On the Shape of the Ability-Earnings Relation*

Marko Terviö

Aalto University School of Business,
Helsinki GSE

First draft – very preliminary

Febuary 19, 2023

Abstract

I describe the shape of the relationship between cognitive (and non-cognitive) ability and earnings using Finnish administrative data. The ability measures are based on standardized tests conducted by the military on conscripts, while earnings are based on population data at age 35–45. The average ability conditional on earnings percentile is roughly flat until about the 40th percentile, clearly increasing thereafter and becomes ever steeper near the top. The average earnings conditional on ability percentile has a clearly positive gradient at all percentiles, again with a steeper slope between the top percentiles. The relationships have a similar overall shape, whether looking at cognitive or non-cognitive ability.

JEL: J24, J31.

---

*E-mail: marko.tervio@aalto.fi. Permission to use the administrative data was granted by Statistics Finland (license TK-53-228-14) and by the Finnish Defense Forces (AJ23378). I thank Ramin Izadi, Mika Päällysaho, and Roope Uusitalo for useful comments and suggestions.

In this short note I describe the shape of the association between cognitive and non-cognitive test scores and later-life earnings in the Finnish administrative data on military conscripts. To probably no one's surprise there is a strong positive relation.[1] Furthermore, the relationship between average ability and average earnings gets steeper near the top of the distribution. This convexity near the top is robust to using standardized variables or percentiles and holds in both directions, that is, whether looking at earnings conditional on ability or vice versa.

The exact shape of the average relationship between earnings and cognitive ability near the top of the earnings distribution has received attention lately due to Keuschnigg, van de Rijt and Bol (2023) who find that the relation plateaus near the top. Using Swedish administrative data on wage earnings and cognitive ability test results of military conscripts they find that average cognitive ability is essentially flat between the top 10 percentiles of wage earners. This surprising finding can lead to some interesting speculations about the economic processes behind wage differences near the top. It is important to see how this relationship looks like in other high-quality population level datasets.

I use Finnish administrative data that is largely comparable to the data available in Sweden. Both countries had comprehensive male conscription for the relevant birth cohorts, so the cognitive ability tests administered by the military achieve a wide coverage. Earnings are measured at age 35–45. The Finnish test score dataset covers the male cohorts born in 1962–79; of these the cohorts 1962–74 are old enough to be in the earnings data for the full age range. The Finnish military also administers a non-cognitive test so I will report results for both cognitive and non-cognitive ability, where these simple terms are shorthand for summary measures that are constructed from subtest scores and explained in detail in Jokela, Pekkarinen, Sarvimäki, Terviö and Uusitalo (2017).

1 Data and definitions

I use the same dataset that was first used and carefully described in Jokela et al. (2017).[2] The only difference is that earnings data has been extended to 2019. The earnings data covers all men and comes from the basic panel data module of Statistics

[1]The literature on the relationships between psychological test results, cognitive and otherwise, and life outcomes including income, is vast. A few surveys to get started with: Borghans, Duckworth, Heckman and Ter Weel (2008), Almlund, Duckworth, Heckman and Kautz (2011). [More cites to come...]

[2]Link to Jokela et al. (2017) (open access): `https://dx.doi.org/10.1073/pnas.1609994114`. Please notice that much of the details are in the supplementary appendix. It is too long to be adequately summarized here.

Finland (FOLK-Perus). Earnings is the sum of wage earnings and entrepreneurial income, deflated by the CPI.

The birth cohorts 1962–79 comprise of 591 thousand native born men, of whom 476 thousand (80%) are observed in the military test data. The military test data was provided by the Finnish Defense Forces (FDF). Using a shorter age range 35–40 of earnings data allows for all birth cohorts to be used, but turns out that this doesn't make much difference to the results either. In most analyses I restrict the cohorts to the range 1962–74 in order to form a panel where all can be observed in the earnings data throughout the age range 35–45; this is the age range used by Keuschnigg et al. (2023). The most restrictive sample I use is a balanced panel for whom both the cognitive ability test and all 11 years of earnings are observed; this covers 336 thousand men. This restriction makes little difference to the shape of the earnings-ability relation.

The FDF administers two separate tests to all conscripts in the early weeks of military service. The cognitive ability test consists of three subtests of 40 multiple choice questions. The visuospatial subtest is based on Raven's matrices and thus measures what is commonly known as the IQ. The two other subtests, Verbal and Arithmetic, measure also skills that students learn in school. The non-cognitive test has 218 yes-no questions, in the data these answers are tallied into 8 subscores such as "self-confidence" and "leadership motivation". Both tests remained essentially unchanged during the data period.

In order to reduce the scores from each test to a one-dimensional variable I use the same anchoring method that we used in Jokela et al. (2017). Briefly, the anchored score is a weighted average of the subscores, where the weights are chosen so that the anchored score is the best linear predictor of later-life earnings. For more context and detail on anchoring see section 2 in Cunha, Nielsen and Williams (2021). Finally, I rescale the anchored scores into base year standard scores using the earliest birth cohort in the data as the base.

Both Finnish and Swedish militaries use a nine-point summary measure of cognitive test results (this measure is colloquially known as "p-luku" in Finland). The stanine (*sta*ndard *nine*) is supposed to divide observations into bins of predetermined width where the middle or 5th stanine is the largest and the extreme stanines the smallest, and where bin borders are 0.5 standard deviations apart. I will show some results that rely on the FDF stanines, because Keuschnigg et al. (2023) use stanines in their study.

Under the pure statistical definition of a stanine the average would always be 5, but at least in Finland the category borders remained fixed for many years while the test results kept creeping up. For the 1962 birth cohort the average stanine is 5.05.

Thereafter scores in top stanines become increasingly common, and apparently the FDF rescaled its definition of stanines in 1996. The average stanine drops from 5.87 to 5.17 between consecutive test years, even while all subtest mean scores keep going up. This means that around the birth cohorts 1975 and 1976 the mapping from FDF stanines to test scores changes abruptly; here the main analysis is not affected because it uses the restricted cohorts sample (1962–74). More generally, the problem of stanine rescaling is easy to avoid by using a summary measure calculated from the raw subscores.

2 No plateauing of ability among top earners

In this section I follow the flipped-axis approach behind the headline "plateauing" result of Keuschnigg et al. (2023). This means dividing individuals into percentile bins by their earnings on the horizontal axes, and depicting the ability measure calculated for each bin on the vertical axis. The main result is Figure 1. The percentile bins are defined with respect to men's average yearly earnings at age 35–45. The percentiles are calculated within all native-born men of the same birth year, including those who are not in the FDF data. The left axis depicts the average stanine in the cognitive test, which is directly comparable with Figure 3.C of Keuschnigg et al. (2023).[3] The graphs look remarkably similar below the top decile of earners. In both countries the average stanine is close to 5 for the lower half of the earnings distribution, and turns upwards around the 40th percentile.

[Figure 1 about here.]

The one conspicuous difference with the Swedish results is the lack of any plateauing in the Finnish data. In fact, here the relation of average ability and earnings percentile becomes steeper near the top of the earnings distribution. That is, measured in standardized test scores, the difference in average ability between neighboring earnings percentiles keeps increasing all the way to the top as we move higher up the earnings distribution. This despite moderate top-coding in the Finnish wage data, which could bias the ability-earnings gradient towards zero near the top. I return to the issue of top-coding in Section 4.

The right axis in Figure 1 depicts the average test results in terms of anchored scores. Anchored scores are essentially continuous and thus much more granular than the 9-point stanine, but this increase in precision makes remarkably little difference

[3] Link to Keuschnigg et al. (2023) (open access): `https://doi.org/10.1093/esr/jcac076`. Figure 3 is on page 8.

to the shape of the relationship. Using the earliest birth cohort as the base year also aligns the zero SDs on the right axis with stanine 5 on the left. In the base year the FDF stanines ("p-luku") still followed the statistical definition of stanines, before a distributional shift to upper stanines in later cohorts, which is why the population average for all earners is clearly higher than the theoretical average 5 for stanines.

Figure 1 also depicts the relation of average non-cognitive score and earnings percentile. The relation looks very similar to that of cognitive ability. The correlation between cognitive and non-cognitive scores is 0.41, so this similarity is by no means mechanical. The FDF data does not include a stanine or other summary measure for the non-cognitive test, hence no left-axis version of the non-cognitive score.

Both skill types show a visible dip among the first few earnings percentiles; a similar dip also shows up in Keuschnigg et al. (2023). At least in the Finnish earnings data one problem is that it can include person-years with very low or zero earnings for individuals that that were not in fact potential earners during the year (or during most of it). There is no reliable measure of months spent unavailable for work. This type of unavailability causes people with earnings potential from all over the earnings distribution to clump near zero. For instance, those who stay abroad without officially moving their residence out of Finland show up with zero or near-zero earnings. This can be thought of as non-classical measurement error, if the point is to uncover a relation between true earnings potential and ability. From the point of view of actual relation of earnings and ability the bottom percentile dip may well just follow naturally from the data-generating process. This is my educated guess for the cause of the bottom-earner dip, but there could also be other types of measurement error in the earnings data that I am not aware of.

The same overall shape of the ability-earnings relationship is visible also in Figure 2, which shows the share of highest-stanine test scores in each earnings percentile. The share is roughly flat until around the median, then increasing and roughly convex, reaching 29% at the top percentile. The contrast with the equivalent figure in Keuschnigg et al. (2023) (Figure A2) is again sharp at the top, as they find the relation to be at first flat above 90th percentile and the steeply negative above the 97th.[4]

[Figure 2 about here.]

Figure 3 shows the interquartile range for ability test scores within each earnings percentile. While the association between earnings and average ability is very strong, there remains much heterogeneity. For example, a man with a cognitive score in

[4]Figure 12 in the Appendix shows a stacked histogram of all stanine shares by earnings percentile.

the 75th percentile among the 10th percentile of earners would place slightly above the median among the cognitive scores of the 80th percentile of earners. Similarly, a 75th percentile cognitive score among the population of 80th percentile earners would amount to a median score among the 98th percentile of earners.

The interquartile range of ability gets increasingly narrow among the higher earnings percentiles. (The same is true for the standard error of the mean, as can be inferred from the confidence intervals in Figure 1; with over 3000 observations in each percentile this is not very surprising nor informative.) It spans about 1.5 SDs in the lower half of the earnings distribution, and progressively narrows down to 0.9 SDs at the top. The lower variability of ability is another way in which the ability-earnings relation gets stronger at the top of the earnings distribution. Again, the relationship is very similar for both cognitive and non-cognitive ability.

[Figure 3 about here.]

Figure 4 shows the average test score percentile plotted against earnings percentiles. Now the variables on both axis are distributed uniformly over $\{1, \ldots, 100\}$ in the population. This unit-free approach reveals the by now familiar shape between earnings and ability: it is flat until about the 40th earnings percentile; increasing and convex thereafter. If anything, the relationship seems to get steeper at the top earnings percentiles. For example, the average cognitive score at the top earnings percentile is in the 80th percentile of the FDF test data, which is 12 percentiles higher than the average cognitive percentile in the 90th earnings percentile. Going down a further 10 percentiles in earnings, there is on average just an 8 percentile difference in cognitive scores.

[Figure 4 about here.]

Figure 5 plots the average score from each of the three subtests by the earnings percentile. The same overall shape of the relationship is apparent in each subscore, being strongest with the arithmetic score.

[Figure 5 about here.]

3 The shape of the earnings-ability relation

The flipped axis approach answers the question of what can we infer about ability just based on observing a person's earnings in the labor market. In this section I flip back the axis to what is more conventional in economics. We would usually think that ability affects potential and actual earnings. Here the abilities are measured after secondary school and years before the later-life earnings, so direct reverse causality can be ruled out but unobserved factors can of course affect both.

The purpose of this section is to give a detailed non-parametric description of the earnings-ability relationship, using the Finnish administrative data on earnings and military conscript test scores. I divide the conscripts who are observed in the relevant age range into percentile bins by their anchored test scores, and then calculate average earnings within each bin. Figure 6 shows the average percentile of earnings at age 35–45 on the vertical axis.

[Figure 6 about here.]

The relation of earnings and ability is strong all along the distribution of test scores, with slight steepening near the top. The relation is almost linear for a wide swathe of both cognitive and non-cognitive abilities, from 15th to 90th percentiles, with a slight steepening at the top. Now the relation is monotonic also at the bottom, but there is a very steep earnings gradient among the lowest cognitive percentiles which does not show up for non-cognitive scores. The difference between very low and extremely low non-cognitive scores seems to be much less informative than a similar difference between cognitive scores.

The lack of a dip at the bottom seen in the flipped axes approach is consistent with my earlier "educated guess" about the source of the dip. There is no equivalent error (or selection process) that would turn middling test scores into zeros as there is for earnings. Rather, it causes a negative bias in the measured euro-valued earnings all along the distribution, probably more so near the top, but causes less mix-ups in terms of the average earnings percentile conditional on test score.

Figure 7 shows the same relation for euro-valued earnings. Despite being deflated to a common base year, here the use of monetary valued earnings will effectively weigh the later cohorts more due to growth in real earnings. The earnings percentile, calculated between men of the same birth year, avoids this problem. Nevertheless, the same overall shape of the earnings-ability relation is again apparent. Compared to the percentile-on-percentile plot the steepening at the top is more visible, which is natural as uniformly distributed percentiles are unaffected by the skewness of the underlying variable.

[Figure 7 about here.]

4 Data coverage

The FDF test data covers 80% of native-born men in the data. The reason for missing test scores cannot be inferred from this data, but the two most common reasons are individuals opting for civilian service instead of military service and exemptions due to pre-existing medical conditions. Some men are exempted from service due to a criminal record. There may also be gaps in the primary data archive of the FDF, e.g., lost microfilms affecting certain military units at certain test dates.

Figure 8 shows the relation of the coverage of the FDF test data with later-life earnings. The FDF data covers a roughly stable 85% of earners above the 30th percentile, below which the coverage drops steeply. Of the bottom decile of earners the FDF data covers only 69%.[5] This is likely due to a positive selection into military service vis-a-vis those exempted from service, which would rarely be an issue for men who later end up earning at median or above.

[Figure 8 about here.]

The Finnish wage data used here is moderately top-coded, with values in the top percentile replaced by the percentile median (sic) each year.[6] This could cause some attenuation in the estimated earnings-ability relation near the top. A similar top-coding applies separately to entrepreneurial income, but this makes less difference to total earnings.

Figure 9 shows the share of top-coded wage observations by the earnings percentile. For men whose earnings are in the top percentile in terms of earnings at age 35-45 almost 90% of their yearly wage observations are top-coded; below the 90th percentile (not shown in the figure) this share is everywhere less then 1%.

[Figure 9 about here.]

Figure 10 shows the share of top-coded wage observations by the percentile of cognitive scores. Among the top percentile of cognitive scores 11% of observed wage-years are affected by top-coding. This graph also serves as another description of the shape of the earnings-ability relation. The depicted variable is the share of calendar years at age 35-45 where the man's wages were in the top 1% of all wage earners (including men and women of all ages).

[Figure 10 about here.]

[5] We showed in Jokela et al. (2017) that there was not significant change in the selection into military service between the birth years in this data.

[6] Non-top-coded income data also exists at Statistics Finland, but is not part of the basic person-year panel "FOLK-Perus" and so was not immediately available for this project.

5 Shorter age range, more birth cohorts

The choice of the age range where the earnings are measured involves a trade-off. The longer the age range, the less noisy is the individual average. However, the longer age range also means that fewer birth cohorts can be included in the analysis as the later cohorts are not yet old enough to be observed in the longer age range. Using the age range 35–45 makes it easier to compare the results with Keuschnigg et al. (2023) and, as it turns out, does not make much difference to the shape of the earnings-ability relation.

The top panel in Figure 11 shows a version of Figure 1 where the age range for calculating the average yearly earnings is reduced to 35–40. This allows for all birth cohorts 1962–79 in the FDF test data to be used. The bottom panel differs only by also restricting the birth years to 1962-74, as had to be done in Figure 1. This shows that the small difference between the age 35–40 and 35–45 are mainly due to the added birth cohorts allowed by the shorter age range. I conclude that, for the purposes of this note, it is fine to stick with the age rage 35–45. The population level dataset is in any case large enough for the shape of the earnings-ability relation to be clear even with a smaller set of birth cohorts.

[Figure 11 about here.]

6 Conclusion and speculation

The relation between earnings and ability is robustly monotonic and increasing in population level administrative data of Finnish men. The relation is stronger near the top of the distribution, whether looking at highest earnings or highest ability scores. It is very similar for cognitive and non-cognitive ability.

Recent findings by Aghion, Akcigit, Hyytinen and Toivanen (2023) point to another "steeper at the top" relationship between a measure of later-life productivity and FDF test scores. They combine individual data on patent filings, and find that the share of inventors is monotonically increasing as a function the cognitive score; see especially Figure 6 of that paper. About 7% of men with the highest score in the visuospatial subtest are inventors, whereas below the median score the share is less than 1%.

The one clear anomaly to the positive relationship between earnings and ability is a dip in the average ability score around the lowest percentiles, when looking at this relationship from the "flipped axes" perspective. That is, there is a significant decrease in average ability scores between neighboring earnings percentiles at the bottom of the earnings distribution. I believe this is due to a specific type of "measurement error" whereby earners from all over the distribution can fall to near or at zero incomes, due to any type of incapacitation or incorrectly measured country of residence. Those whose observed earnings significantly differ from their ability-dependent earnings potential tend to clump close zero earnings. No other part of the earnings distribution is in the same way a common destination for mismeasured earnings potential.

The shape of the flipped axis relationship of ability conditional on earnings looks mostly very similar here to what was found in similar Swedish data by Keuschnigg et al. (2023), with one very notable difference. Whereas they find that the relationship flattens out at about the 90th percentile of earnings, I find that in Finnish data the relationship becomes if anything steeper among top earners. The reason for this difference remains a mystery at the time of this writing, but I will offer a couple of speculative directions for closer investigation.

The first possibility is some mundane difference in data between Finland and Sweden. For example, the set of birth cohorts in Keuschnigg et al. (2023) is partly different, and the Swedish cognitive test has four subtests instead of the three in Finland. The hard part is to find a difference in data that would mostly affect the top earnings decile, while at lower deciles the picture looks so remarkably similar.

One such candidate is the definition of what counts as earnings as opposed to capital income—this distinction can be problematic in the best of cases. It is conceivable that differences in such categorizations would make much more of a difference among the top earners. Those with significant entrepreneurial income may be able to transform much it into capital income or private asset appreciation; the Finnish tax system certainly results in incentives to do so. Incidentally, in this Finnish data, changing the definition of earnings so as to exclude entrepreneurial income made little difference to the results (not shown).

Another possibility is that there were differences in test-taking behavior between the countries that affected the eventual top decile of earners differently from those below. In any case it is probably true that men with low motivation produce a noisier test result. If noisy-test-taking behavior is positively correlated with eventual top decile earnings in Sweden but not in Finland then this could explain why the ability-earnings relation plateaus only in Sweden.

One fact that makes the difference in test-taking behavior a plausible candidate for country differences is that in Sweden the cognitive ability test was taken before the start of actual military service, so the data covers also men who end up not doing the military service (it covers 94% of native-born men according to Keuschnigg et al. (2023)). Yet, as seen in Figure 8, of the top decile of Finnish male earners over 85%

did serve in the military, and this propensity is about the same all along the top half of the earnings distribution. The top earners that are missing from the FDF data would have had to have an inordinate impact for their test results to flatten out the top of the ability-earnings curve, and again this would have to work out differently for the earners below them.

Finally, one possibility is that, despite their many similarities, the economic processes that determine the mapping from cognitive ability to top earnings really are that different in Finland and Sweden. However, I cannot come up with a mechanism what would cause it to be so.

Appendix

[Figure 12 about here.]

References

Aghion, Philippe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen (2023) "Parental Education and Invention: The Finnish Enigma," *Helsinki GSE Discussion Papers*, No. 3.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz (2011) "Personality psychology and economics," in Hanushek, Eric A, Stephen Machin, and Ludger Woessmann eds. *Handbook of the Economics of Education*, Vol. 4: Elsevier, pp. 1–181.

Borghans, Lex, Angela Lee Duckworth, James J Heckman, and Bas Ter Weel (2008) "The economics and psychology of personality traits," *Journal of human Resources*, Vol. 43, No. 4, pp. 972–1059.

Correia, Sergio (2017) "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator,"Technical report, Working Paper.

Cunha, Flavio, Eric Nielsen, and Benjamin Williams (2021) "The econometrics of early childhood human capital and investments," *Annual Review of Economics*, Vol. 13, pp. 487–513.

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*, Vol. 114, No. 25, pp. 6527–6532.

Keuschnigg, Marc, Arnout van de Rijt, and Thijs Bol (2023) "The plateauing of cognitive ability among top earners," *European Sociological Review*, p. jcac076.

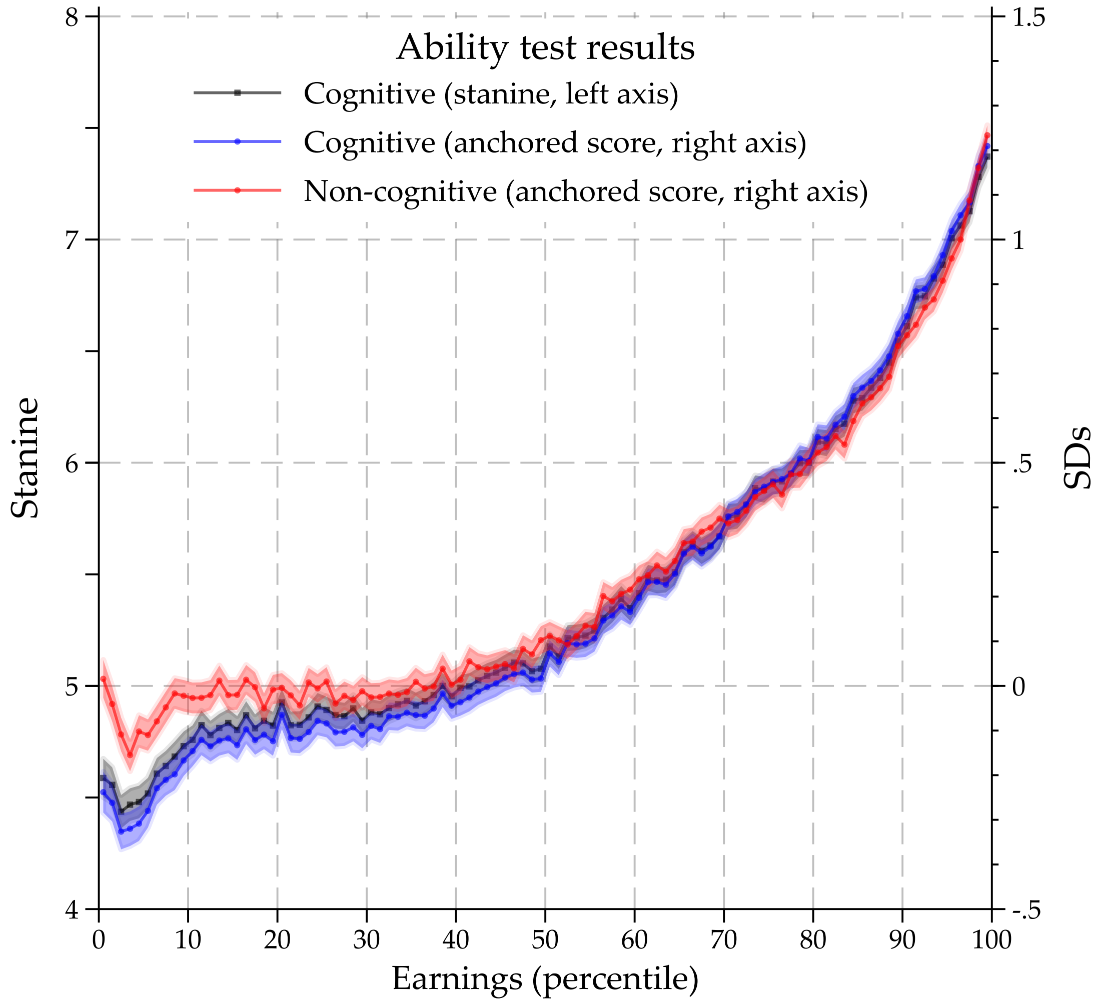

Figure 1: Cognitive and non-cognitive test results plotted against the percentile of average yearly earnings at age 35–45. The stanine ("p-luku") is an integer summary measure of the cognitive test calculated by the FDF. An anchored score is a weighted average of the subscores of a test, here depicted in base-year standard deviations. Shaded bands depict the 95% confidence intervals for the within-percentile means.

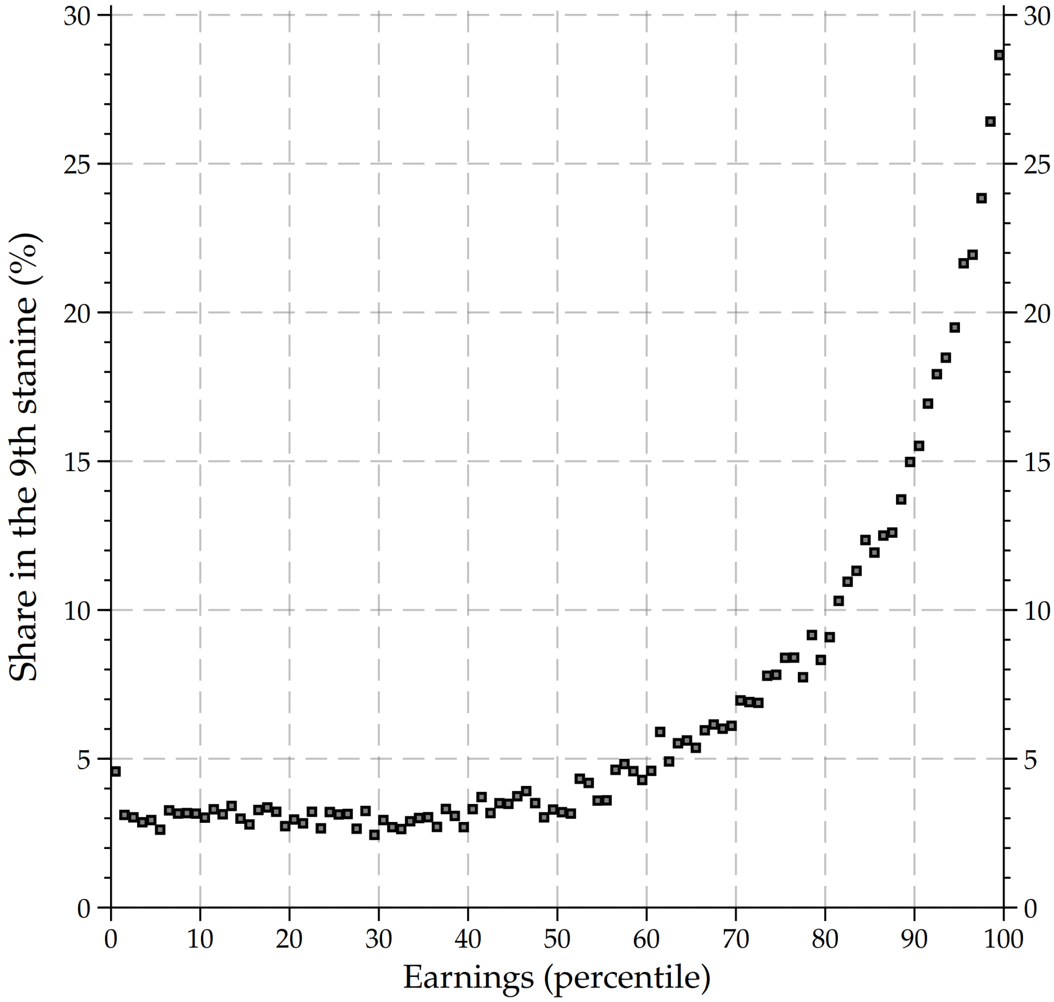

Figure 2: The share of conscripts in the highest stanine of cognitive ability test results by the earnings percentile at age 35–45.

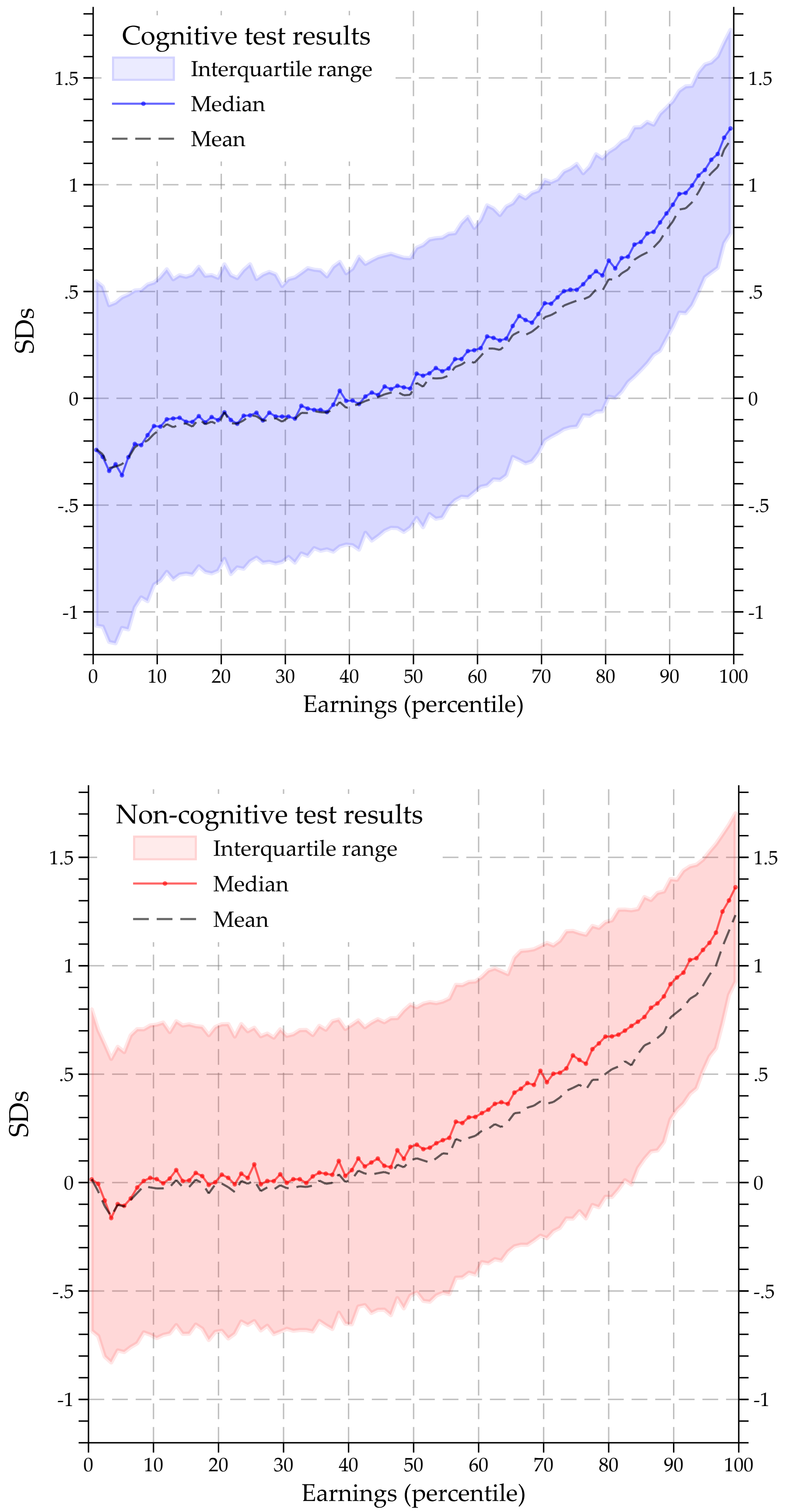

Figure 3: Interquartile range of cognitive and non-cognitive ability scores by the earnings percentile at age 35–45.

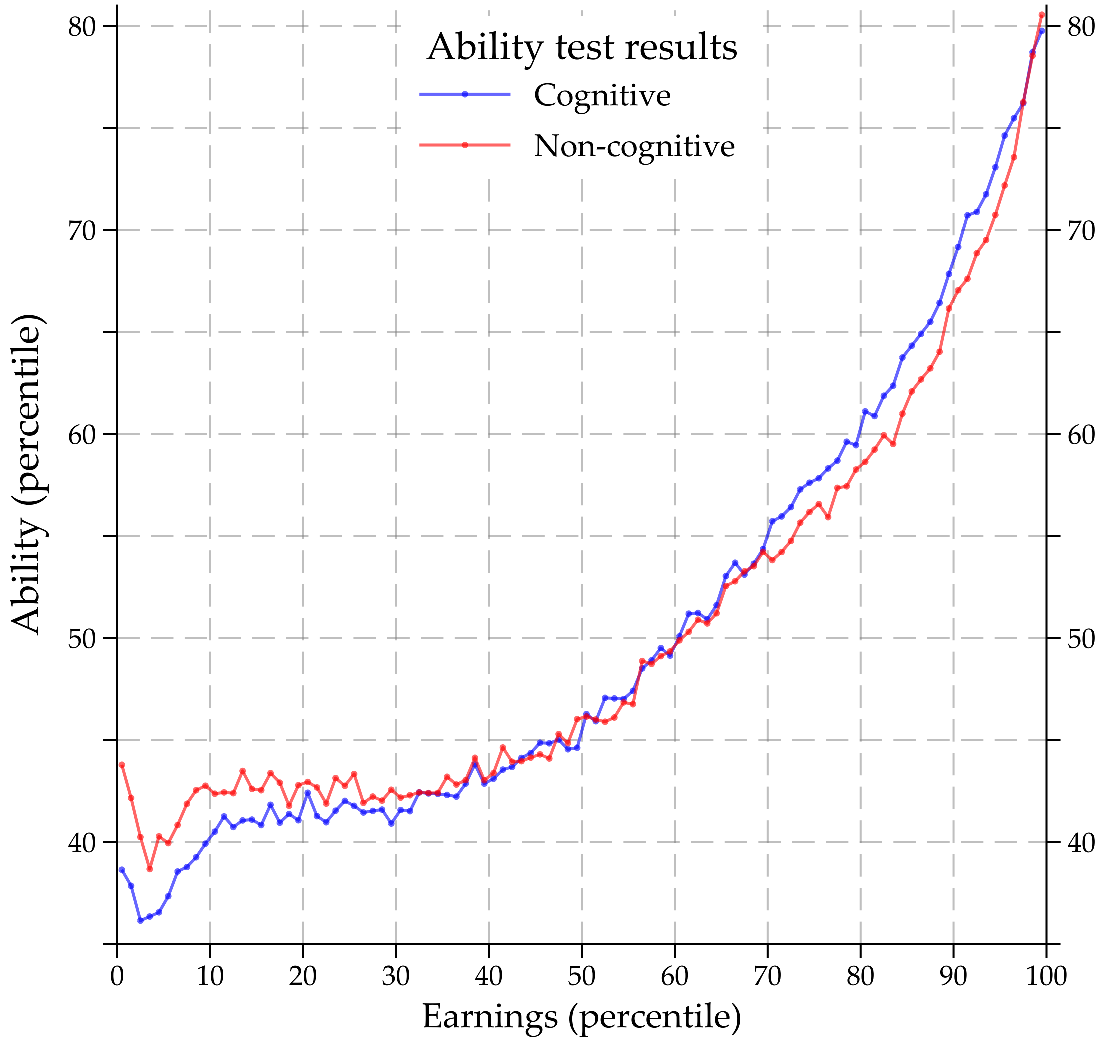

Figure 4: Average percentile in cognitive and non-cognitive tests by earnings percentile at age 35–45.

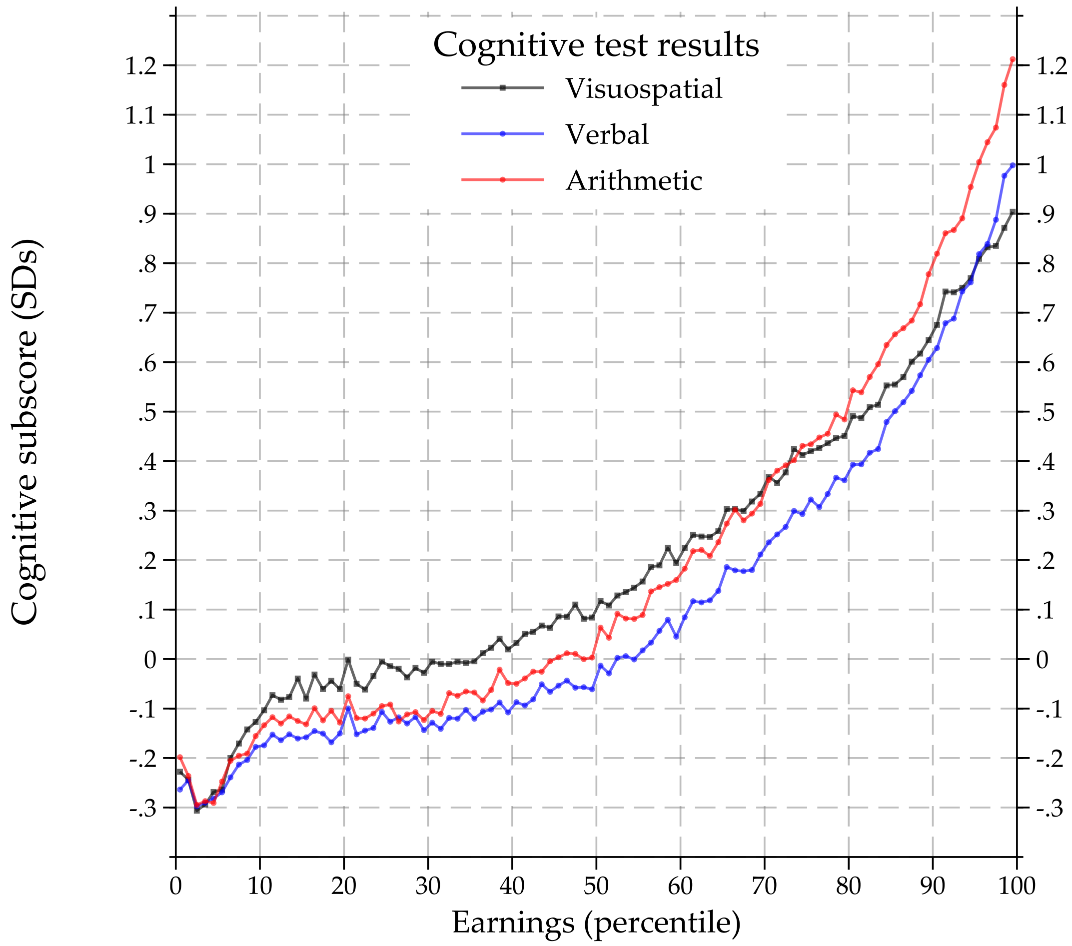

Figure 5: Average subscores of the cognitive ability test by the earnings percentile at age 35–45. The subscores are standardized in base cohort SDs. The visuospatial subtest measures similar abilities as typical IQ tests.

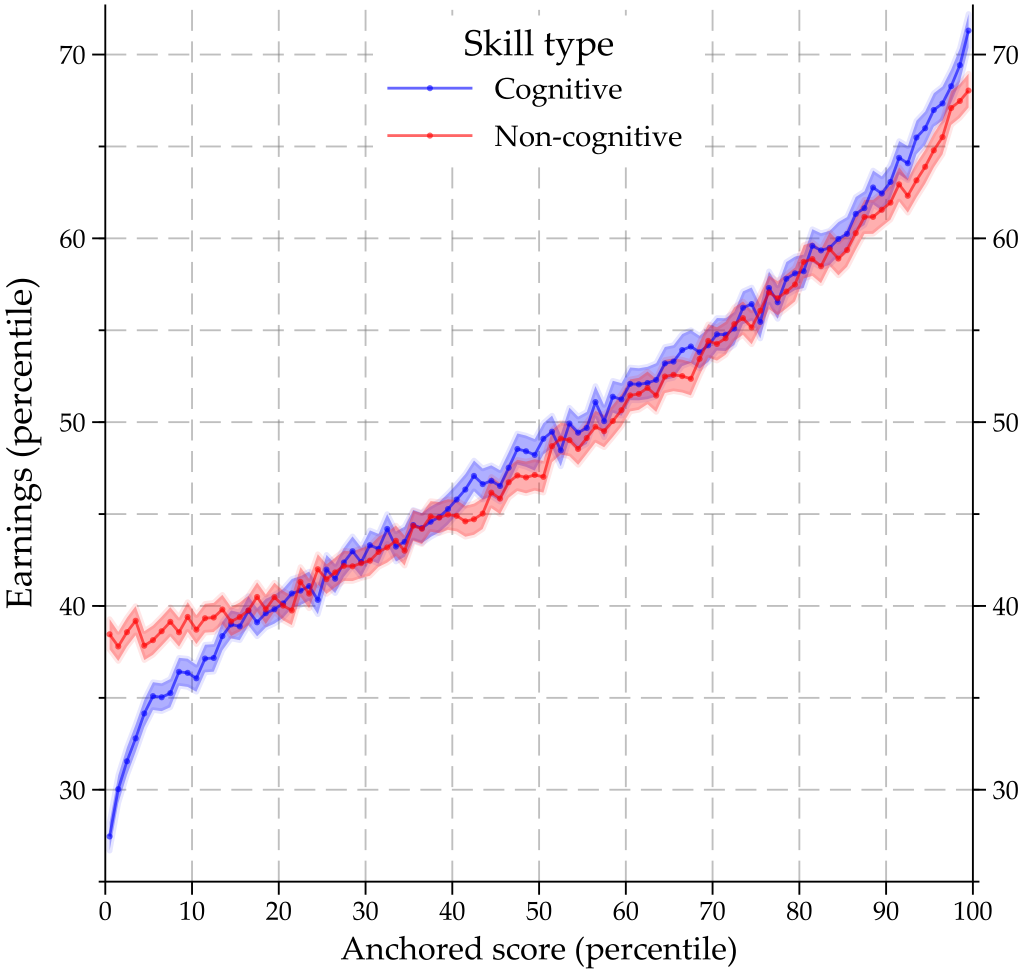

Figure 6: Earnings percentile (at age 35–45) by test score percentiles. Shaded bands depict the 95% confidence intervals for the within-percentile means.

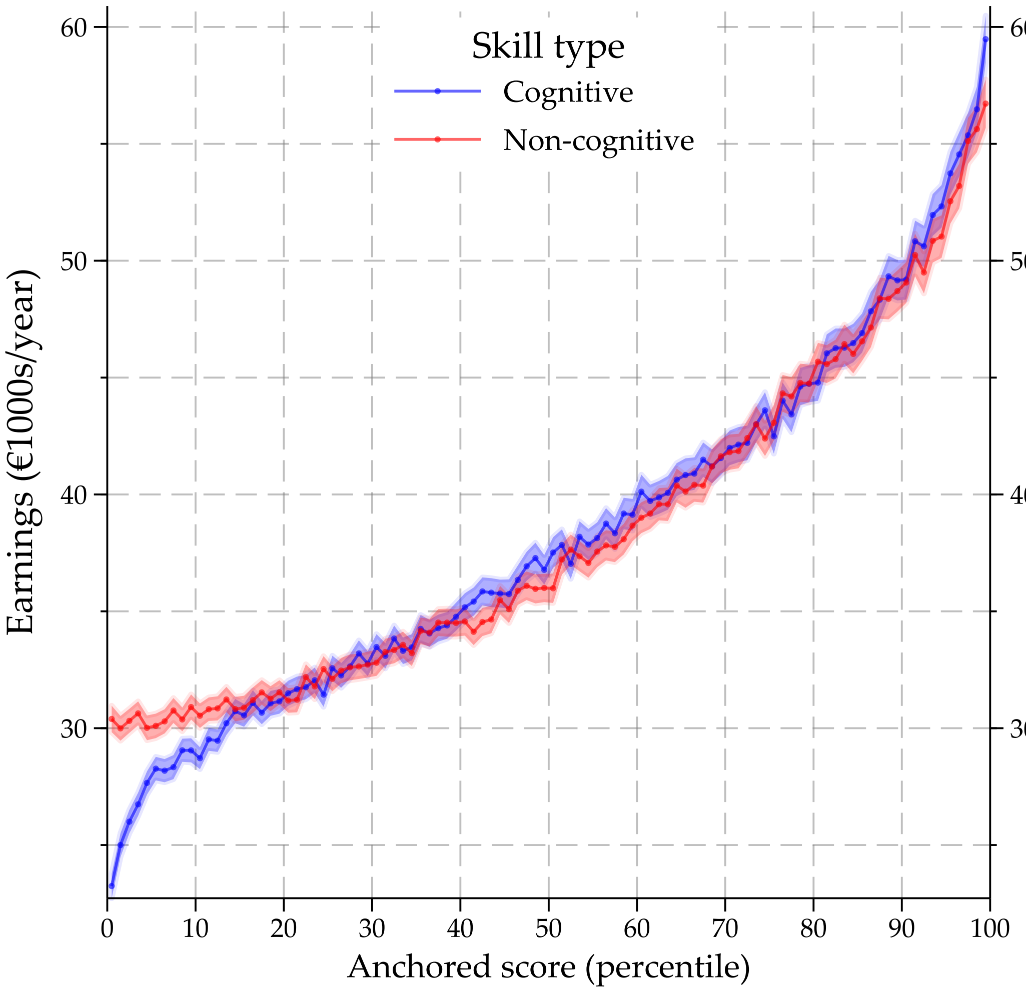

Figure 7: Average yearly earnings (at age 35–45) by test score percentile. Shaded bands depict the 95% confidence intervals for the within-percentile means. Earnings deflated by CPI to 2014 euros.

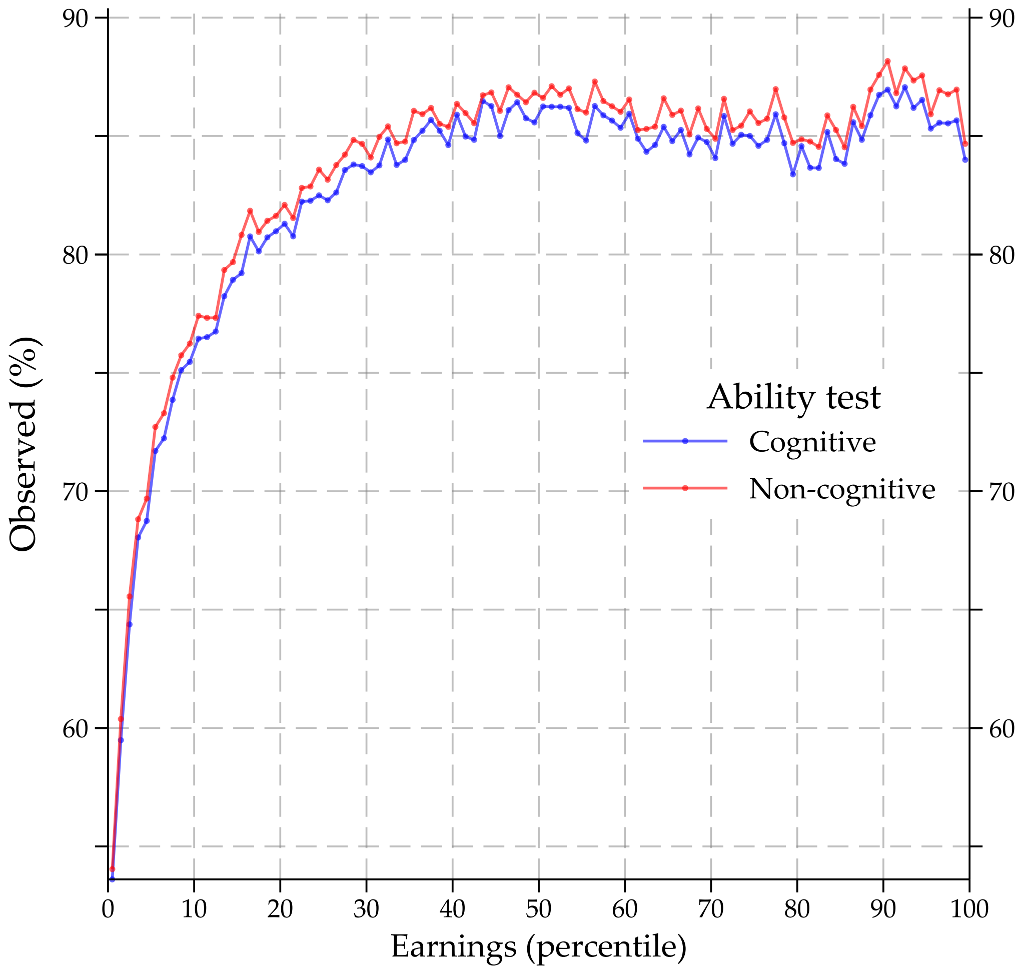

Figure 8: FDF data coverage by earnings percentile at age 35–45.

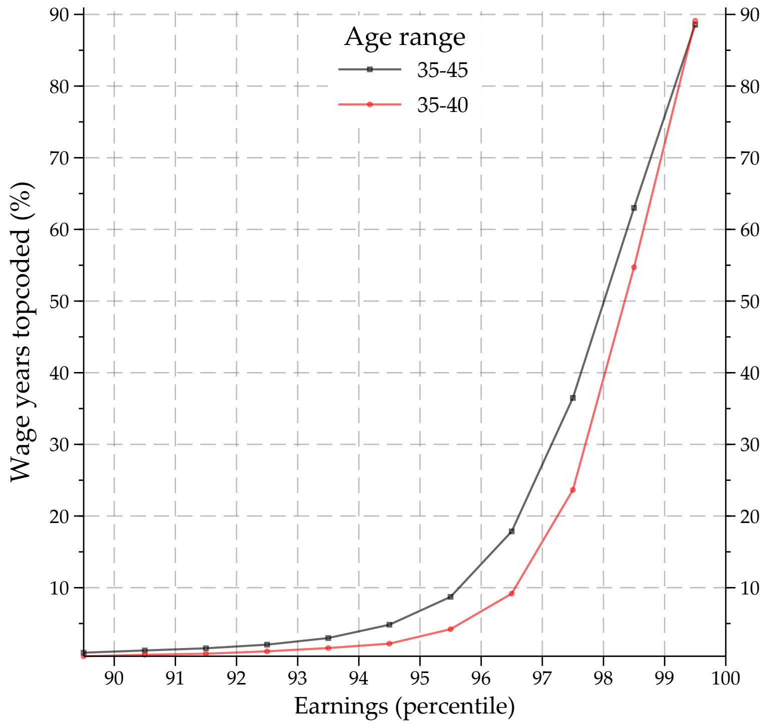

Figure 9: The average share of wage-years at a given age range affected by the top-coding of wages, by earnings percentile.

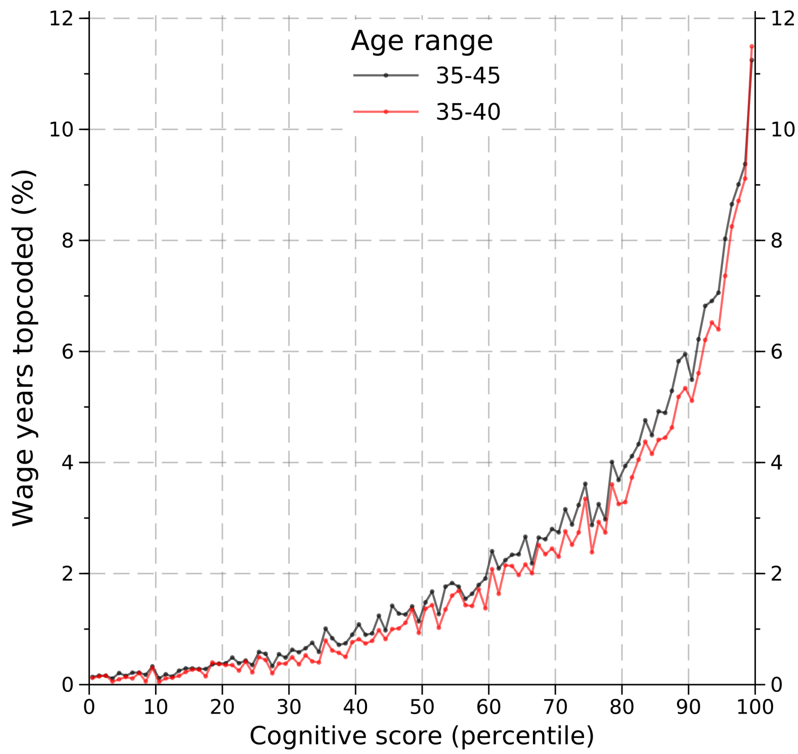

Figure 10: The average share of wage-years at a given age range affected by the top-coding of wages, by the percentile of cognitive test scores.

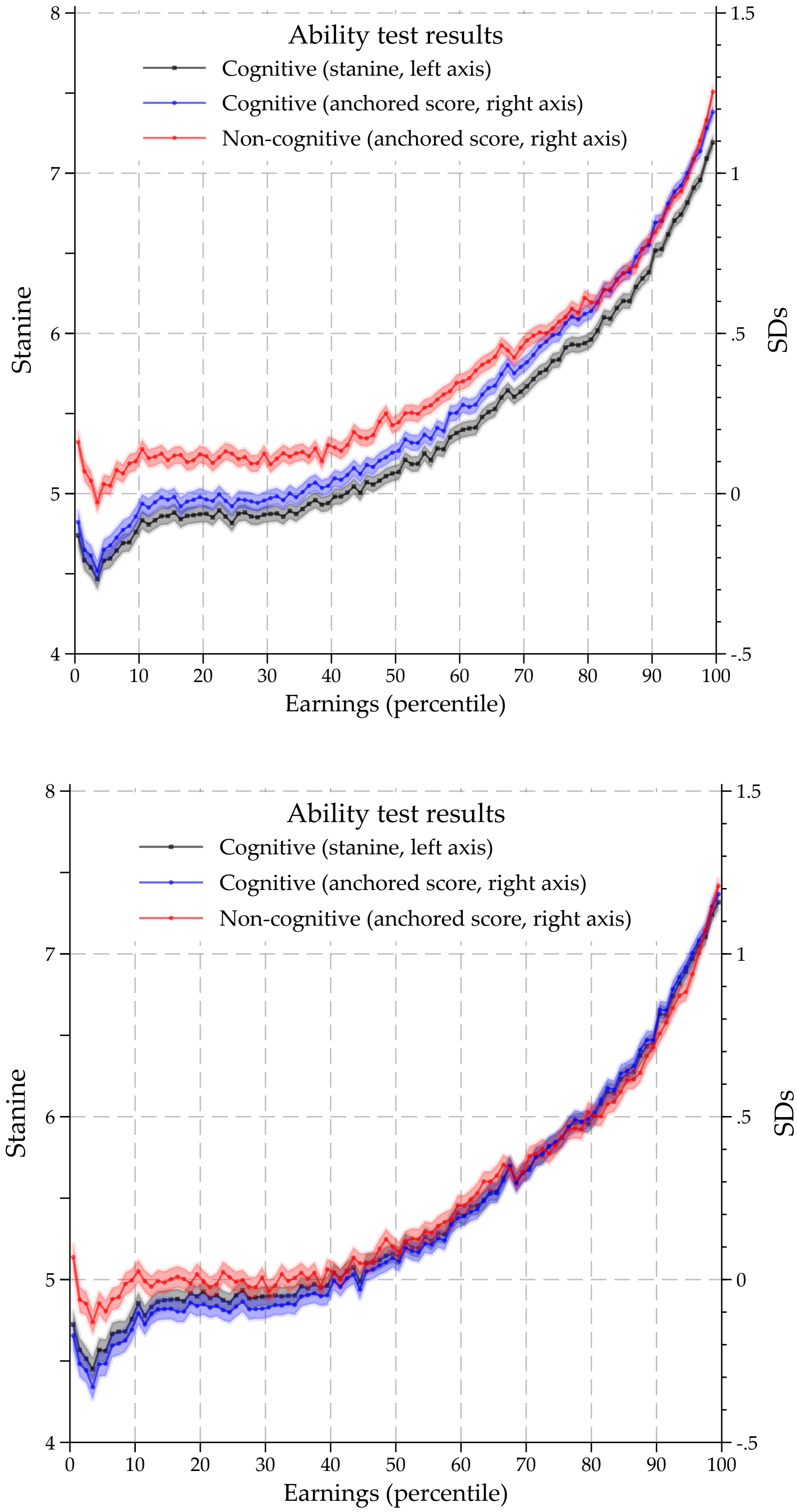

Figure 11: The top graph replicates Figure 1 for a shorter age range, 35–40. This allows for all birth cohorts 1962-79 to be included in the analysis, as the 1979 birth cohort turns 40 in the last year of the earnings data (2019). The bottom graph shows the results for the same 35–40 age range but while restricting the birth years to 1962-74 as in Figure 1.

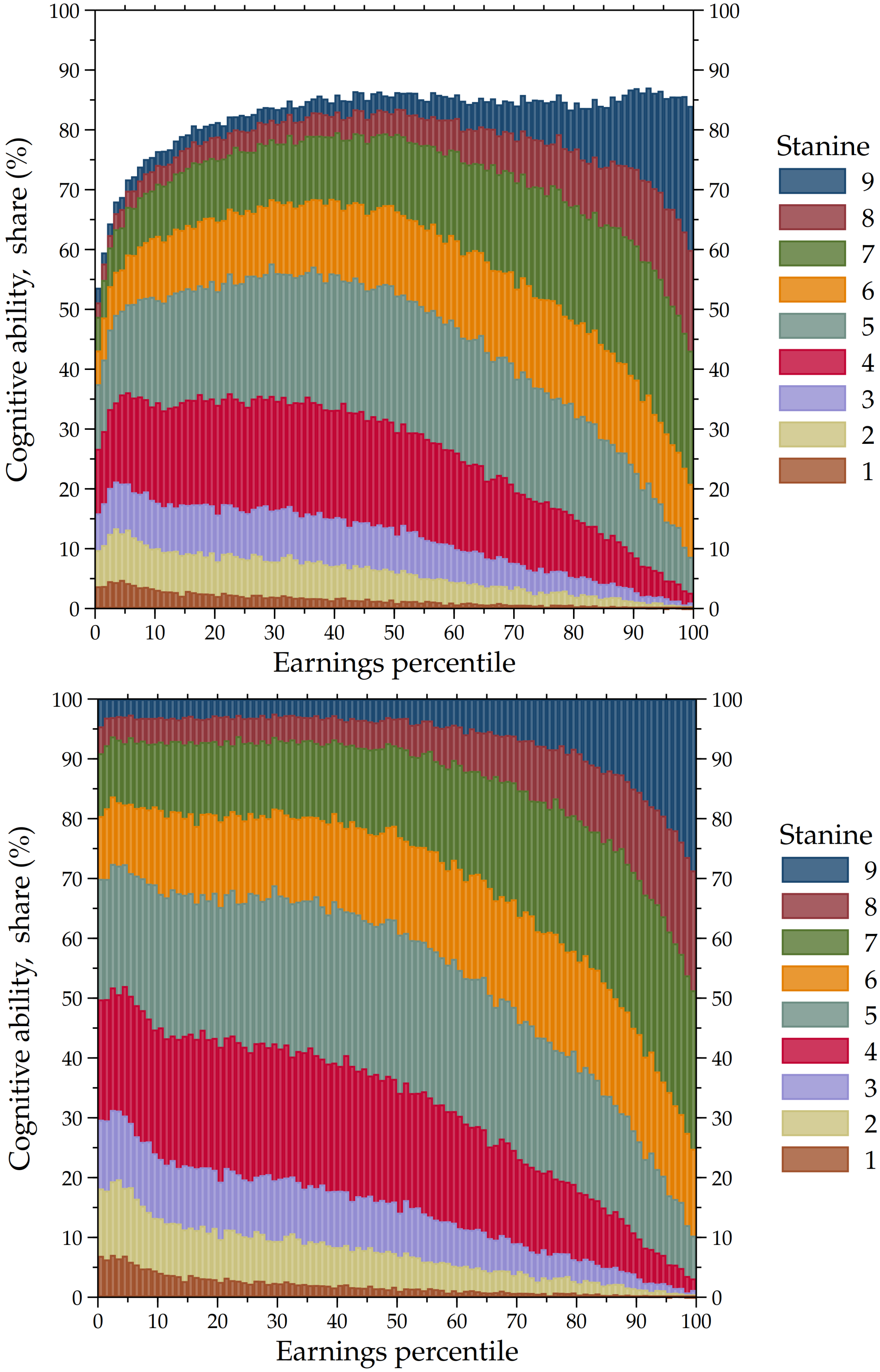

Figure 12: Distribution of cognitive ability test results in terms of FDF stanines ("p-luku") by earnings percentile (for average yearly earnings at age 35–45, birth years 1962–74. Top panel: share of men in earnings data, $N = 415k$. Bottom panel: share of men in both earnings data and in FDF data, $N = 347k$.