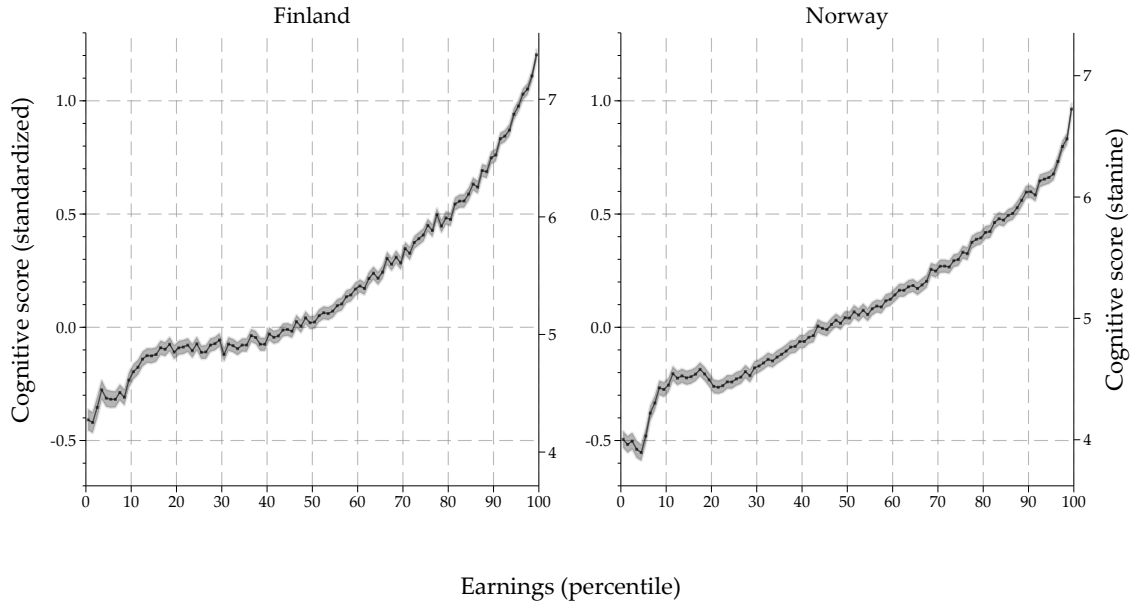


Figure 2: Average cognitive ability by earnings percentile, with 95% confidence intervals shown by shaded bands. The left axis depicts test scores in terms of base year standard deviations, the right axis shows the stanines before standardization. In Appendix A we present a version of this figure with levels of earnings instead of percentiles.

In Figure 3, we plot the shares of cognitive test scores in the lowest and highest stanines by earnings percentile. Recall from Figure 1 that the distribution of scores by these officially defined stanines differs between countries, with Finland having more mass in the upper stanines. Nevertheless, the key qualitative features are again similar in both countries. Figure 3 reveals a clear asymmetry in the shares of “mismatched” individuals. While a significant fraction of high-scoring men ends up with very low earnings, the lowest cognitive scores are virtually absent among the top earners.

Figure 3 also shows that the dip in average cognitive scores at the bottom is due to men with the lowest test scores being disproportionately represented in the lowest earnings percentiles. In both countries, there is a conspicuous downward jump in the share of men with bottom-stanine test scores around the 5th-10th earnings percentile. This jump is more pronounced in Norway. This may be due to men with various problems being selected out of the Finnish test score data; unlike in Norway, in Finland the tests are only administered to men deemed fit for the military service.

Next, we examine the variability of cognitive ability across the earnings distribution in a more concise manner. We summarize this variability by calculating the

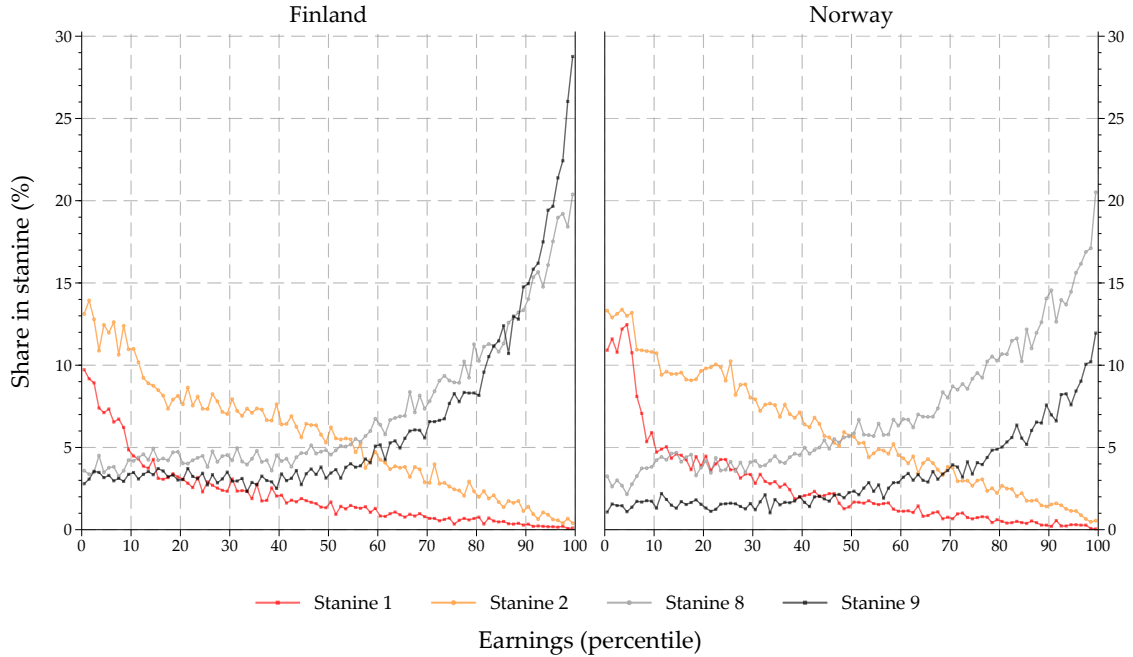


Figure 3: Shares of men with cognitive ability scores in lowest and highest stanines across the earnings distribution.

standard deviation of cognitive scores within each earnings percentile. (These standard deviations are calculated for the standardized stanine-scale scores depicted in Figure 2.) Figure 4 illustrates the results, showing that variability is decreasing in earnings at both ends of the distribution and remains roughly constant in the middle. Once again, the pattern is consistent in both Finland and Norway, though with higher magnitudes for the negative slopes in Finland. The results imply that observed earnings are more informative about cognitive ability among high earners than among low earners.

So far, we have explored the question of what can be inferred about cognitive ability based solely on labor market earnings. Next, we flip the perspective between the two variables. Figure 5 illustrates this by plotting earnings against cognitive ability. Specifically, this figure shows the average earnings percentile within each cognitive test score bin, plotted against the corresponding cognitive score quantiles. We define the cognitive quantile of each bin to be the midpoint of the percentile range that it covers. With stanine-scale data we only get nine plot points for this analysis. With the raw test scores available for Finland, we can calculate many more percentiles. Raw test scores measure the number of correctly answered questions (out of 120) and are thus much more granular than the 9-point stanines. Fortunately, these more detailed results line up well with the stanine-based plot, suggesting that the picture

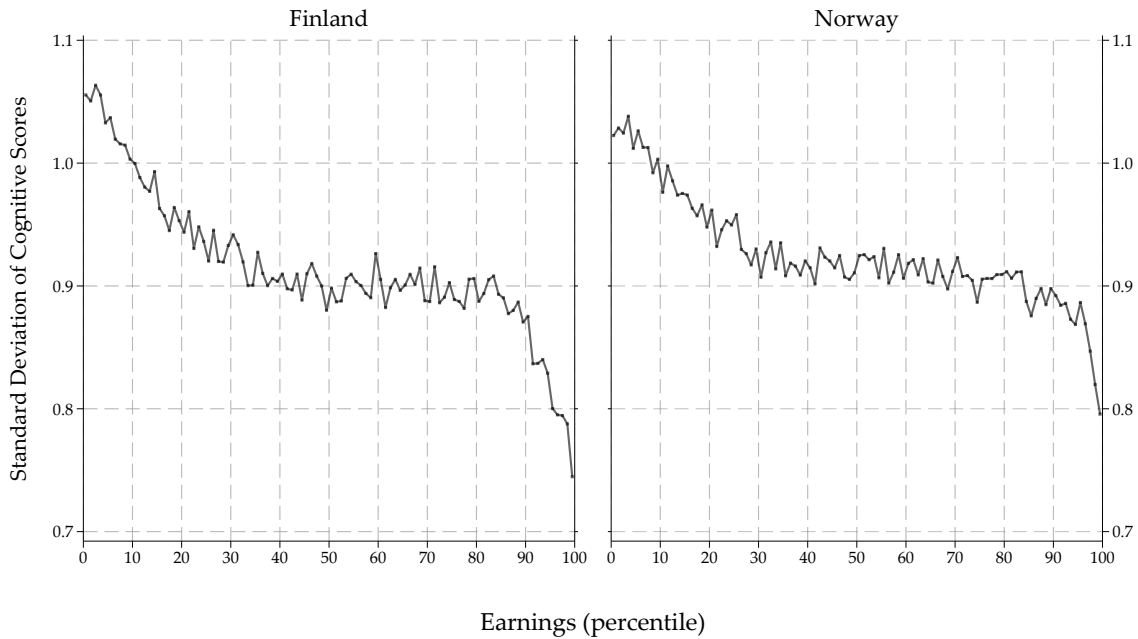


Figure 4: Standard deviation of cognitive scores by earnings percentile.

for Norway is unlikely to be distorted by the stanine scale’s coarseness.

Figure 5 answers the question: How does ranking by cognitive ability predict ranking by earnings? The resulting graph has a reverse-logistic shape in both countries. It is close to linear between the 20th and 80th percentiles, and has steeper slopes near both extremes of the distribution. Men with test scores in the top stanine outearn on average about 70 percent of their birth cohort, whereas for men in the bottom stanine for this figure is less than 30 percent. Although the purpose of our analysis is purely descriptive, direct reverse causality can easily be ruled out here, as ability is measured many years before earnings. However, it is important to note that many unobserved factors are likely to affect both cognitive scores and earnings.

Figure 5 does not support Gladwell’s (2008) assertion that “additional IQ points [above 120] doesn’t seem to translate to any measurable real-world advantage”.² IQ points are standardized at mean 100 and standard deviation 15. In our data a threshold of 120 points corresponds to a point just below the top decile of cognitive scores: in the Finnish test score distribution this corresponds to 88th percentile.

²Gladwell attributes the idea to Arthur Jensen (1980), who phrased it more mildly but at a lower threshold. “Beyond [about 115] the IQ level becomes relatively unimportant in terms of ordinary occupational aspirations and criteria of success. That is not to say that there are not real differences between the intellectual capabilities represented by IQs of 115 and 150 or even between IQs of 150 and 180. But IQ differences in this upper part of the scale have far less personal implications [...] and are generally of lesser importance for success in the popular sense than are certain traits of personality and character.”

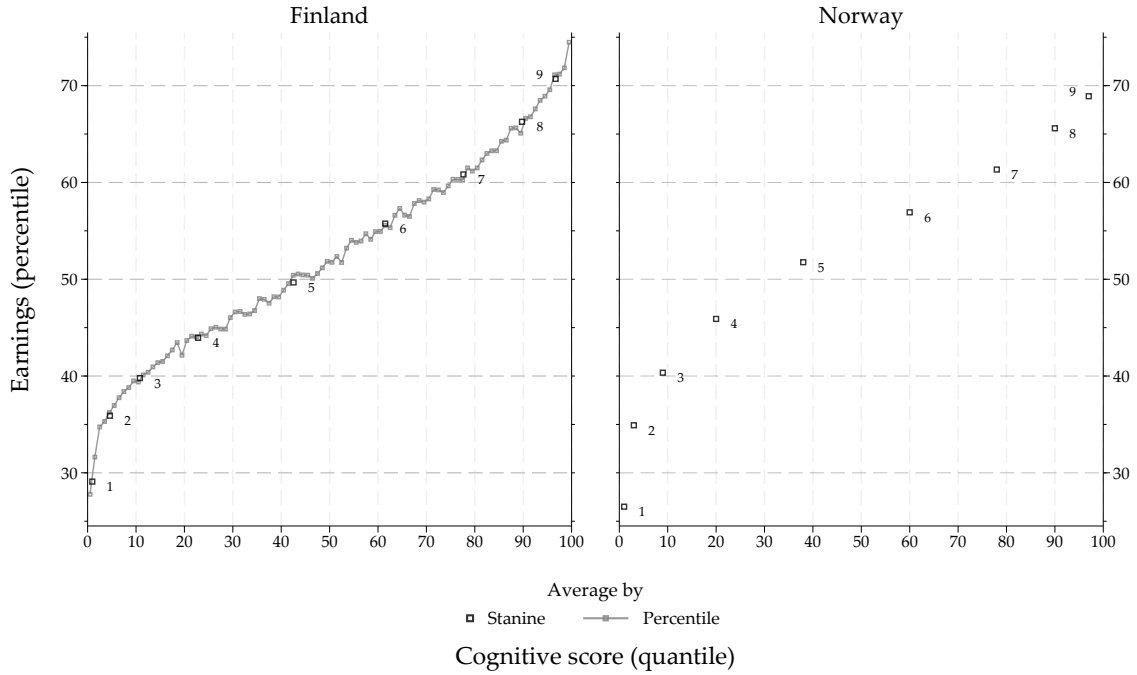


Figure 5: Average earnings percentile by the percentile of cognitive ability. The percentiles for cognitive score stanines are defined to be at the midpoint for each stanine. In Appendix A we present a version of this figure with the levels of earnings instead of percentiles.

(Using casual interpolation between the stanines, the Gladwell threshold is located roughly at that same percentile in Norway). We can measure an advantage that keeps getting larger at higher levels of cognitive ability in our dataset, but, to be sure, the advantage is not vast even at the top.

In Appendix B we present additional results that, due to data limitations, can only be obtained for one country. From Norway, we use cognitive test scores for 12 additional birth cohorts (1950 through March 1961) to show that the patterns described here hold also for this earlier set of cohorts, and also at a later age (45-55). For Finland, we show that the shape of the relationship between cognitive ability and earnings is very similar for each of the three subscores, although slightly stronger for mathematical ability and slightly weaker for abstract (visuospatial) reasoning. With Finnish data, we also show that the anchored score in a separate non-cognitive test has a very similar “steeper-at-the-top” relationship with earnings as cognitive scores.

Comparison with Sweden (Keuschnigg, van de Rijt, and Bol 2023)

The above results from Finland and Norway differ strikingly from the headline results of an ability plateau for high earners recently reported using similar data from Sweden (Keuschnigg, van de Rijt, and Bol 2023). Specifically, the Swedish paper finds that average cognitive ability “plateaus” or levels off across the top earnings percentiles, with indications of a decline (negative ability-earnings gradient) across the top three percentiles. Not only do we fail to replicate this plateau result in the Finnish and Norwegian data, we find the opposite: the relation between cognitive ability and earnings gets stronger at the top, and is at its steepest between the top percentiles. Beyond this, the shape and magnitude of the ability-earnings relationship appear largely similar across all three countries, with the exception of the bottom decile where the Swedish curve has a negative slope across percentiles.

It is difficult to understand why the earnings-ability relationship should differ so markedly for top earners across these three neighboring countries. The main methodological difference to our study is that Keuschnigg et al constrain their analysis to males identified as entering the labor market within the 1991-2003 period. This involves a number of sample selection rules based on individual labor market history that serve to reduce their analysis sample to about 20% of the full birth cohorts. In this subsection we attempt to mimic these sample selection criteria so as to assess their impact on our main results.

Another difference in our approach is the definition of the income variable. We used all earnings, including entrepreneurial income, whereas Keuschnigg et al use wage earnings only. Entrepreneurial income (distinct from capital income) is the analogue of wage earnings for small-scale entrepreneurs, who may be classified as zero earners if only wage earnings are considered. In what follows we will also modify our income variable and use wage earnings. Finally, they calculate the percentiles by pooling observations in their final sample, whereas we have calculated the percentiles within the birth cohort of all men not just those with cognitive scores. To ease the comparison we calculate wage percentiles using this same approach.

Keuschnigg et al (2023) use labor market data from calendar years 1991-2012.³ To be included in their sample a man must have entered the labor market in the years 1991-2003. They operationalize this requirement with two sample selection rules: all

³Our description of the Keuschnigg et al (2023) sample selection rules is based on the article and personal communication with Marc Keuschnigg.

subjects (i) must not be classified as full-time employee in 1990 and (ii) must have at least one year of full-time employment during 1991-2003. Their individual income measure is the yearly average at age 35-45, requiring all years to be observed (for this choice we followed their example already in our main analysis). Together these choices imply that the men in their analysis sample must have been born between 1956 (to not exceed age 35 in 1991) and 1967 (to reach age 45 by 2012). From these birth years they end up with a sample of 59,387 men, which is about the number of males in a single Swedish birth cohort.⁴ We infer that their sample must cover less than 20% of Swedish-born men from the birth cohorts that could in principle satisfy both the employment history-based requirements and have their earnings observed throughout age 35-45.

To construct our restricted sample of conscripts, we drop those who fail to satisfy an analogous set of employment history requirements, but we shift the calendar years of all requirements forward by six years. This shift accounts for the fact that the first cohort in our data is born six years later than the first cohort in the Swedish sample. The requirement not to be employed in the year 1990 means that each Swedish birth cohort had to be non-employed at a different age, ranging from age 23 for the youngest cohort to age 34 for the oldest. By shifting the calendar years in the sample selection rules, we attempt to mimic the implied age distribution of their employment and non-employment requirements. Therefore, we require men to be defined as not in full-time employment in 1996 and employed in at least one of the years between 1997-2009.

For Norway, in alignment with the Swedish sample, we use a variable from administrative employer-employee records that specifies whether the employee in 1996 held a full-time position. In Finland, we employ a classification by Statistics Finland that records a person's main activity at the end of the calendar year as employment (as opposed to study, unemployment, non-employment, etc).

The data coverage is illustrated in Figure 6. We measure coverage as a percentage of the full sample, i.e., all native-born men from birth years 1962-1975 who are present throughout age 35-45. The full sample is divided into earnings percentiles within their own birth cohorts. Since administrative data includes earnings data for everyone (even if zero) the coverage for the full sample is by definition 100 %. However, the coverage of cognitive scores is, of course, lower, and varies by earnings.

⁴They omit test results from years 1978-1979 in order to reach 90% test data coverage for every conscription cohort, which should affect the cohorts that reached age 18 or 19 in these years.

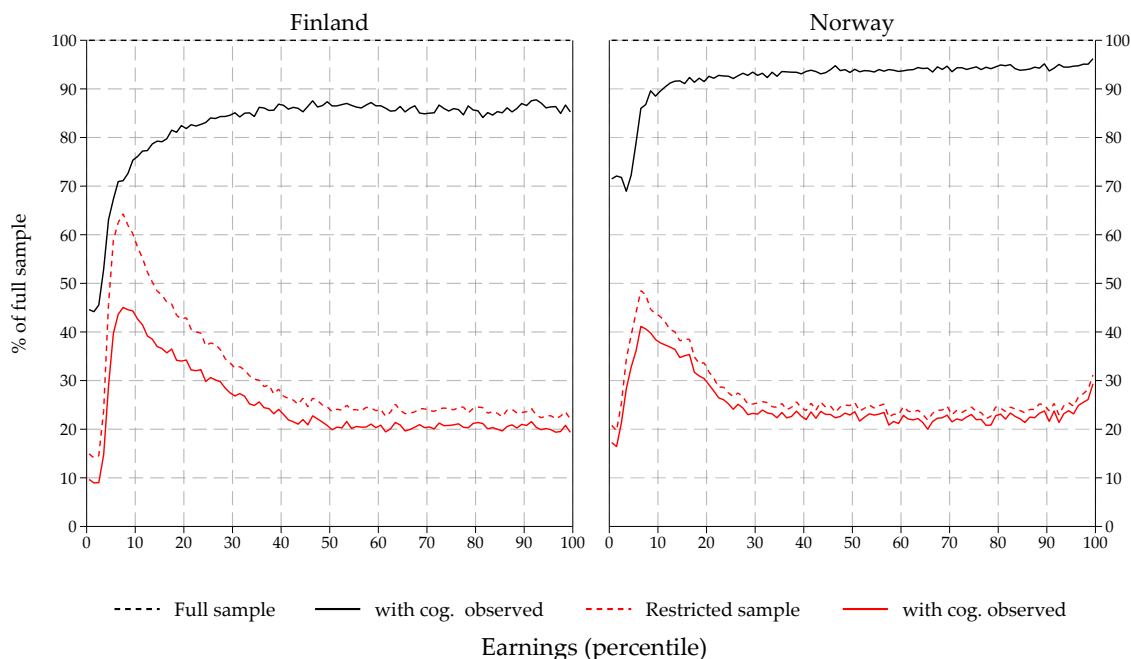


Figure 6: Data coverage by earnings percentile relative to the full sample (native-born men observed at age 35-45). Restricted sample attempts to replicate the sample selection rules in Keuschnigg et al (2023).

The bottom decile of earners is clearly less covered, and this selection is much more pronounced in Finland where the cognitive test is taken during military service. In Norway, test scores are observed for 92 % of the full sample, and for over 90% in all but the lowest decile of earners. In Finland the test score data covers 82 % of the full sample, and hovers around 85% above the bottom two deciles of earners.

The application of the sample selection rules leads to a significant reduction in the sample coverage, resulting in approximately 25% of the full sample remaining in both countries. There is a noticeable bump in the coverage curve just above the segment with the lowest coverage at the bottom of the earnings distribution. This bump is more pronounced in Finland but peaks just below the 10th percentile in both countries.

This bump seems to stem from the requirement of non-employment in the first calendar year (1996). Those not employed in a given year as young adults—ranging between ages 21 and 34, depending on the cohort—are more likely to have relatively low earnings later on, contributing to this feature. For those at the very bottom of the earnings distribution, this requirement is less significant, as they fail to appear in the test score data for other reasons and are more likely to never have a year of full-time employment.

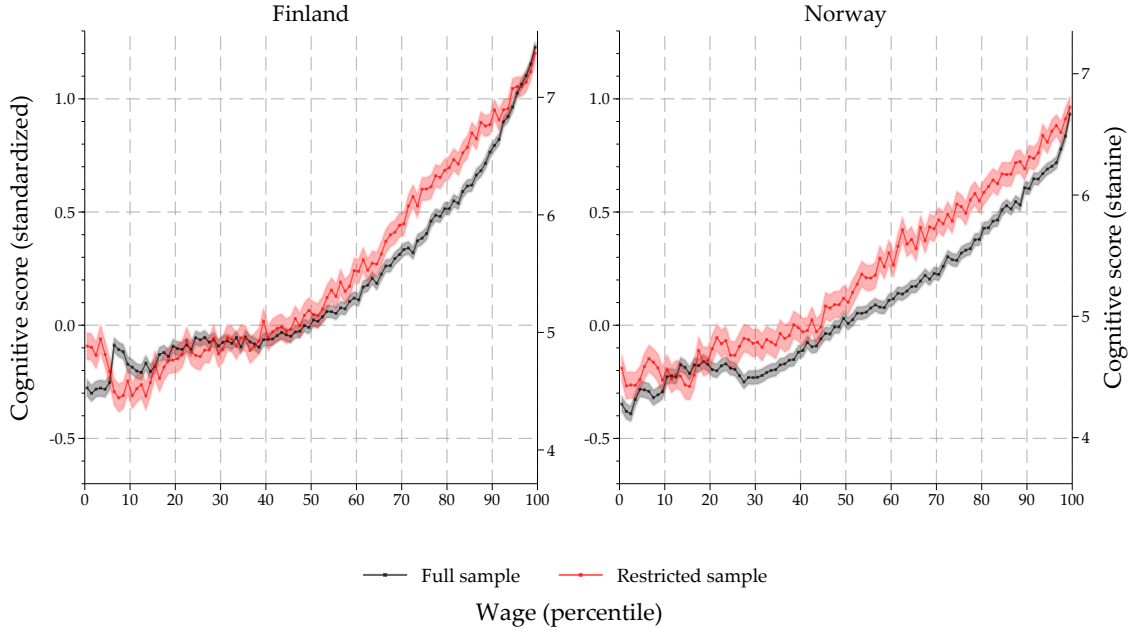


Figure 7: Average cognitive ability by wage percentile, with 95% confidence intervals shown by shaded bands. Full sample refers to the same sample as in Figure 2. Restricted sample attempts to replicate the sample selection rules in Keuschnigg et al (2023).

Finally, Figure 7 shows the average cognitive score by wage percentile for both the full sample and the restricted sample. Here the percentile is calculated within the entire respective samples (not by birth cohort). The restrictions make very little difference to the overall shape of the relationship in either country. In Norway the restricted sample has about 0.1 to 0.2 standard deviations higher test scores in a swathe of middle percentiles, but the mean scores in the top percentile are almost unaffected by the restrictions. Naturally the confidence intervals become wider with the smaller sample, but there is no sign of any kind of plateauing at the top. (The impact of sample restrictions on other figures is equally underwhelming so we omit them). We conclude that sample selection criteria are unlikely to explain the difference between our results and those of Keuschnigg et al (2023).

While we do not have a theory as to why the ability-earnings relationship should get steeper at the top, it is worth noting that Keuschnigg et al (2023) interpret the plateau at the top as evidence for two concurrent regimes of social stratification. First, the bulk of the population are sorted into earnings ranks based on “merit” (cognitive ability), resulting in a clear gradient. Second, a subset are pushed into the tails of the earnings distribution based on factors other than cognitive ability, such as luck and family connections. This leads to a type of mean reversion in ability at

both ends of the earnings distribution, creating a plateau or even a decline across the top percentiles given a sufficient number of such “non-meritocratic” high earners. The lack of a plateau in Finnish and Norwegian data could simply mean that these countries have *too few* non-meritocratic high earners to depress the *average* cognitive ability in the top percentiles. Even so, they would be expected to increase the ability *dispersion* at the top. As shown in both Finnish and Norwegian data, however, the ability dispersion falls markedly towards the top, with the top percentiles showing the *lowest* variation in cognitive ability (see Figure 4).

4 Conclusion

Using register data on earnings and cognitive test scores for 14 birth cohorts of men in Finland and Norway, we find average cognitive ability increasing with earnings—and more so at the top. The gradient in average cognitive ability is three to four times higher within the top decile of earners than below (in standard deviations per earnings percentile). The relationship of ability and earnings grows strongest at higher levels of earnings also in the sense that the variability of cognitive ability gets smaller and the share with the lowest scores becomes vanishingly small.

Our empirical results were obtained in parallel using administrative data in two countries, and our findings are very similar in both countries. We believe this is the best descriptive evidence to date on the shape of the relationship between labor market earnings and cognitive ability. The most significant limitation is that the cognitive test scores in our datasets only cover the male population. With that caveat we conclude that the overall relationship between earnings and ability is increasing and more so at the top. The relationship is at its strongest among top earners also in the sense of ability having the lowest variability in the top earnings percentiles.

The Finnish and Norwegian evidence is inconsistent with recent results from Sweden, which found average cognitive ability plateauing across higher earnings percentiles. The Swedish pattern was argued to reflect an inflow of lower-ability individuals into the top earning brackets on the basis of luck or family wealth and connections (Keuschnigg et al, 2023), which would also result in increased ability dispersion in the tails. In the Finnish and Norwegian data, however, the dispersion is substantially reduced in the upper tail. The country differences are surprising given the institutional similarities of these three neighboring countries, but similar results have been

reported for Sweden in the past: Björklund, Roine, and Waldenström (2012), when investigating mechanisms for intergenerational persistence in top decile incomes for Swedish father-son pairs, also noted a flat or declining cognitive-ability gradient within the top decile of fathers' incomes.

We do not have an explanation for why the relationship between cognitive test scores and earnings should differ so markedly for Swedish high earners compared to Finnish and Norwegian. There are no obvious major differences in institutions or data quality between these countries. The most straightforward explanation is that the economic processes that determine top earnings just are that different in Sweden. Another possibility is that the cognitive test used by the Swedish military is not as discriminating between high levels of ability as its Finnish and Norwegian counterparts. If this were the case, then it might still be possible to find other measures of ability in Sweden that do not “plateau” near the top.

A third possibility is that top earnings are measured with significantly more error in Sweden than in Finland or Norway; such measurement error would have similar empirical implications as “luck” in the model of earnings invoked by Keuschnigg et al (2023) and originally proposed by Denrell and Liu (2012). Recent findings by Bastani and Waldenström (2021), who also use the Swedish military test data, suggests a possible channel that would be based on peculiarities in the Swedish tax code. They study the relationship between individual tax responsiveness and cognitive ability. They find that men with high cognitive ability react more to tax incentives than do low-ability individuals; the main mechanism is a smart use of methods to transform labor earnings into business income in a way that reduces the tax burden.

In their conclusion, Keuschnigg et al (2023) note that their analysis is “limited to a single country” and call for research from other countries “to evaluate to what extent [the] findings generalize.” Based on our work, we conclude that their Swedish results fail to generalize to Sweden’s two neighboring countries. This illustrates the importance of replication and evidence from multiple cases before general conclusions can be drawn.

Finally, it is important to note that, while the average relation between cognitive ability and earnings in the population is very strong, most variation in earnings does not relate to cognitive ability. Individual ability cannot be reliably inferred just from observed earnings or vice versa. For example, even among the top 1% of the earnings distribution, where the variation in cognitive ability is at its lowest, 9%

of men in Finland and 14% in Norway scored below the population median in the cognitive test. Conversely, among the bottom percentile of the earnings distribution, about a 30% of men scored above the median in both countries. (We present these shares for all percentiles in Figure A3 in Appendix A.) The larger point made by the earlier literature is that the observed ability-earnings relationship is compatible with multiple blends of causal stories and interpretations, and the fact that there is an ability-earnings gradient does not tie us to any specific normative judgment.

References

Barth, Erling, Karl O. Moene, and Fredrik Willumsen. 2014. “The Scandinavian Model—An Interpretation.” *Journal of Public Economics* 117 (September): 60–72. <https://doi.org/10.1016/j.jpubeco.2014.04.001>.

Bastani, Spencer, and Daniel Waldenström. 2021. “The Ability Gradient in Tax Responsiveness.” *Journal of Public Economics Plus* 2 (January): 100007. <https://doi.org/10.1016/j.pubecp.2021.100007>.

Björklund, Anders, Jesper Roine, and Daniel Waldenström. 2012. “Intergenerational Top Income Mobility in Sweden: Capitalist Dynasties in the Land of Equal Opportunity?” *Journal of Public Economics* 96 (5): 474–84. <https://doi.org/10.1016/j.jpubeco.2012.02.003>.

Bratsberg, Bernt, Ole Rogeberg, and Vegard Skirbekk. 2023. “The Relevance of Cognition In The Context of Population Ageing.” In *The Routledge Handbook of the Economics of Ageing*. Routledge.

Calvin, Catherine M., Ian J. Deary, Candida Fenton, Beverly A. Roberts, Geoff Der, Nicola Leckenby, and G. David Batty. 2011. “Intelligence in Youth and All-Cause-Mortality: Systematic Review with Meta-Analysis.” *International Journal of Epidemiology* 40 (3): 626–44. <https://doi.org/10.1093/ije/dyq190>.

Denrell, Jerker, and Chengwei Liu. 2012. “Top Performers Are Not the Most Impressive When Extreme Performance Indicates Unreliability.” *Proceedings of the National Academy of Sciences* 109 (24): 9331–36. <https://doi.org/10.1073/pnas.1116048109>.

Gladwell, Malcolm. 2008. *Outliers: The Story of Success*. Hachette UK.

Hegelund, Emilie Rune, Trine Flensburg-Madsen, Jesper Dammeyer, and Erik Lykke Mortensen. 2018. “Low IQ as a Predictor of Unsuccessful Educational and Occupational Achievement: A Register-Based Study of 1,098,742 Men in Denmark 1968–2016.” *Intelligence* 71 (November): 46–53.

<https://doi.org/10.1016/j.intell.2018.10.002>.

Jantti, Markus, Bernt Bratsberg, Knut Roed, Oddbjorn Raaum, Robin Naylor, Eva Osterbacka, Anders Bjorklund, and Tor Eriksson. 2006. “American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States.” SSRN Scholarly Paper ID 878675. Rochester, NY: Social Science Research Network.

<https://papers.ssrn.com/abstract=878675>.

Jensen, Arthur Robert. 1980. *Bias in Mental Testing*. Free Press.

Jokela, Markus, Tuomas Pekkarinen, Matti Sarvimäki, Marko Terviö, and Roope Uusitalo. 2017. “Secular Rise in Economically Valuable Personality Traits.” *Proceedings of the National Academy of Sciences* 114 (25): 6527–32.

<https://doi.org/10.1073/pnas.1609994114>.

Keuschnigg, Marc, Arnout van de Rijt, and Thijs Bol. 2023. “The Plateauing of Cognitive Ability among Top Earners.” *European Sociological Review*, January, 39 (5): 820–833.

<https://doi.org/10.1093/esr/jcac076>.

Lindqvist, Erik, and Roine Vestman. 2011. “The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment.” *American Economic Journal: Applied Economics* 3 (1): 101–28.

<https://doi.org/10.1257/app.3.1.101>.

Sundet, Jon Martin, Dag G. Barlaug, and Tore M. Torjussen. 2004. “The End of the Flynn Effect?: A Study of Secular Trends in Mean Intelligence Test Scores of Norwegian Conscripts during Half a Century.” *Intelligence* 32 (4): 349–62.

<https://doi.org/10.1016/j.intell.2004.06.004>.

Appendix A

In Appendix A we show additional results that were briefly mentioned in the main text. Figures A1 and A2 show versions of Figures 2 and 5 where the unit-free earnings percentiles have been replaced with average monetary values within each bin. We use the average exchange rate in 2020 to convert Norwegian kroner to euros. In these figures the plot points are the same but their positions along the earnings-axis are stretched differently depending on the shape of each country’s earnings distribution.

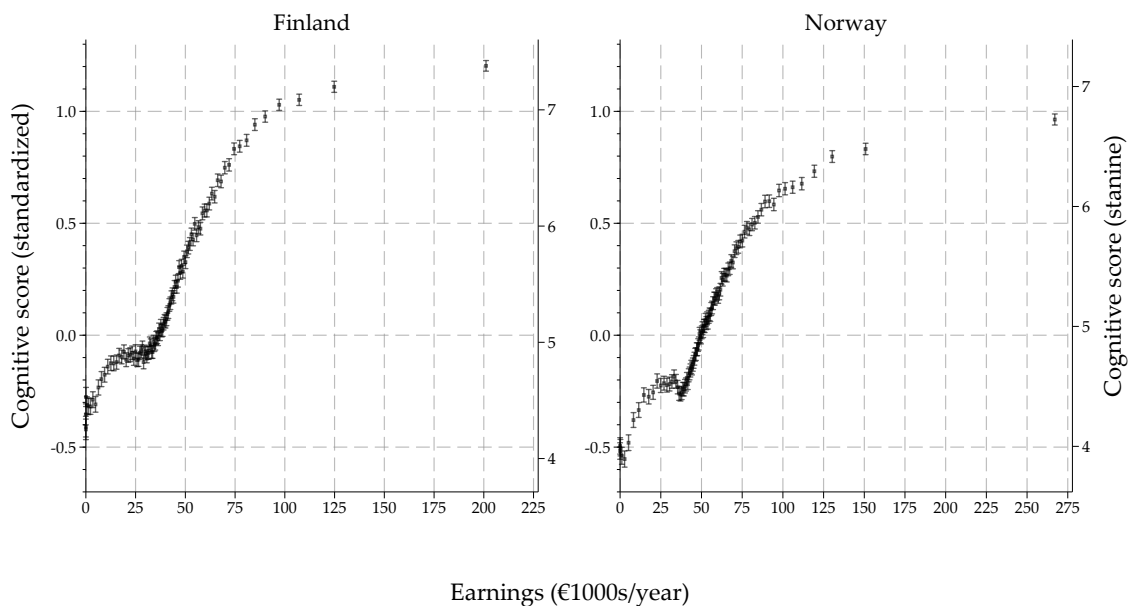


Figure A1. This figure is otherwise the same as Figure 2, but horizontal axis depict the average level of earnings within each earnings percentile.

Figure A3 illustrates how the “mismatch” between earnings and cognitive ability varies along the earnings distribution. Here we plot the shares of men who have cognitive scores above or below median, and the (absolute value) of their difference, in order to contrast the behavior of “mismatch” at upper and lower deciles of the earnings distribution. It follows almost directly from the fact that average ability is monotonically increasing in earnings that this difference gets larger towards the tails of the earnings distribution, because there is less and less “space” to be an outlier on one side of the average. However—and this is the point of the graph—there is a clear asymmetry in the behavior of the mismatch between top and bottom halves of the earnings distribution. It is steeply increasing along the earnings percentiles in the top half of the earnings distribution, but flatter among the bottom earnings percentiles.

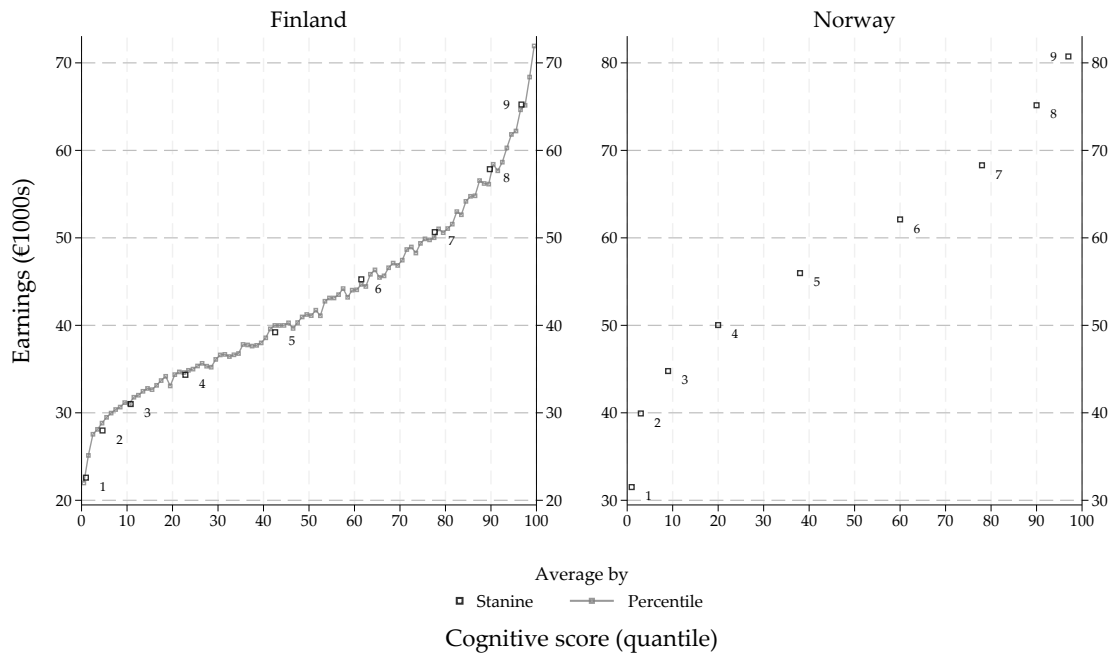


Figure A2. This figure is otherwise the same as Figure 5, but vertical axis depict the average level of earnings within each cognitive score quantile.

Low cognitive scores are much rarer among top earners than high cognitive scores among bottom earners. This asymmetry is slightly more pronounced in Finland.

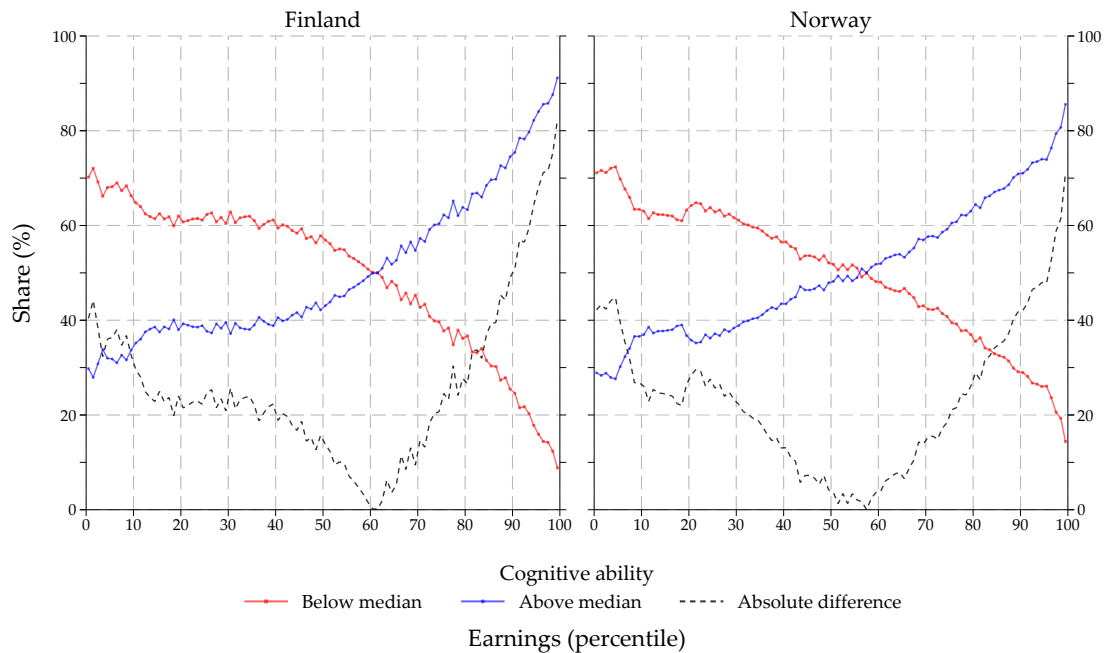


Figure A3. Share of men with cognitive ability scores below or above median within each earnings percentile, and the absolute value of the difference between these shares.

To construct Figure A3 we assigned a status “below” and “above” median for those at the median value (stanine 5) by assuming that, at every earnings percentile, an underlying continuous ability level is uniformly distributed within the median stanine.

Appendix B

In Appendix B, we present supplementary results that can only be obtained for one country, due to data limitations. These results all show the same analysis as Figure 2, but with different variables or in different datasets. All exhibit the same “steeper at the top” feature in the relationship between average ability and earnings.

For Finland, we first use the subscores of the cognitive test. One of the subscores, “Visuospatial”, measures abstract reasoning ability using an approach akin to Raven’s matrices, which are commonly used to measure fluid intelligence or “IQ”. The correlations between these subscores range from 0.61 to 0.71. The left panel of Figure A4 shows that all subscores have separately a very similar relationship with earnings as the coarser stanine measure in Figure 2. Arithmetic (mathematical) ability has a slightly higher gradient within the top decile of earnings than have the other subscores. There is, however, no indication of plateauing in any of the subscores.

The right panel of Figure A4 uses data from the non-cognitive test, which is a separate standardized test administered to military conscripts in Finland. It has 8 subscores, which we aggregated to a single measure using the same anchoring methodology as Jokela et al 2017; see that article also for a detailed description of the non-cognitive test. Briefly, the anchored score is a fitted regression value: it is that weighted average of the subscores that best predicts earnings. Measured in base year standard deviations, the shape and magnitude of the ability-earnings relationship is once again remarkably similar: relatively flat below median earnings, then ever steeper towards the top. The main difference is that, for non-cognitive scores, the dip at the bottom decile is milder than for cognitive scores.

For comparison, the right panel of Figure A4 plots also anchored scores for the cognitive test, and a combined anchored score that incorporates all 11 subscores of cognitive and non-cognitive tests into one measure. The correlation between the anchored scores of the two tests is 0.42. The measure that combines information from both tests shows the steepest gradient of any measure, but the difference to

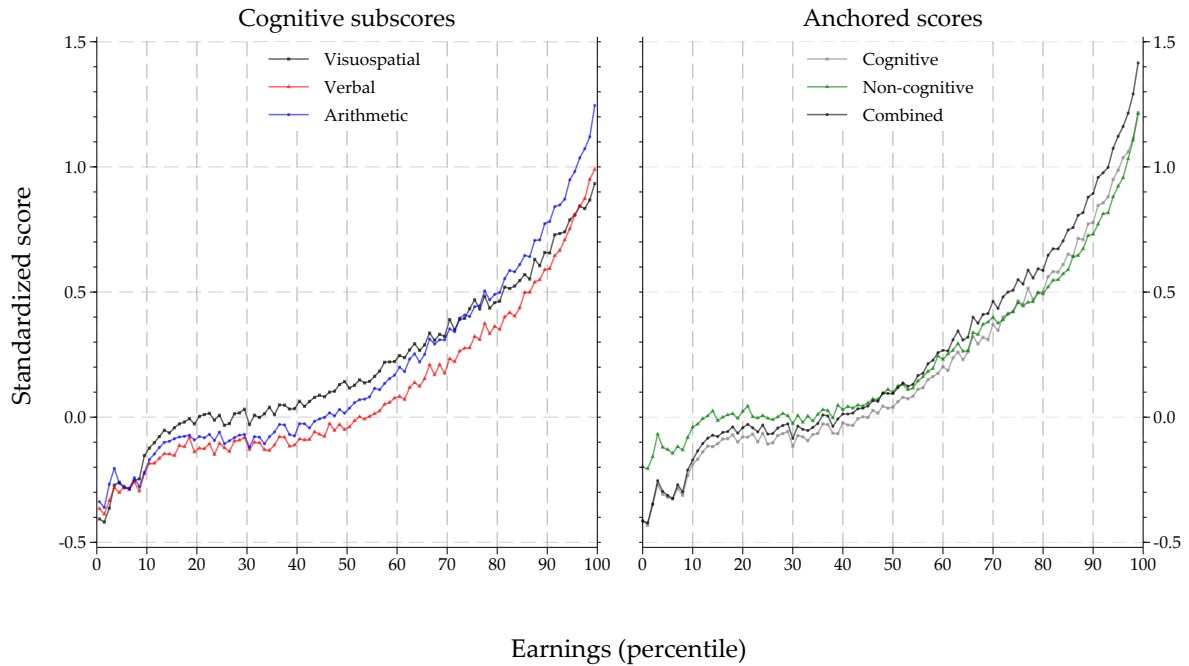


Figure A4. Additional results with Finnish data. Average ability score by earnings percentile in terms of base year standard deviations. Left panel: Subscores of the cognitive test. Right panel: Anchored scores from cognitive, non-cognitive, and combined test results.

the test-specific scores is small.

Similar Swedish data has been analyzed by Lindqvist and Vestman (2011), who found that cognitive ability is a stronger predictor of wages for conscripts who earn above the median, whereas noncognitive skill is more important for low earners. However, in Sweden the noncognitive skill measure is based on an interview by a psychologist.

For Norway, conscription data are available for male birth cohorts since 1950. Figure A5 shows the association between earnings rank and cognitive conscription scores for those born before the cohorts used in the main analysis (where the oldest cohort was born in 1962). Because test scores were re-normed in 1980, affecting those born in April 1961 and later, in this appendix we restrict the data to males born between 1950 and March 1961. These cohorts can be followed into a higher age, so we supplement the 35-45 earnings rank measure used in the main analysis with ranks based on earnings between ages 45 and 55, applying equivalent sample restrictions to the higher age bracket. As the figure shows, whether based on earnings over the age span 35-45 or 45-55, the earnings-cognitive ability relationship of the older cohorts displays the same features as that of the 1962-1975 cohorts (see Figure 2). If

anything, the slope of the earnings-ability curve is slightly steeper at the top when based on earnings observed between ages 45 and 55 than between ages 35 and 45.

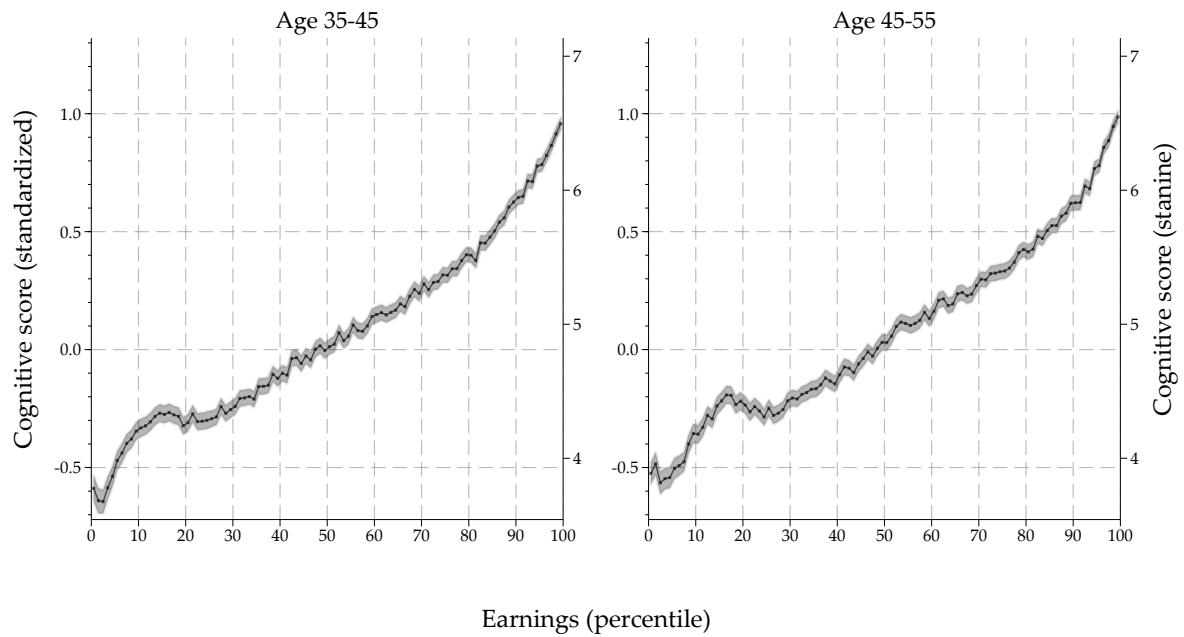


Figure A5. Additional results with Norwegian data, male birth cohorts from 1950 through March 1961. While the test remained unchanged, Norwegian conscription test scores were re-normed for those born after March 1961.