Reading Skills and Earnings: Why Do Doing Words Good Hurt You're Wages?

Carl Sanders*

December 15, 2014

Abstract

In a reading comprehension test administered to American youth in 1980 by the National Longitudinal Survey, a one standard deviation increase in reading scores is associated with a 1.5% decrease in later wages after conditioning on the youths' other scores on math, science, and personality tests. This goal of this paper is to explain this negative conditional relationship between the reading test score and wages, which I call the *reading penalty*. In the first section of the paper, I consider and reject a number of statistical objections to the existence of the reading penalty. In the second section, I construct a simple generalized Roy model that offers two distinct economic explanations for the reading penalty. The first explanation is the *starving artist,* where reading skills are a proxy for preferences for low-wage jobs, while the second is the *bad sign*, where high reading test scores proxy for a lack of other productive attributes. In the final section, I use the NLSY data and the identification from the model to determine which economic explanation of the reading penalty is most plausible. I find weak support for the starving artist hypothesis. On the other hand, I find evidence in favor of the bad sign explanation: higher reading comprehension scores may signal a lack of organizational skills.

JEL Codes: J24, I21, J31

*Department of Economics, Washington University in St. Louis. carlsanders@wustl.edu
Thanks to David Rivers, Lance Lochner, Ronni Pavan, Shintaro Yamaguchi, and seminar participants at McMaster University and the University of Western Ontario for helpful comments and suggestions.

1 Introduction

Reading comprehension is a skill. More precisely, the ability to quickly and correctly interpret written language is universally considered a key productive capability of workers. Anyone with even a basic knowledge of marginal productivity theory should then be able to correctly predict the correlation of reading test scores with wages: higher reading test scores are associated with higher wages. In the data used in this paper and explained in detail later, in 1980 a sample of American youth between the ages of 16 and 23 took a reading comprehension test. Students scoring one standard deviation above the mean had average real wages $13 \pm 0.5\%$ higher over their careers. The test score itself is not causing higher wages, but the positive correlation between the score and earnings has a simple economic interpretation: workers score higher on reading tests because they are better at reading, workers who are better at reading are more productive, and more productive workers get paid more.

There is a problem with this simple story linking test scores to productivity, however. The reading tests administered to the youth were a part of a battery of tests which were intended to measure a variety of skills and knowledge: e.g. math, vocabulary, and even non-academic subjects such as auto shop knowledge and self-esteem. All of these individual scores are highly positively correlated with each other. Students who did well on one test tend to be generally smart, and so tended to do well on all the other tests. But conditional on all the other test scores, the relationship between reading skills and wages reverses: a one standard deviation increase in reading test scores is associated with a $1.5 \pm 0.7\%$ *decrease* in wages over their careers.

The goal of this paper is to understand this negative conditional correlation of reading comprehension test scores with wages, which I call the *reading penalty*. In the first part of the paper, I consider and reject a "laundry list" of potential reasons for the reading penalty that have little economic content, such as potential problems with the tests themselves, functional form assumptions in the wage regression, or differences in socio-economic and educational background across test-takers. In the second part of the paper, I consider two different economic explanations for the reading penalty. The first explanation I call the *starving artist* story. Simply put, workers with good reading skills may prefer working as a barista and writing a novel in their spare time to working as an accountant. If reading test scores are a proxy for individual preferences for the work environments of low-wage jobs and those preferences are later reflected in sorting in the labor market, higher reading test scores can be associated with lower wages later even if it is also measuring a productive skill. The second explanation I call the *bad sign* explanation. An as example, say reading scores are a measure of ability to read but also proxy for introversion, which is not tested for otherwise. Then if the returns to reading are sufficiently low relative to returns to people skills, higher reading test scores can be associated with lower wages. More generally, if reading test scores are sufficiently negatively correlated with an important skill that is not measured by the other tests, later labor market outcomes can be worse.

Whether the reading penalty is due to preferences or productivity has implications for our understanding of what skills matter in the labor market and why. If preferences for low-wage jobs turn out to be the cause for the reading penalty, the tight connection between skills and outcomes that labor economists have typically assumed may need to be re-evaluated. On the other hand, if the reading penalty is proxying for a lack of some other useful skill, it is not clear that we know what we have been measuring through standardized tests, since just as easily math scores could be proxying for an abundance of that same unknown skill.

While the reading penalty has been noticed in passing a few time in the labor literature, it is so counter-intuitive that each time it is typically either ignored or dismissed as a statistical artifact. Altonji (1995) documented that the returns to social studies and English classes in high school were negative under a number of specifications, although the effects were sometimes statistically insignificant Using samples of college attendees, Taber (2001), Kinsler and Pavan (ming), and Seki (2013) all found that conditional on math SAT score, the wage coefficient on verbal SAT is negative and statistically significant. More generally, there is basically no evidence that reading skills (above basic literacy) have positive effects on wages and a significant amount of evidence that the effects may be negative; see for example Hastings et al. (2013) who document that the causal return to a journalism major is negative.

In the first section of the paper, I consider and reject a plethora of statistical objections to the existence of the reading penalty. Some objections are simple to consider: for example, concerns over the data mining of this single negative coefficient are answered by noting that the negative relationship between reading skills and wages holds in a number of data sets. Similarly, concerns about the construction of the test itself are inconsistent with huge amount of previous research that has verified the validity of the ASVAB. The primary objection that requires addressing is functional form considerations, but after estimating a number of specifications of the baseline wage regression I conclude that the reading penalty does exist and requires an economic explanation.

In the second section, I create a simple generalized Roy model where workers with two skills, reading skill and an "unknown" skill, choose to either be an accountant or a barista.

Workers have preferences over both wages and the job itself. The model can generate a negative relationship between wages and reading skills in exactly two different ways that correspond to the starving artist and bad sign explanations discussed above. Analyzing the model allows me to show how observables can be used to distinguish the two proposed explanations of the reading penalty. Under the starving artist story, we would expect the negative relationship between reading test scores and wages to move closer to 0 as we added additional (endogenous) controls for sectoral choice, and if workers were identical within sectors the reading penalty would fall to 0. To identify the bad sign explanation, I show that in sectors where the unknown skill is not used, the reading penalty should be 0, and in sectors where they skill is most required the reading penalty would be largest. Combining data with skill usage in occupations with the wage and test score data will then allow me to determine what the unknown skill actually is.

In the third and final section, I use the identification results from the model and data on later choices and outcomes of the youth to determine which of the starving artist and bad sign explanations of the reading penalty is most plausible. I find little support for the starving artist story. Instead, the reading penalty only marginally decreases after including controls for the worker's chosen sectors no matter how I define sectors, whether by one-digit or three-digit occupational codes or measures of the tasks performed in occupations. On the other hand, the bad sign explanation finds support in the data. The results suggest that workers in jobs where they perform high levels of organizational tasks face the largest reading penalty, while those in intensely interpersonal or creative jobs see a significantly lower wage reading penalty. I take this as preliminary evidence that reading skills are negatively correlated with organizational skills and that this drives the wage penalty.

My results imply that measurement of skills is currently a larger problem than labor economists believed. Factor-analytic methods typically choose a low number of skills based off the simple observation that most measures of skills are incredibly highly correlated. The reading penalty and the productivity explanation that best fits the data implies that residual skills can be important drivers of labor market outcomes even after factoring out the joint correlation , and those additional skills need to be measured directly to recover consistent estimates of which skills actually matter.

2 The ASVAB and the NLSY79

The Armed Services Vocational Aptitude Battery is a series of tests for potential military entrants and used by all US Armed Forces Services since 1968. The goal of the ASVAB is twofold: first, to determine whether an applicant has sufficient skills, knowledge, and abilities to begin military training, and second, to facilitate assignment of individuals into different services and jobs within those services. The more well-known AFQT score is simply a function of scores on four of the of the ASVAB subtests. The AFQT score is meant to measure whether an applicant is potentially trainable, but the whole ASVAB battery is primarily intended to assign applicants to different tasks. The AFQT has been widely used in the labor literature as a measure of unobserved ability. Labor economists have focused on the AFQT as a single-dimensional measure of ability, but there is only around a 10% failure rate on the ASVAB, so the assignment function is considered the primary function of the ASVAB.

The ASVAB[1] consists of a battery of 10 timed subtests:

	Title	Comments	Time (Minutes)
1	General Science		11
2	Arithmetic Reasoning*	Word Problems	36
3	Word Knowledge*	Vocabulary, Analogies	11
4	Paragraph Comprehension*	Reading Comprehension	13
5	Numerical Operations	Speed of Arithmetic	3
6	Coding Speed		7
7	Auto & Shop Information	Cars, Tools	11
8	Mathematics Knowledge*	Geometry, Trigonometry	24
9	Mechanical Comprehension	Basic Machines: Gears, Pulleys, etc.	19
10	Electronics Information	Electric Circuits	9

The tests marked with an * are components of the AFQT score.

In 1980, the ASVAB was administered to the representative sample of American youth who were a part of the National Longitudinal Survey of Youth 1979. The conditions while taking the ASVAB were the same as for potential military applicants. The participants were told they would take a test that 1) can help them determine their optimal career and 2) could help them enter the military if they so desired. The Armed Forces agreed that the NLSY individuals could "save" their AFQT score and ASVAB score for up to 5 years after taking

[1]Unless otherwise mentioned, all references to the ASVAB are to the 1980 version (Form 8A) that was administered to the NLSY79 cohorts.

the exam if they wanted to enter the military. While there was no direct incentive for different performance levels, all test-takers were paid $50 for their participation.

The reading comprehension test I discussed in the introduction is Section 4, Paragraph Comprehension. Test-takers were given a set of questions that consisted of short reading passages followed by three or four multiple-choice questions about the content of the passage, the author's arguments, and the tone of the article. Some question were of sufficient generality that reading the actual passage would not be required to answer, while others required more subtle understand of the author's intentions. While the actual questions answered in this version of the ASVAB are property of the U.S. Department of Defense and unavailable, an example Paragraph Comprehension question is shown in Figure 1.

The primary sample of the NLSY79 I use consists of white males where were never in the military, and I do not drop the non-representative oversample the NLSY includes. Within that subgroup, 94% of them took the ASVAB. Later in the results section I will compare results for white males not in the military to other races, women, and white males who enter the military. Notably, I do not restrict the sample on education. The only control variable used in the baseline regressions that is not predetermined at time of the test is conditioning on the worker never entering the military, which is endogenous. Sample statistics of ASVAB scores in the NLSY are shown in Table 1, and other general summary statistics of the NLSY sample are shown in Table 2.

Dropping men who ever enter the military for the baseline results is standard in labor economics, but here it is not innocuous. In particular, applicants planning to go into the military have an incentive to perform well on the ASVAB compared to those who will not, since if they get a sufficiently high AFQT score they can use it to enter the military within the next 5 years. Additionally, the ASVAB has an actual direct effect on the careers of those who enter the military, because once the individual enters the military the scores on individual subtests are used for assignment. For example, in the Marine Corps a higher score on Paragraph Comprehension decreases the relative odds of being assigned to combat versus a technical occupation. Different assignment within the military will then have a direct effect on skill acquisition and later outcomes post-military. There is no analogous effect of ASVAB scores for non-military individuals.

Results from the baseline regression of wages onto the ASVAB individual subtests and a quadratic in the current age of the worker is shown in Table 3, Model 2. The regression is the pooled cross-section at the yearly level, so every individual is potentially in the sample up to 25 times. Additionally, wage observations are restricted to non-outlier wages and full time work (35+ hours per week). Given the number of repeated observations of individuals, clustering standard errors at the individual level increases the standard errors dramatically. I return to this later in the functional form part of Section 3. The results from the baseline regression demonstrate the basic idea behind the reading penalty: conditional on the other test scores and current age, one standard deviation in reading scores is associated with a 1.5% decrease in real wages. In the next section, I deal with a series of potential objections to the empirical finding of the reading penalty.

3 A Laundry List of Potential Objections to the Reading Penalty

In this section I deal with a series of potential problems that have little economic content, but could be the true cause of the estimated reading penalty. The first issue is that the test itself may have errors or problems with its administration that could have caused problems with reading comprehension scores. Second, at time of test administration the sampled youth already had a variety of different education experiences, which could be affecting the relative ranking of youth even with the same "true" ability. Third, I address concerns that perhaps those with higher reading scores have careers that begin at lower wages than others, but see faster growth. The fourth issue I consider is a concern that this finding is simply data mining, particularly since it is about the sign of a single coefficient in a regression. The fifth concern, which is the most serious given the small cross-sectional size of the sample, is that functional form assumptions are driving the result. I find that all five of these issues cannot explain any significant part of the reading penalty, which leads me to consider explanations more grounded in economic theory.

Problems with the ASVAB

The ASVAB tests themselves may lead to some spurious results if there are coding errors or other problems with the test itself. Given the ASVAB is used for both admission to and assignment into the US Armed Forces, it is the subject of probably the most extensive testing and analysis of any test in the world. The particular version of the ASVAB given to the NLSY79 cohort (Form 8A) was exhaustively documented and subject to a number of validity experiments and studies. There are hundreds of pages of technical documentation for the ASVAB, and potential problems with the Paragraph Comprehension section of the ASVAB are never mentioned. See Welsh et al. (1990), who review data from 176 distinct studies confirming the ASVAB's test validity under a number of different criterion.

There were two minor issues with the administration of the ASVAB to the NLSY79 cohorts. First, the initial scoring of the exams turned out to be inconsistent with the actual scoresheets, but this was corrected in later releases of the NLSY79. The current scores in the data come directly off the original scoresheets with an answer key constructed by the NLS using the original test. The second issue has to do with the final question in the Paragraph Comprehension section, PC#15. The question was poorly posed (it was an unclear "all of the above" type of question) and it was on a page by itself, which may have led to some test-takers missing it. However, all of my results are robust to dropping PC#15.

Educational Background

One significant problem with the administration of the ASVAB in the NLSY79 is that as the test-takers were between ages 16-23 in 1979, when they took in the test in the summer of 1980 there were individuals ranging from 10th grade to college graduates all taking the exact same test. While there is no obvious story why this would lead to the observed reading penalty, the interpretation of the reading penalty would change dramatically if it were only something that showed up among older or younger test-takers or only among those in college. Table 4 shows the baseline regression run 4 times, each on a different NLSY79 subsample. Subsample 1 consists of individuals who were not enrolled in school in May 1980 and had never finished high school; this would be high school dropouts and youth taking some sort of other break from schooling. Subsample 2 consists of youth who were currently enrolled in high school in May 1980. Subsample 3 is those who were enrolled in college in May 1980, and Subsample 4 is those who were not enrolled in school in May 1980 but had a high school degree.

The results from the subsample regressions show that for high school dropout types, there is no observed reading penalty, but for all the other groups there is, with the largest penalty seen among college students. . In general, this shows that the educational background of the students does not eliminate the reading penalty. In all later regressions I include controls for all preexisting characteristics of test-takers in 1980: their year of birth, their educational level at the time, and they current enrollment status.

Wage Levels/Growth Tradeoff

One potential issue may be that higher reading scores have potentially a net 0 (or positive) effect by having some negative effects early in the life cycle but positive effects later on, which would lead to a net 0 or positive effect on the present discounted value of wages. I deal with the concern about different career paths in two ways. First, I interact all the ASVAB scores with individual age. For the career concern explanation to make sense, there should be a levels/growth tradeoff for individuals with higher reading scores. This is not the case: the negative coefficient on reading score does not change significantly depending on the age of the individual. There is a statistically insignificant fall in the reading penalty as workers age, but even by age 60 the model estimates the effect is still negative. Second, and perhaps more simply, I repeat the baseline regression with only data for year 2005 and after. If individuals with higher reading scores started with lower wages (or more time out of the labor market in school), by 20 years later they should be earning more in order to compensate for their earlier career. Again, there are no changes in the results. B While the reading penalty has statistical significance due to the reduced sample size, the negative coefficient is almost identical to the full sample version of the baseline wage regression. There does not seem to be any positive long-run effect of reading scores on wages. For the sake of space I do not report the full results in the Tables; the results are available on request.

Data Mining

If I simply regressed wages onto thousands of variables that "should" have a positive coefficient, many of them would have a negative estimated coefficient. This may lead to questions about whether the reading penalty is simply a statistical artifact. Additionally, this is an impossible question to answer within any one data set.

There are a number of reasons to believe that data mining is not the cause of the observed reading penalty, however. Primarily, the exact same phenomenon is observed in the sister data set of the NLSY79, the National Longitudinal Survey of Youth 1997. The results from the NLSY97 regression are available on request, but simply show the same pattern with effectively the same magnitude. The samples are completely independent, and as in this case there are no other tests that have any significant negative relationship with wages. Additionally, if for some reason the NLS surveys were all considered suspect, as mentioned in the introduction there are a number of papers that have noted an analogous phenomenon using SAT takers from the Baccalaureate and Beyond data set. The primary difference in this paper is that the ASVAB was administered to the sample unconditionally on education, whereas the SAT is of course a selected sample of youth who are more likely to go to college.

Functional Form / Full Support Condition

This is the most serious issue with interpreting the reading penalty. To have the perfect non-parametric estimate of the relationship between getting an additional question right on the Paragraph Comprehension section given scores on all the other subtests, I would need to compare two individuals who scored the same on every subtest but different on PC by one question. Given there are 10 total sections and an average of 20-25 questions per section, no such pair of people exist. The linear regression in the baseline case imposes the strict functional form assumption of linearity and additivity to get around this problem.

The linearity assumptions is the easiest one to relax. Figures 2 and 3 relax linearity in an unconditional and conditional way, respectively. Figure 2 simply plots mean log wages net of non-ASVAB score observables as a function of each possible number of questions right on the PC section. Clearly, higher reading scores are associated with higher wages. Then Figure 3 plots the results from the same regression but also netting out the effects of the other ASVAB scores. While I still retain additivity, this version of the conditional model no longer puts any requirements on the relationship between predicted wages from scoring a 3 on the Paragraph Comprehension test (a very bad score) and a 15 (a perfect score). The result seems to clearly show that the positive relationship between reading test scores and wages reverses over the majority of the support when other test scores are controlled for. This exercise still imposes additivity, but suggests that odd shapes are not the cause of the existence of the reading penalty.

To relax additivity assumptions, I use alternative estimation methods that use local rather than global information to recover the conditional mean of wages as a function of test scores. To do this, I first construct aggregated measures of each individual's score on the non-Paragraph Comprehension test. I use two different weighting schemes: the first is a simple sum of all the other scores, while the second uses relative weights given by the coefficients of each test in the baseline wage regression. Then I separate these scores into 15 different dummy variables, one for each 1/15th of the score distribution. I run the wage regression allowing for a different slope for PC score for each segment of the Other Tests distribution. This has a simple interpretation: take everyone who scored between (say) 55-62% on average on the other tests, and run the baseline regression for them, and repeat this for every other decile. The results (available on request) are that the average reading penalty across the sample stays effectively the same, but there are some differential effects depending on the level of the individual's other test scores. The reading penalty is largest for those who score highly on all the other exams, but it is also largest for those who score the lowest on Paragraph Comprehension. This counter-intuitive finding has some interpretation: individuals who are very smart on math and science but mediocre at reading comprehension earn less than those with similar math and science scores but who perform horribly at reading comprehension.

4 How Can A Standard Labor Model Explain the Reading Penalty?

The reading penalty can actually arise naturally from a simple generalized Roy (1951) model where workers care about both their wages and some utility that may differ across sectors.

Workers draw $(r_i, u_i) \sim F_{r,u}$. There are two sectors, A and B, and workers choose $S \in \{A, B\}$. The utility of the two sectors for a worker of type (r, u) is given by

$$U(r,u,B) = w(r,u,B),$$

$$U(r,u,A) = w(r,u,A) + k_A.$$

The function w gives the wages in each sector, while k_A is the utility bonus for being in sector A.

Workers choose the higher utility sector,

$$S^\star(r,u) = \begin{cases} A & \text{if } w(r,u,B) - w(r,u,A) < k_A \\ B & \text{if } w(r,u,B) - w(r,u,A) \geq k_A \end{cases}.$$

For simplicity, I assume that for every r there is a cutoff value of u that sorts all workers with a higher level of u into sector B and a lower level into sector A. This is an easily generalizable assumption but makes the algebra and interpretation much simpler. The **cutoff function** $u^{\star}(r)$ gives the level where a worker of type r is indifferent between sectors and is defined by

$$\forall r, \exists u^{\star}(r) \ s.t. \ w(r,u^{\star},B) = w(r,u^{\star},A) + k_A.$$

Expected wages of a worker with reading skills r is then

$$E[w|r] = \int_{-\infty}^{\infty} w(r,v,S^{\star}(r,v)) f_{u|r}(v)\, dv.$$

Dividing the integral into sector A below $u^{\star}(r)$ and sector B above it, a straightforward application of Leibniz Rule gives the relationship between changes in reading skills and expected wages:

$$\begin{aligned} \frac{\partial E[w|r]}{\partial r} &= E_{u|r}\left[\frac{\partial w(r,u,S^{\star}(u,r))}{\partial r}|r\right] \\ &+ [w(r,u^{\star}(r),B) - w(r,u^{\star}(r),A)] \cdot f(u^{\star}|r) \cdot \frac{du^{\star}(r)}{dr} \\ &+ E_{u|r}\left[w(r,u,S^{\star}(u,r)) \cdot \frac{\partial \log f(u|r)}{\partial r}|r\right]. \end{aligned} \tag{1}$$

The first term is the direct effect of reading skills on wages, integrated out over all possible levels of the unknown skill. The generalized Roy model also generates two additional channels relating skills and wages. The second component, $[w(r,u^{\star}(r),B) - w(r,u^{\star}(r),A)] \cdot f(u^{\star}|r) \cdot \frac{du^{\star}(r)}{dr}$, is the mathematical expression of the starving artist explanation. $w(r,u^{\star}(r),B) - w(r,u^{\star}(r),A)$ is the net wage difference between sector B and sector A for those who are indifferent between the two sectors. If workers are only wage-maximizing, this will be 0. The larger the compensating differential, the larger the wage gap will be for workers on the margin between the sectors. The remainder of the starving artist term, $f(u^{\star}|r) \cdot \frac{du^{\star}(r)}{dr}$, represents the mass of workers who are actually on the margin of moving between sectors with a marginal change in r, which leads to a marginal change in the cutoff function $u^{\star}$.

The final term represents the bad sign explanation. The properties of the bad sign explanation are not obvious. This function has four useful properties:

1. If $u \perp r$, $E\left[w^{\star} \cdot \frac{\partial \log f}{\partial r}\right] = 0$.

2. If $w^\star(r,u,S^\star(u,r))$ is constant over u, $E\left[w^\star \cdot \frac{\partial \log f}{\partial r}\right] = w^\star \cdot E\left[\frac{\partial \log f}{\partial r}\right] = 0$.

3. If the joint distribution of reading skills and the unknown skill $f(r,u)$ is a bivariate standard normal with correlation coefficient ρ,

$$E\left[w^\star \cdot \frac{\partial \log f}{\partial r}\right] = \frac{\rho}{1-\rho^2}\operatorname{cov}(w^\star, u|r).$$

4. If $\frac{\partial F_{u|r}(z|r)}{\partial r} > 0\ \forall z$ and the sector-specific wage functions $w(r,u,A)$ and $w(r,u,B)$ are both strictly increasing in u, then for some sufficiently small positive level of compensating differential k_A, $E\left[w^\star \cdot \frac{\partial \log f}{\partial r}\right] < 0$.

    - $\frac{\partial F_{u|r}(z|r)}{\partial r} > 0$ is effectively assuming that every conditional distribution of the unknown skill given reading skill $f(u|r)$ is first order stochastically dominated by $f(u|r')$ for all $r' > r$.
    - Result 3 is a special case of this result.

Result 3 corresponds most closely to the intuition of the bad sign explanation. If the level of the unknown skill u is positively correlated with optimized wages, then there can be a reading penalty if the correlation between reading skills and the unknown skills are negative. A practical example of this would be if, conditional on math ability, young people who read more tend to be less organized and organization skill was more important in the labor market than reading skill.

The point of the model is identification. The remainder of this paper uses data on education and sectoral choice, which are both endogenous. The model skill gives some intuition for ways to try to reduce the selection bias:

1. If only the starving artist force is in effect, this is saying that the reason we estimate a negative effect is that we see that the effect of raising reading scores is confounding increases in reading skills with movements out of the sector that rewards reading skills due to the utility bonus in the other sector. Conditioning on sectoral choice, however, this bias should be largely eliminated: the within-sector effect of the starving artist story compares the wages of the marginal workers in a sector with the average wage

in that sector (given reading skills). If "sector" is narrowly enough defined in the data so workers are performing the same jobs within sectors, the wage structure should be basically the same for all workers in that sector and the measured reading penalty should no longer be negative if the true returns to reading are non-negative. Even if this isn't strictly true, the intuition that "zooming in" on workers in their particular job should reduce the measured importance of preferences, since most of the sorting on preferences has already happened at that point through education choice, college major choice, choice of which jobs to apply for, etc.

2. Only bad sign: consider some hypothesis for the unknown skill, e.g. organizational skills. Now look only at workers in jobs where those skills should have no effect on wages, that is, every sector S where $\frac{\partial w(r,u,S)}{\partial u} = 0$. The model then predicts that the bad sign term is 0 within that sector. Similarly, the stronger w moves with u in a sector, the larger the reading penalty within that sector. Taking this to the data requires knowledge of the returns to whatever skill u is hypothesized to represent, but if the returns are actually known the exclusion restriction can recover estimates of the return to reading scores cleaned of selection bias.

5 What Does the Data Say About Explanations?

In this section I use the intuition from the model to determine to what extent the starving artist and bad sign explanations are consistent with the career outcomes of the ASVAB test-takers. Up to this point, almost every result has only conditioned on variables that were predetermined at the time of the test (except for later entry into the military). Conditioning on endogenous variables later invites strong selection problems with interpreting the results. However, the generalized Roy model I analyzed above has given some answers to what kind of selection each explanation of the reading penalty would produce. Unless otherwise noted, the results from this section typically have the implicit assumption that the true returns to reading skills are zero in every sector. There is little evidence to suggest they are positive for more than a small fraction of jobs.

The Starving Artist

The results from the model suggested that if the starving artist explanation for the reading penalty were the dominant one, the reading penalty should be smaller within-sector than in the pooled regression. Additionally, the reading penalty should decrease the more narrowly the sector is chosen. This leads to an almost absurdly simple suggestion for determining if the starving artist explanation is the primary explanation of the reading penalty: condition on all measures of sector available.

A necessary condition for both the starving artist story and the bad sign story is that workers actually sort into different sectors based on their reading skills. To verify this, I regress non-wage outcomes such as education, occupation choice, and task characteristics of occupations onto ASVAB subtest scores and all other fixed characteristics I used in the wage regressions. Table 5 shows the results. There is nothing counter-intuitive in these results. Math is the most important predictor of later education, but even conditional on math the reading test score is highly correlated with completed education. In terms of occupational selection, conditional on the other test scores higher reading score individuals select into jobs such as teaching and administrative work, which also tend to have lower wages. At first glace, we could hope that this provides a basis for the starving artist theory: workers with higher reading scores may just becomes low-wage teachers because that's the job they like.

The starving artist explanation is not enough to explain the reading penalty, however. Table 6 shows a series of results that control for sector choice at different levels of aggregation. The first model adds dummy variables for each two-digit occupation to the wage regression (of which there are 22). The second model instead includes all 567 distinct occupation IDs. The other two models both include controls for a large set of occupational task descriptions from the O*NET database. In all specification, the reading penalty does not change. the easiest-to-interpret results from regressing log wages onto ASVAB subtest scores, the fixed characteristics, completed education, and the categorical O*NET occupation ID.

There may be within-occupational characteristics of jobs that individuals actually select on due to preferences, and this occupation-level analysis would miss any of those factors. For example, perhaps individuals with higher reading skills prefer jobs where they work fewer hours so they have more time to read in their free time. While the NLSY does not contain particularly fine on-the-job characteristics, hours worked is indeed a possible compensating differential I can check. Table 6 includes conditional on hours worked, and there is effectively no change in the reading penalty. I am still restricting the sample to 35+ hours, but the restriction on maximum hours is 90, which is binding on only an incredibly small part of the sample, so there is still significant variation in hours worked.

The Bad Sign

The analysis of the bad sign explanation will proceed a bit differently than that of the starving artist. First, I suggest three different possibilities for what the excluded skill u may represent.

1. People Skills
2. Organizational Skills
3. Ability to Think Creatively

Each of these three different possible skills have different implications for the relationship of the reading penalty across occupations. If People Skills were the true unmeasured skill, the model predicts that in jobs where people skills are irrelevant, the reading penalty should be basically 0. This is simply because if the reading penalty in the overall sample were because reading skills are negatively correlated with people skills, in jobs where people skills don't matter reading skills no longer give any information about output-relevant skills. On the other hand, the jobs where people skills are used the most would then be expected to have the highest reading penalty.

Using this argument, I run regressions interacting reading test scores with measures of the types of activities performed in occupations from the O*NET occupational characteristics database. The results from the specifications I use to test potential unknown skills are shown in Table 7. The first column shows the estimated reading penalties within each two-digit occupation. While all were estimated, the ones reported in the table were of the most interest. In particular, there is actually a positive measured return to reading test scores in the Construction sector, and essentially a 0 return within the Computer and Mathematical jobs sector. On the other hand, within Education and Social Sciences the reading penalty is even larger than it is in the overall sample.

The second model in Table 7 shows the results from interacting all the ASVAB variables with the set of occupation-level task descriptions from the O*NET. This regression has hundreds of covariates, but the most relevant ones for this question are that individuals in jobs with more data analysis face a larger reading penalty, those who spend more time "Getting Information" (which typically involves calling and talking to people) see a basically 0 wage penalty, and those who spend more time making decisions also see a bigger wage penalty.

These results to not suggest clearly whether the unknown skills is people skills, organizational skills, or creative skills, but this regression does seem to suggest it is not people skills. If higher reading scores proxied for a lack of people skills, we would expect to see the reading penalty most strongly in jobs intensive in people-related tasks, e.g. getting information.

For a simpler-to-interpret measure of the same idea, I create aggregate variables that represent measures of how important skills 1-3 are in any occupation. The third model in Table 7 shows the effect of interacting reading test scores with standardized (mean 0, variance 1) versions of my aggregated measures. The results here indicate rejection of the "lack of creativity" idea: in very creative occupations, reading test scores have a positive effect. Neither of the other two interactions are statistically significant, but organization is the most promising. Using the estimated values, individuals who are in the least organizationally-intensive jobs (conditional on all other tasks) see a reading penalty of basically 0, and those in the highest organizationally-intensive tasks see a statistically significant reading penalty.

The results currently do not deliver any clear picture of what types of skills the reading skill may be proxying for. On the other hand, given the lack of evidence for the starving artist explanation, the most promising avenue for future work is the ability to measure other particular skills along with reading skills to be able to more precisely estimate the cause of the reading penalty.

6 Conclusion

Reading comprehension is universally considered a skill. Indeed, in the reading comprehension test in the NLSY79, students scoring one standard deviation above the mean had average real wages $13 \pm 0.5\%$ higher over their careers. But conditional on all the other test scores, the relationship between reading skills and wages reverses: a one standard deviation increase in reading test scores is associated with a $1.5 \pm 0.7\%$ decrease in wages over their careers.

The goal of this paper was document and understand this negative conditional correlation of reading comprehension test scores with wages, which I called the *reading penalty*. I considered and rejected a "laundry list" of potential reasons for the reading penalty that have little economic content, such as potential problems with the tests themselves, functional form assumptions in the wage regression, or differences in socio-economic and educational background across test-takers. In the second part of the paper, I developed a simple generalized Roy model and showed it could deliver the reading penalty in two different ways: through the *starving artist* mechanism and the bad sign mechanism. In the starving artist story, workers with good reading skills may prefer working as a barista and writing a novel in their spare time to working as an accountant. In the bad sign explanation, if reading scores are a measure of ability to read but also proxy for (say) introversion, then if the returns to reading are sufficiently low relative to returns to people skills, higher reading test scores can be associated with lower wages.

Whether the reading penalty is due to preferences or productivity has implications for our understanding of what skills matter in the labor market and why. If preferences for low-wage jobs turn out to be the cause for the reading penalty, the tight connection between skills and outcomes that labor economists have typically assumed may need to be re-evaluated. On the other hand, if the reading penalty is proxying for a lack of some other useful skill, it is not clear that we know what we have been measuring through standardized tests, since just as easily math scores could be proxying for an abundance of that same unknown skill.

In my results, I find little support for the starving artist story. Instead, the reading penalty only marginally decreases after including controls for the worker's chosen sectors no matter how I define sectors, whether by one-digit or three-digit occupational codes or measures of the tasks performed in occupations. On the other hand, the bad sign explanation finds support in the data. The results suggest that workers in jobs where they perform high levels of organizational tasks face the largest reading penalty, while those in intensely interpersonal or creative jobs see a significantly lower wage reading penalty. I take this as preliminary evidence that reading skills are negatively correlated with organizational skills and that this drives the wage penalty.

My results imply that measurement of skills is currently a larger problem than labor economists believed. Factor-analytic methods typically choose a low number of skills based off the simple observation that most measures of skills are incredibly highly correlated. The reading penalty and the productivity explanation that best fits the data implies that residual skills can be important drivers of labor market outcomes even after factoring out the joint correlation , and those additional skills need to be measured directly to recover consistent estimates of which skills actually matter.

References

Altonji, J. G. (1995). The effects of high school curriculum on education and labor market outcomes. *Journal of Human Resources*, pages 409–438.

Hastings, J. S., Neilson, C. A., and Zimmerman, S. D. (2013). Are some degrees worth more than others? evidence from college admission cutoffs in chile. *Working Paper*.

Kinsler, J. and Pavan, R. (Forthcoming). The specificity of general human capital: Evidence from college major choice. *The Journal of Labor Economics*.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.

Seki, M. (2013). Heterogeneous returns to us college selectivity and the value of graduate degree attainment.

Taber, C. R. (2001). The rising college premium in the eighties: Return to college or return to unobserved ability? *The Review of Economic Studies*, 68(3):665–691.

Welsh, J. R., Kucinkas, S. K., and Curran, L. T. (1990). *Armed Services Vocational Aptitude Battery (ASVAB): integrative review of validity studies*. Air Force Human Resources Laboratory, Air Force Systems Command.

A Tables and Figures

Table 1: ASVAB Subtest Summary Statistics

	Mean	Std. Dev.	Max.
Science	16.8	(5.2)	25
Arithmetic	19.1	(7.5)	30
Word Knowledge	25.8	(8.0)	35
Paragraph Comp.	10.6	(3.6)	15
Numerical	33.1	(11.2)	50
Coding Speed	42.1	(15.9)	84
Auto and Shop	17.0	(5.2)	25
Mathematics	14.1	(6.7)	25
Mechanical	16.2	(5.3)	25
Electricity	12.9	(4.3)	20
Observations	2526		

Table 2: NLSY Sample Summary Statistics

	Mean	Std. Dev.	Min.	Max.
Year of Birth	1960.52	(2.2)	1957	1964
Years in Sample	26.38	(10.6)	1	36
Completed Years of Educ.	13.30	(2.9)	3	20
Took ASVAB	0.94	(0.2)	0	1
Took Rotter	0.99	(0.1)	0	1
Took Rosenberg	0.55	(0.5)	0	1
Took Pearlin	0.68	(0.5)	0	1
Cross-Sectional Observations	2682			

Table 3: ASVAB and Wages

	(1) Log Real Wage	(2) Log Real Wage	(3) Log Real Wage
Paragraph Comp.	0.131	-0.015	-0.019
	(0.002)	(0.004)	(0.004)
Current Age (Years)	0.085	0.081	0.080
	(0.002)	(0.002)	(0.002)
Current Age (Years)^2	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)
Science		-0.001	-0.008
		(0.004)	(0.005)
Arithmetic		0.004	0.005
		(0.005)	(0.005)
Word Knowledge		0.032	0.029
		(0.005)	(0.005)
Numerical		0.066	0.063
		(0.004)	(0.004)
Coding Speed		0.022	0.022
		(0.003)	(0.003)
Auto and Shop		0.001	0.011
		(0.004)	(0.004)
Mathematics		0.076	0.063
		(0.004)	(0.004)
Mechanical		0.016	0.012
		(0.004)	(0.004)
Electricity		0.008	0.014
		(0.004)	(0.004)
Highest Grade Completed, 1980			0.009
			(0.002)
Enrollment Status, 1980			
Not in HS			-0.105
			(0.012)
In HS			0.026
			(0.010)
In College			0.000
			(.)
No College, HS Grad			-0.052
			(0.007)
Observations	37363	37363	36828
R^2	0.22	0.27	0.28

Robust standard errors in parentheses.

Table 4: ASVAB Tests by Enrollment Status in 1980

	(1) Log Real Wage	(2) Log Real Wage	(3) Log Real Wage	(4) Log Real Wage
Paragraph Comp.	-0.025	-0.008	-0.068	-0.024
	(0.011)	(0.006)	(0.013)	(0.008)
Numerical	0.074	0.044	0.030	0.114
	(0.011)	(0.006)	(0.009)	(0.006)
Mathematics	-0.017	0.062	0.039	0.055
	(0.015)	(0.007)	(0.012)	(0.008)
Observations	4005	13965	6509	12349
Enroll Status	HS Dropout	In HS	In College	HS Grad, No Coll.
R^2	0.18	0.32	0.23	0.19

Robust standard errors in parentheses.

Other controls were other ASVAB tests, current age, highest grade completed in 1980.

Table 5: Selection on ASVAB Scores

	(1) Highest Completed Schooling	(2) 2-Digit Occ: Computers	(3) 2-Digit Occ: Teachers	(4) Task Intensity: Using Information
Science	0.446	-0.000	0.011	0.004
	(0.124)	(0.002)	(0.002)	(0.011)
Paragraph Comp.	0.355	-0.011	0.009	-0.018
	(0.114)	(0.002)	(0.002)	(0.010)
Numerical	0.120	-0.010	-0.001	-0.024
	(0.103)	(0.002)	(0.001)	(0.009)
Auto and Shop	-0.587	-0.014	-0.021	0.003
	(0.100)	(0.002)	(0.001)	(0.009)
Mathematics	0.878	0.034	-0.007	0.164
	(0.126)	(0.002)	(0.002)	(0.011)
Observations	952	32407	32407	30844
R^2	0.47	0.06	0.04	0.17

Robust standard errors in parentheses.

Other controls were other ASVAB tests, current age, highest grade completed in 1980, and year of birth.

Table 6: The Starving Artist

	(1) Log Real Wage	(2) Log Real Wage	(3) Log Real Wage	(4) Log Real Wage
Paragraph Comp.	-0.019	-0.021	-0.022	-0.022
	(0.004)	(0.004)	(0.004)	(0.004)
Mathematics	0.019	0.018	0.016	0.015
	(0.005)	(0.005)	(0.005)	(0.005)
Two-Digit Occupation				
Management	0.000		0.000	0.000
	(.)		(.)	(.)
Business	0.116		0.105	0.101
	(0.014)		(0.016)	(0.016)
Computer and Mathematical	0.197		0.168	0.161
	(0.013)		(0.018)	(0.018)
Legal	0.131		0.203	0.197
	(0.045)		(0.049)	(0.049)
Education	-0.108		0.038	0.030
	(0.018)		(0.023)	(0.023)
Admin/Office	-0.082		0.054	0.048
	(0.012)		(0.015)	(0.015)
Construction	-0.183		0.006	0.012
	(0.024)		(0.024)	(0.024)
Production	0.055		0.199	0.194
	(0.012)		(0.018)	(0.018)
Observations	32407	32407	30844	30844
Three-Digit Dummies	No	Yes	No	No
Task Intensity Controls	No	No	No	Yes
R^2	0.36	0.45	0.40	0.40

Robust standard errors in parentheses.
Other controls were other ASVAB tests, current age, highest grade completed in 1980,
highest completed education, and year of birth.

Table 7: The Bad Sign

	(1) Log Real Wage	(2) Log Real Wage	(3) Log Real Wage
Paragraph Comp.		-0.084	-0.035
		(0.057)	(0.020)
Mathematics		0.050	-0.098
		(0.061)	(0.020)
Paragraph Comprehension × 2-Digit Occupations			
PC × Computer and Mathematical	0.018		
	(0.025)		
PC × Social Sciences	-0.107		
	(0.034)		
PC × Education	-0.121		
	(0.031)		
PC × Construction	0.076		
	(0.029)		
PC × Production Workers	-0.045		
	(0.011)		
Paragraph Comprehension × Task Intensities			
PC × Analyzing Data		-0.037	
		(0.014)	
PC × Getting Information		0.055	
		(0.012)	
PC × Making Decisions		-0.036	
		(0.010)	
Paragraph Comprehension × Aggregated Tasks			
PC × Agg. Interpersonal			-0.001
			(0.001)
PC × Agg. Organize			-0.003
			(0.002)
PC × Agg. Creative			0.006
			(0.003)
Observations	32407	30844	30844
R^2	0.39	0.42	0.38

Robust standard errors in parentheses.
Other controls were other ASVAB tests, current age, highest grade completed in 1980, highest completed education, and year of birth.

Figure 1: Example Paragraph Comprehension Question

Hearsay evidence, which is the secondhand reporting of a statement, is allowed in court only when the truth of the statement is irrelevant. Hearsay that depends on the statement's truthfulness is inadmissible because the witness does not appear in court and swear an oath to tell the truth. Because his or her demeanor when making the statement is not visible to the jury, the accuracy of the statement cannot be tested under cross-examination, and to introduce it would be to deprive the accused of the constitutional right to confront the accuser. Hearsay is admissible, however, when the truth of the statement is unimportant. If, for example, a defendant claims to have been unconscious at a certain time, and a witness claims that the defendant actually spoke to her at that time, this evidence would be admissible because the truth of what the defendant actually said is irrelevant.

1. The main purpose of the passage is to

a. explain why hearsay evidence abridges the rights of the accused
b. question the probable truthfulness of hearsay evidence
c. argue that rules about the admissibility of hearsay evidence should be changed
d. specify which use of hearsay evidence is inadmissible and why

(Source: ASVAB: Core Review, LearningExpress, 1990)

Figure 2: Unconditional Wages by Number of Correct Paragraph Comprehension Answers

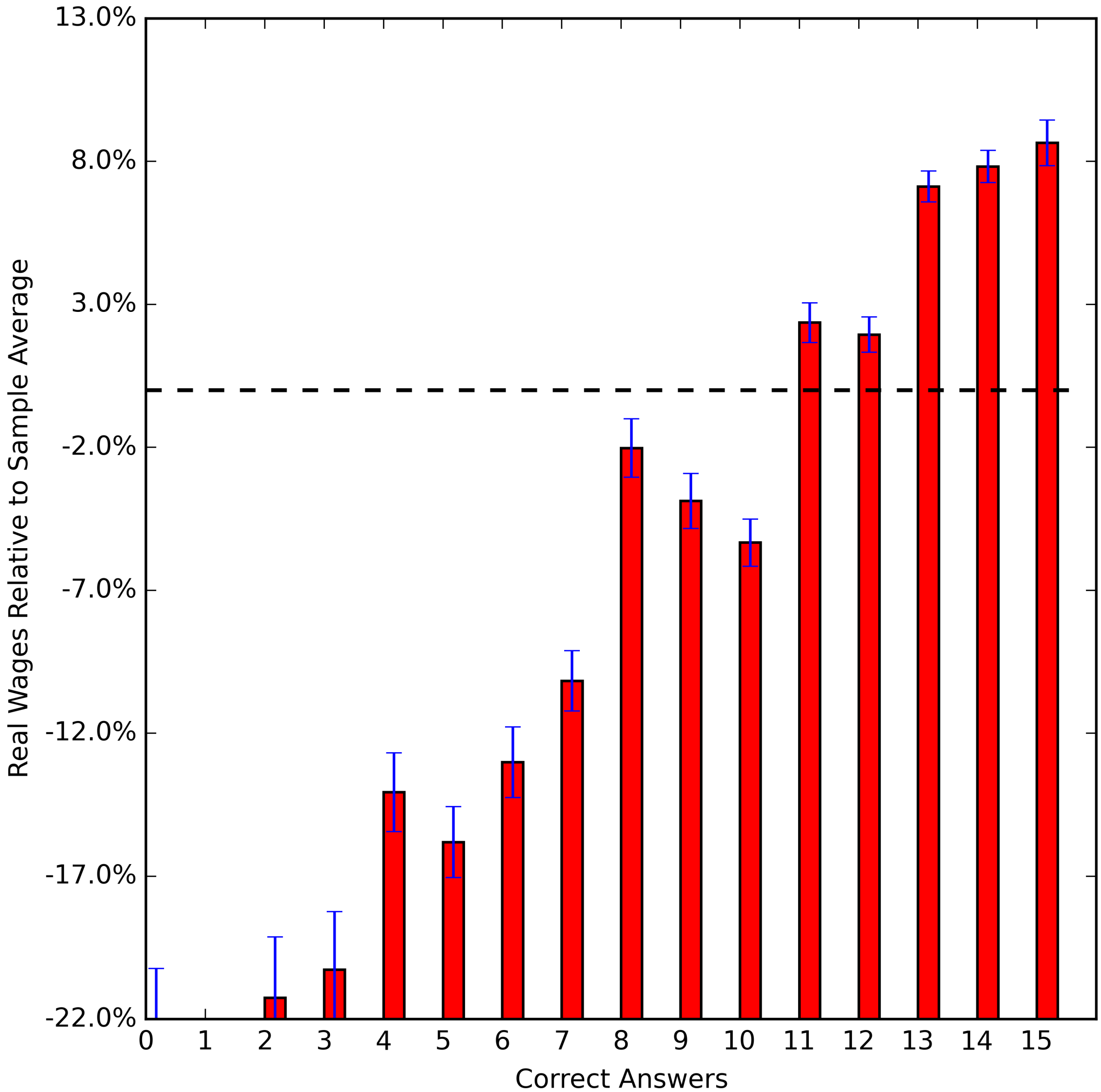

Figure 3: Conditional Wages by Number of Correct Paragraph Comprehension Answers