Wage and test score dispersion: Some international evidence

Kelly Bedard[⋆†] and **Christopher Ferrall**[⋆⋆]

[⋆]**Department of Economics**
University of California, Santa Barbara
Santa Barbara, CA 93106-9210
USA
kelly@econ.ucsb.edu

[⋆⋆]**Department of Economics**
Queen's University
Kingston, Ont K7L 3N6
Canada
ferrallc@post.queensu.ca

Received 23 January 2001; accepted 2 August 2001.

Abstract

Previous research has shown that wage-setting institutions help explain international differences in wage inequality. We expand on this theme to explore the role that educational institutions play in determining differences in wage dispersion across countries and within countries over time. We compare the distribution of test scores at age thirteen in 1964 and 1982 and wages later in life across eleven countries. We find that wage dispersion later in life is never greater than test score dispersion. In particular, Lorenz curves for a cohort's wages always lie above or on top of the cohort's test score Lorenz curve. Furthermore, wage dispersion, as summarized by Gini coefficients, is significantly related to test score dispersion in the country. A general fall in test score dispersion between 1962 and 1982 appears to be reflected in reduced wage dispersion within cohorts. For three countries with available data (the U.S., the U.K., and Japan), we find evidence of skill-biased changes in wage dispersion between the early 1970s and the late 1980s.

JEL Classification: I2, J3

Keywords: educational economics, human capital, salary wage differentials

1. Introduction

There is a vast literature documenting the rising wage inequality during the 1980s and early 1990s, differences in inequality across countries, and the potential explanations for the level differences and rate changes (see Levy and Murnane, 1992 and Gottschalk and Smeeding, 1997 for reviews of the literature). While there is widespread agreement about the actual levels and trends in wage inequality across countries, there is no consensus as to the underlying causes. Changes in industrial structure, skill-biased technical change, and increases in immigration and foreign trade have been put forward as candidate explanations for the upward trend in earnings inequality (see Danziger and Gottschalk, 1995 for a review). On the other hand, differences in the centralization of wage-setting institutions is the most common explanation for international variations in wage inequality.[1]

Although institutions such as unions and minimum wages have obvious direct and indirect affects on wages, there are many other differences across countries. Largely ignored in the discussion of differences in distribution of wages across countries are the significant differences in educational institutions across countries. In fact, the distribution of educational experience is one of the greatest differences in adult populations across countries (see Wößmann 2000). Major differences in education systems include whether and at what age students are streamed into different programs and whether or not expenditures per pupil are evenly distributed. For example, a school system that strives to bring all students to some pre-specified standard will focus its efforts and expenditures much differently than a system that is designed to produce a small educated elite. If countries differ in the way in which they distribute education and skills across students there will be cross-country differences in the distribution of basic skills brought to the labor market. As a result, the distribution of wages within a country depends on both the distribution of skills brought to the market and the labor market institutions more often studied in international comparisons.

In this paper we compare international data on the distribution of schooling outcomes and the distribution of wages later in life. The analysis is based on test score and wage observations across eleven countries and two birth cohorts. The test score data come from the First (1964) and Second (1982) International Mathematics Examinations (IME). The exams define two cohorts for which we calculate dispersion in annual wages for full time workers later in life using the Luxembourg Income Study (LIS) and other sources. These data sources, as with all other sources we know of, do not provide test scores and wages later in life for individuals in several countries.[2] Despite this limitation, the wage and test Lorenz curves provide novel evidence about the link between school systems and labor market outcomes.

Assuming that the distribution of mathematical ability is identical across countries at the point of school entry,[3] the distribution of math test scores at various points in the educational process are an important indication of the level of inequality inherent in an education system. We then compare the test score inequality with wage inequality within the same cohort later in life across countries.

A related literature studies the relationship between *current* skills and wages using the International Adult Literature Survey. Devroye and Freeman (2001) and Blau and Kahn (2001) both find that greater skill dispersion contributes to greater earnings inequality. While the relationship between current skills and wages is clearly interesting and important, it does not pin down when and where skills are acquired. For example, many basic skills are learned after one leaves school altogether. Indeed, Keane and Wolpin (1997) and Eckstein and Wolpin (1999) find that unmodeled and unobserved endowments of skill are critical elements for explaining the early schooling and career decisions of teenagers and young adults.[4] Our study of early school outcomes allows us to focus on the relationship between the distribution of skills during the schooling process and inequality in labor market outcomes later in life.

The empirical analysis indicates that wage and score dispersion follow predictable patterns across countries and cohorts. We find that the Lorenz curve for wages always dominates the Lorenz curve for test scores. That is, except for some borderline cases, the test scores of thirteen-year-olds are never less disperse than the distribution of wages later in life. Using

Gini coeficients to summarize the Lorenz curves, we find that countries with more unequal test score distributions tend to have more unequal wage distributions. Our descriptive results are important for policy, because they indicate scope for educational systems to affect the distribution of labor market outcomes later in life.

2. Data

2.1. Mathematics tests

In making international comparisons, mathematics is perhaps the best subject to focus on. First, Paglin and Rufolo (1990) find that math scores are the best academic predictor of future wages in the United States. Secondly, most countries devote twelve to fifteen percent of class time to the study of mathematics (Robitaille, 1990). Mathematics exams therefore measure the performance of a major component of the curriculum. The only other subject of similar importance to mathematics is language, which is inherently much more difficult to assess internationally.

The First and Second International Mathematics Examinations (IME) were conducted in 1964 and 1982 by the International Association for the Evaluation of Educational Achievement (IEEA). Twelve countries participated in the first IME (cohort I), and twenty-two countries participated in the second (cohort II). While the first IME test instrument and sampling procedure were initially criticized (Freudenthal, 1975), more recent studies have supported its validity (Keeves 1988). A committee of national representatives used curricula outlines and test construction recommendations submitted by participating countries to develop a test guideline that included the test format, topics, and prototype questions. The committee then invited participating countries to submit section-specific questions that complied with the guidelines. Using new and submitted questions the committee of representatives produced a preliminary exam which was circulated to national testing centers for preliminary testing and feedback. After reviewing the results, the final test instruments were agreed upon.[5]

To ensure that the samples would be representative, each country was stratified by geographic region and schools were randomly selected from within each region.[6] Students were selected to take exams at two points in their schooling when they could be identified consistently across countries. Namely, the IEEA gave an exam to the grade containing the most thirteen-year-olds and another exam to students enrolled in mathematics during their pre–university year. Since differences in high school curricula and participation rates make it difficult to define comparable samples of older students, we focus on the exams given to thirteen-year-olds. All participating countries had 100% school participation at this age in both 1964 and 1982.[7]

Table 1 lists the cohort I and II participants included in our study. Of the twelve countries participating in the first IME, we combine England and Scotland, and we exclude Israel because the high rate of immigration to Israel makes it dubious to link test scores in the 1960s to wages in the 1980s. Of the twenty-two countries participating in the second IME, we use all countries for which we could obtain the necessary wage data. The Canadian provinces of British Columbia and Ontario joined the second IME, and we were able to gather wage data separately for them. As with the 1964 scores, English and Scottish scores are combined as are Flemish and French scores for Belgium. The only countries that fall out from the first IME are West Germany and Australia.

[Table 1 here]

The 1964 (cohort I) data are taken from Husen (1967) and the 1982 (cohort II) data come directly from the New Zealand Ministry of Education.[8] Figure 1 graphs the test score distributions for cohorts I (thick line) and II by country (thin line). The percentages for cohort I have been converted to a density per one-unit score over the same 0 to 35 range as cohort II. Figure 1 tells a familiar cross-country story: the distribution of math test scores differs greatly across countries. For example, the 1964 Japanese distribution is nearly uniform, the Finnish distribution is nearly bell-shaped, and the U.S. and U.K. distributions are strongly skewed to the right. A comparison across cohort I and cohort II shows that the distributions within countries are typically pushed to the right. That is, mode scores are greater and most distributions are less right-skewed in cohort II than cohort I. For countries that participated in both tests the degree of push appears to differ. For Finland the mode is very similar in 1964 and 1982 while for France the cohort II distribution is close to a simple translation of the cohort I distribution.

[Figure 1 here]

The differences across cohorts in Figure 1 reflect an unknown combination of changes in mathematical skills and changes due to the testing instrument itself. Further, the striking differences in the shape of the distributions across countries imply that simple measures of dispersion, such as the coefficient of variation, may be incomplete if not inaccurate statistics with which to summarize and compare the distributions. To reduce the spurious effects of the testing instrument and to facilitate comparisons between test score and wage distributions that are robust to quite distinct distributions, we construct Lorenz curves for each country by cohort. Because the data are grouped (for example, in 1964 we know the number of students who answered 1 to 5 questions correctly) we linearly interpolate the test score percentiles. [9] The grouped data report the upper and lower score range $r = (u(r), l(r))$ and the number of students n_r scoring in that range. Let N_r be the cumulative number scoring in or below range r. A linear interpolation of the score percentile assumes that scores are uniformly distributed within a range. For a score $x \in \{0, 1, \ldots, X\}$, define $r(x) = r$ such that $l(r) < x \leq u(r)$. Then the linearly-interpolated distribution of scores is

$$F_s(x) \approx \frac{N_{r(x)-1} + \frac{n_{r(x)}}{u(r)-l(r)+1}}{N_{r(X)}}. \tag{1}$$

The percentiles are defined as usual:

$$Q_s(y) \equiv F_s^{-1}(y), \tag{2}$$

for $y = 1, 2, \ldots 99$. Figures 2 and 3 graph the test score Lorenz curves for cohorts I and II respectively. The score Lorenz curves are given by the thin lines in both figures. Similar to Figure 1, the Lorenz curves clearly document the substantial differences in test score distribution across countries.

[Figure 2 here]

[Figure 3 here]

2.2. Wages

In all cases wages are defined as annual wages rather than weekly or hourly earnings. This is the most consistent definition across countries. To further ensure comparability across countries, all samples are restricted to full-time male workers who are not self-employed and who would have been approximately thirteen years old at the time of the IME exam.[10] In other words, we use the sampling criteria to control for many of the individual characteristics that Blau and Kahn (1996) control for using variables in wage regressions. It is also possible to control for years of education and occupation. We purposely do not do this, because we are interested in the net link between early test scores and wages later in life. As Grogger (1996) points out, differences across school systems may ultimately encourage different educational attainment and different occupational choices. Our approach allows for the possibility that math skills feed directly into wages or instead lead to different educational paths that then feed into wages.

Table 1 lists all wage sample information for both IME cohorts, including data sources, sample sizes and sample years. The data for all but two countries are drawn from the Luxembourg Income Study (LIS). The Canadian wage data are from the Public Use File of the Census of Canada. The Japanese data come from Japanese Ministry of Labour (1989). Matching the test scores of individuals who are approximately thirteen years of age in 1964 and 1982 to wages for the same cohorts later in life is complicated by differences in wage sample years across countries. For example, the LIS contains data from the *French Income Survey of Taxes* for 1979 and 1984 and *Survey of Income and Program Users* data from the Netherlands for 1983, 1987, and 1991. In an attempt to be a thorough as possible Table 1 reports wage information for multiple years where possible. The wage observations are reported in chronological order for each cohort. In the case of the first French cohort, 1979 is designated as wage observation number one and 1984 is designated as wage observation number two. The reporting of multiple wage observations also allows us to ensure that results are not driven by the peculiarity of an individual survey year.

Table 2 summarizes each country-year specific Lorenz curve for wages by its Gini coefficient. All inequality measures are based on five year age groups centered on the age listed in column A for wage observations one through five. For example, the wage Gini coefficient for the second Finish cohort is based on full-time male workers aged 23-27 in 1987. The shaded observations in this table are the year/age group that capture each cohort at as similar ages as possible across countries. More specifically, shaded observations capture cohort I in their late thirties (in the mid to late 1980s) and cohort II as late as possible.[11] Throughout the remainder of the paper we focus on the shaded observations. However, all results are similar if alternate observations are used instead.

In addition to graphing the test score Lorenz curves, Figures 2 and 3 also graph the Lorenz curves for the shaded wages in Table 2. The wage Lorenz curves are given by the thick lines in both figures. As with test scores, the wage Lorenz curves differ significantly across countries and accord with stylized facts. For example, wages are more unequally distributed in the United States than in Finland.

[Table 2 here]

3. Results

3.1. Wage and test score dispersion within IME cohorts

The central question is whether or not there is a systematic relationship between test score and wage dispersion across countries. Referring to Figure 2, notice that the entire test score Lorenz curve lies below the wage Lorenz curve for every country in the cohort I, except for some slight overlap in Finland where the curves are nearly identical.[12] The important point is that within countries the wage curve dominates the score curve: At each percentile the the wage Lorenz curve for a country lies on or below the value of test score Lorenz curve for that country. However, this relationship does not hold between arbitrary pairs of curves. In many cases the test score Lorenz curve from one country crosses the wage Lorenz curve from another country. Thus, wage Lorenz curve dominance found *within* countries is not a spurious result that holds *across* countries.

Consider a few specific countries in cohort I: the United Kingdom, the United States, Australia, and Finland. The United Kingdom has both the most unequal distribution of test scores and the largest proportion of students with very low scores. However, wages in the United Kingdom are much more equally distributed than test scores. In contrast, test scores in the United States are somewhat less disperse than in the United Kingdom but the gap between wage and test score dispersion is much smaller in the United States. In this sense, the United Kingdom has the potential to have the greatest wage dispersion within this cohort but in fact has only the third greatest wage dispersion. Australia and Finland make for a similar comparison at the other end of the inequality spectrum. Finland has the most equal distribution of test scores and very few people scoring at the bottom. However, Finland's wage distribution is nearly the same as its score distribution and is more unequal than many other countries including Australia, which like the United Kingdom has substantially less wage inequality than score inequality.

The story is somewhat different for the second IME cohort (Figure 3). Test score dispersion appears to have fallen consistently across countries; Finland is the one exception. Despite this general shift upwards in the relative position of test score Lorenz curves, the selected wage observations still lay on top of or inside the score curves. Here the one exception is the United States, where the wage curve drops slightly below the score curve at the top of the distribution, leading to a slightly higher Gini coefficient.[13] This observation motivates our further exploration of the data.

As a cautionary note, because the two IME cohorts took different exams one should be dubious about any deep interpretation of changes in the test score Lorenz curves between

1964 and 1982. As a result, it is possible that changes in test score dispersion across cohorts may be due to differences in the test instrument rather than changes in educational practices. However, evidence presented later suggests that this is not the case.

Table 2 summarizes the test score and wage distributions shown in Figures 2 and 3. In both IME cohorts the United Kingdom had the most disperse, or unequal distribution of scores as measured by the Gini coefficient. Finland had the lowest amount of dispersion in the 1964 exam, and France had the lowest amount in 1982 (Japan follows quite closely). In terms of median scores, Sweden and Japan had the lowest and highest on both tests, although some countries switched rankings between exams. From Table 2, we see that the Gini coefficient in wages is smaller than that of the corresponding test score in all but four cases. In each of these cases the difference is not numerically large, and a comparison of the full Lorenz curves strongly suggests that the Gini coefficients are essentially identical in these four cases. In both cohorts the United States had the highest degree of wage dispersion. Australia had the least disperse wages in the first cohort, but it did not participate in the second IME. Of the cohort II participants, Japan had the lowest wage dispersion.

3.1. The score-wage gap

The patterns in Figures 2 and 3 are surprisingly consistent, but since Lorenz curves are functions it is not straightforward to establish whether wage and score dispersion move together using standard tools such as least squares or analysis of variance. One way to map two paired Lorenz curves into a bivariate relationship is to first summarize the dispersion inherent in the curves using the Gini coefficient. Then the wage and test score coefficients (G_w and G_s, respectively) can be included in a linear regression. A regression of G_w on G_s and a cohort indicator suggest that dispersion in wages and scores are correlated:

$$\begin{aligned} G_w \quad = \quad & .0488 \quad + \quad .3585 \times G_s \quad - \quad .0008 \times \text{CohortII} \\ & (1.61) \qquad\quad (1.96) \qquad\qquad\qquad (0.06) \\ & F_{2,16} = 4.08 \quad [p = 0.0371], \end{aligned}$$

where t-statistics in absolute value are reported in brackets. Surprisingly, the relationship is stable across the two cohorts, as the coefficient on cohort II is near zero and insignificant, despite the fact that wages are measured at different ages for the two cohorts and they took different exams. Figure 4 shows the stable relationship between G_s and G_w within and across cohorts. This finding suggests that the general drop in test score dispersion between 1964 and 1982 is not an artifact of the test instrument; wage dispersion decreased along with test score dispersion.

The regression line in Figure 4 also seems to separate countries with more decentralized wage setting (the United States and the United Kingdom) from those with more centralized wage setting (Sweden and the Netherlands). Dropping the insignificant cohort indicator and adding the country's union density (listed in Table 1) as a measure of wage centralization results in the regression:

$$\begin{aligned} G_w \quad = \quad & .062 \quad + \quad .342 \times G_s \quad - \quad .0003 \times \text{Union} \\ & (3.07) \qquad\quad (2.80) \qquad\qquad\quad (1.42) \\ & F_{2,16} = 5.60 \quad [p = 0.0143]. \end{aligned} \tag{3}$$

When controlling for test score dispersion, we find that greater union density is marginally related to lower wage dispersion. This result is consistent with Blau and Kahn's (1996) findings that the degree of wage centralization is negatively related to the wage inequality. While we find a (marginally) significant role for labor market institutions, the relationship between test score dispersion and wage dispersion is both statistically and economically significant. Based on our estimates, a 0.05 higher test score Gini coefficient is implies a 0.02 higher wage Gini coefficient.

[Figure 4 here]

3.3.An alternative approach

The Gini coefficient is just one of many accepted ways of summarizing a Lorenz curve and inequality in a distribution. As a check, considered the evidence obtained by comparing the inter-percentile ranges of the score and wage distributions. Recall that $Q_s(y)$ has been defined in (2) as the $100y^{th}$ percentile of the score distribution. Define $Q_w(y)$ for the corresponding wage percentiles, and let $d \in \{s, w\}$ be index for the score and wage distributions. Since wages are measured in local currencies and scores are measured in completely different units, it only reasonable to compare differences in percentiles that are normalized. In essence the Lorenz curve normalizes by the overall mean of the distribution. We can also normalize by $Q_d(.50)$ to define inter-percentile ranges expressed as a fraction of the median:

$$I_d(y_l, y_u) \equiv \frac{Q_d(y_u) - Q_d(y_l)}{Q_d(.50)}.$$

Table 3 reports $Q_d(y)$ for $y = .10, .25, .50, .75$, and .90 by cohort and country, and it reports $I_d(y_l, y_u)$ for $(y_l, y_u) = (.50, .10)$, $(.90, .50)$, and $(.90, .10)$. The table then reports the linear correlation $\text{corr}[I_s(y_l, y_u), I_w(y_l, y_u)]$ both within cohort and overall. As might be expected with such small samples and the moderate precision of the estimates in the Gini regression (3), the results are not highly significant. But they do exhibit some consistent and intriguing patterns. First the 50–10 ranges are negatively correlated (but not significantly different from uncorrelated). That is, countries with more dispersion in the lower part of the score distribution have less dispersion in the lower part of the wage distribution. The correlation turns to positive for the 90–50 ranges, and this correlation is large enough to make the overall 90–10 range always positive. Only when the two cohorts are combined can the hypothesis of no correlation between the 90–10 interpercentile ranges be rejected at the 10% level.

[Table 3 here]

3.4. Partial summary

We have presented three types of evidence for a relationship between score and wage dispersion. First, the Lorenz curve dominance in Figures 2 and 3. Second, the significant (and stable across cohort) relationship between Gini coefficients in equation (3) and Figure 4. Third, the consistent correlations between inter-percentile ranges in Table 3. In all cases, the small sample sizes and the tenuous (not at the individual level) pairing of outcomes due to data limitations leave the evidence at best suggestive. However, the evidence is consistent that the score distribution at age thirteen may have some direct or indirect effect on the wage distribution much later in life. It also seems self-evident that differences in schooling systems across countries would be a major component to any explanation of the wide differences in test score distributions across countries (Figure 1). Combined this with the score and wage pattern and we now have novel, albeit tentative, evidence that the schooling system affects the dispersion in labor market outcomes later in life.

3.5. Three observations holding age constant

Although there appears to be a steady relationship between wage and score dispersion across the two cohorts, the comparison is based on observations of wages at different ages, when the older cohort is in its late thirties and the younger cohort is in its twenties. For three countries (Japan, the United Kingdom, and the United States) we have wage data from the early 1970s when the first cohort was also in its twenties. For these countries Figure 5 compares wage Lorenz curves for the two cohorts at the same age. The wage Lorenz curves are labeled by cohort (test score curves are not shown). Wage dispersion (for men in their mid-twenties) increased slightly in the United Kingdom and the United States and remained nearly constant in Japan across the two cohorts, despite the fifteen year difference and the large shift in test scores. We define $L^{\star}_w(z)$ as the predicted value of the Lorenz curve for cohort II given their test scores and the relationship between $L_w(z)$ and $L_s(z)$ for the earlier cohort. That is, we define

$$\lambda(z) \equiv \frac{z - L^I_w(z)}{z - L^I_s(z)}$$

as the proportion of the test-score gap that is made up in wages by the bottom z percent of cohort I in the early 1970s. If we hold $\lambda(z)$ constant and treat the second cohort's score distribution, $L^{II}_s(z)$, as the new distribution of the same skills as cohort I, then the predicted Lorenz curve for cohort II is

$$L^{\star}_w(z) \equiv z - \lambda(z)\Big(z - L^{II}_s(z)\Big).$$

For all three countries $L^{\star}_w(z)$, the unlabelled thick curve in Figure 5, lies well within both of the actual wage Lorenz curves. That is, holding the distribution of math skills constant, there is more wage dispersion among the second cohort than for the earlier cohort at the same age. There are two possible explanations for this finding. One is that the drop in score dispersion among the second IME cohort is primarily an artifact of the tests, and the 'excess' wage dispersion in Figure 5 experienced by the second cohort while in their twenties is not excessive. This explanation depends on the steady relationship in Figure 4 between G_w and G_s being a coincidence in how the exam changed between cohorts. However, evidence presented earlier in the paper does not support this interpretation. An alternative explanation is that the drop in test score dispersion between the two cohorts was real and the excess wage dispersion in Figure 4 reflects a general increase in skill-related wage dispersion. Contrary to most studies of wage inequality, our results are conditioned upon a measure of skill dispersion independent of labor market outcomes, and may be interpreted as alternative evidence for skill-biased changes from the early 1970s to the late 1980s. This further suggests that wage inequality measures understate the true increase in inequality arising from changes in labor market institutions and technological change because younger workers are actually arriving at the labor market with more evenly distributed skills.

[Figure 5 here]

4. Conclusion

The labor markets in the industrialized countries examined in this paper differ in a variety of ways. They vary in their degree of unionization and centralized bargaining, minimum wage laws, and immigration policies. But it is difficult to argue that any of these institutional differences are greater than the differences in educational structures. It may seem somewhat surprising that researchers interested in income inequality differences across countries have not addressed the role of educational systems. The explanation for this gap in the research is the unrealistic demands that such a study places on available data. The on-going development of international testing and comparable longitudinal data sources makes us optimistic that the descriptive analysis presented in this paper is but a first step in exploring the power of international differences in educational systems to explain international differences in labor market outcomes.

We have compared the distribution of scores on math test given at age thirteen to the distribution of wages later in life for two cohorts of people in eleven countries. Our main results are as follows: Wage dispersion is consistently lower than test score dispersion across both countries and time; test score dispersion fell between 1964 and 1982; this fall in dispersion seems to be at least partially reflected in wage dispersion in the late 1980s; union density in the country helps explain the gap between wage and test score dispersion.

Our results suggest that countries historically differed greatly in the distribution of skills they imparted to students, that these skills help determined wages later in life, and that international differences in pre-market skills help explain international differences in wage dispersion. Our results also suggest that the distributions of skills brought to the labor market have been converging over the last thirty years across several industrialized countries. If so, increases in wage inequality may *understate* the effect of conventional explanations such as government wage policy and technological change, because as compared to their predecessors young people now bring more equal skills to the labor market but receive less equal wages.

Acknowledgements

Ferrall acknowledges research support from the Social Sciences and Humanities Research Council of Canada. For comments and suggestions on earlier drafts we thank Charles Beach, Russell Davidson, Allan Gregory, Peter Kuhn, Craig Riddell, and participants in seminars at Brock, Edinburgh, Glasgow, Lakehead, and the Canadian International Labour Network.

Notes

1 Gottschalk and Joyce (1998) and Blau and Kahn (1996), among others, document that large differences exist in wage dispersion across countries. For example, Blau and Kahn find that the difference between the 90th and 10th percentiles in log wages is more than twice as large in the U.S. as in Norway. They attribute a large portion of the differences in dispersion across countries to wage–setting institutions, particularly unions.

2 The test scores for individual students are not available to us, and to our knowledge this is the first attempt to link the IME to wages later in life. Extensive studies of mean test scores for the first IME cohort are reported in Husen (1967) and Heyneman and Loxley (1983) and for the second cohort in Robitaille (1989).

3 We are not assuming that parental inputs do not affect mathematical aptitude, only that the distribution of parental involvement is same across developed countries.

4 Using data from the British National Development Survey, Currie and Thomas (1999) show that test scores as early as age seven are related to subsequent labor market outcomes.

5 The 1982 test format is slightly different in that each student wrote a core form of forty questions as well as one of four rotated forms of thirty-five questions. The scores reported in this paper are the average of the rotated forms. All results are similar if the core form is used instead.

6 West Germany is the only exception. West Germany made participation voluntary and then randomly selected schools from the set of volunteers.

7 In countries where streaming of students had already occurred by age thirteen repre-

sentative samples were constructed.

8 The raw test score data are available from the authors.

9 For the purpose of constructing Lorenz curves it is important that we do not transform the raw data in any way that depends upon the distribution itself.

10 In the published IME data we cannot separate test score distribution for boys and girls, so the test score distributions include all students. Husen (1967) reports a significant difference in means for boys and girls in cohort I within only three countries: Belgium, Japan, and the Netherlands. For cohort II, Robitaille (1989) reports "practically significant" differences across boys and girls by subtest. Four countries had significant differences in more than one of the four sub-tests: Belgium, Finland, France, and the Netherlands. At this age, when there are differences in mean scores neither sex always performs better than the other.

11 Because the 1982 IME cohort is so young, for most countries we are comparing their score distribution to the wage distribution of people five to ten years older than them. Table 2 reports the year and age bracket used. We are implicitly assuming that national test score distributions for thirteen-year-olds did not change greatly between the mid 1970s and 1982.

12 Since test score dispersion is greater than wage dispersion, adjusting raw test scores for question difficulty (holding the mean score constant) will tend to push test score Lorenz curve down even further. This of course presumes that the questions answered correctly by low scoring students were also answered correctly by high scoring students so that quality-adjusted scores are more disperse than raw scores.

13 The Gini coefficients reported in this paper are the usual definition divided by 2.

Figure Captions

Figure 1: Test Score Distribution by Country, Cohort I (thick) and Cohort II

Figure 2: Wage (thick) and Score Lorenz Curves, Cohort I

Figure 3: Wage (thick) and Score Lorenz Curves, Cohort II

Figure 4: Wage and Test Score Dispersion, Both Cohorts

Figure 5: Wage Dispersion Early in Life, Both Cohorts

References

Blau, F. & Kahn, L. (1996). International differences in male wage inequality: institutions versus market forces, *Journal of Political Economy* 104(4), 791-836.

Blau, F. & Kahn, L. (2001). Do cognitive test scores explain higher us wage inequality? NBER, Working Paper No.8210.

Currie, J. & Thomas , D. (1999). Early test scores, socioeconomic status and future outcomes, NBER, Working Paper No. 6943.

Danziger, S. & Gottschalk, P. (1995). *America unequal,* New York: Russell Sage Foundation.

Devroye D., & Freeman, R. (2001). Does inequality in skills explain inequality of earnings across countries? NBER, Working Paper No.8140.

Eckstein, Z. & Wolpin, K. (1999). Why youths drop out of high school: the impact of preferences, opportunities, and abilities, *Econometrica* , 67(6), 1295-1339.

Freudenthal, H. (1975). Pupils' achievement internationally compared - the IEA, *Educational Studies in Mathematics 6.*

Gottschalk, P. & Joyce, M. (1998). Cross-national differences in the rise in earnings inequality: market and institutional factors, *Review of Economics and Statistics* 80(4), 489-502.

Gottschalk, P. & Smeeding, T. (1997). Cross-national comparisons of earnings and income inequality, *Journal of Economic Literature* XXXV, 2, 633-688.

Grogger, J. (1996). School expenditures and post-schooling earnings: evidence from high school and beyond, *Review of Economics and Statistics* , 78(4), 628-637.

Heyneman, S. & Loxley., W. (1983). The effect of primary school quality on academic achievement across twenty-nine high- and low-income countries, *American Journal of Sociology 88(6)*, 1162-1194.

Husen, T. (1967). *International study of achievement in mathematics, a comparison of twelve countries*, Stockholm: Almqvist and Wiksell, Volumes 1 and 2.

Japanese Ministry of Labour (1989). *Yearbook of labour Statistics.*

Keane, M. & K. Wolpin, K. 1997. The careers of young men, *Journal of Political Economy* ,105(3), 473-522.

Keeves, J. (1988). Cross-national comparisons in educational achievement: the role of the international association for the evaluation of educational achievement (IEA), in S.P. Heneman and I. Fagerlind (eds.), *University Examinations and Standardized Tests: Principles, Experience, and Policy Options*, World Bank Technical Paper No. 78.

Levy, F. & Murnane, R. (1992). U.S. earnings levels and earnings inequality: a review of recent trends and proposed explanations, *Journal of Economic Literature* 30(3), 1333-1381.

OECD (1981). *OECD Studies in Taxation. Income tax schedules : distribution of taxpayers and revenues*, Paris: OECD Publications and Information Center.

OECD (1989-1990), *Education in OECD Countries.*

Paglin, M. & Rufolo, A. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences, *Journal of Labor Economics* 8(1), 123-144.

Robitaille, D. (1990). Canadian participation in the second international mathematics study, working paper, Economic Council of Canada No. 6.

Robitaille, D. (1989). Students' achievements: population A, in D. Robitaille and R. Garden (eds) *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematcis*, Oxford: Pergamon.

Traxler, F. (1996). Collective bargaining and industrial change: a case of disorganization? a comparative analysis of eighteen OECD countries, *European Sociological Review 12(3)*, 271-287.

UNESCO (1964, 1986, 1987), *Yearbook.*

Wößmann, L. (2000). Schooling resources, educational institutions, and student performance: the international evidence, Kiel Institute of World Economics, Working Paper No. 983.

Table 1 Sources of wage data, sample sizes, pupil-teacher ratios, union density

	Country	Original data source for wage data	LIS	Full time A	Full time B	Full time C	IME Coh.	Wage obs 1 N	Wage obs 1 Year	Wage obs 2 N	Wage obs 2 Year	Wage obs 3 N	Wage obs 3 Year	Wage obs 4 N	Wage obs 4 Year	Wage obs 5 N	Wage obs 5 Year	PTR	Union
1	B. C.	*Census of Canada Public Use Files*		*	*		II	598	86	1410	91							.	35.8
2	Japan	*Census of Wages*				*	I	1904	69	2741	74	1775	79	1679	83	23243	89	25	25.4
							II	4650	89									25	25.4
3	Ontario	*Census of Canada Public Use Files*		*	*		II	3355	86	6254	91							.	35.8
4	U. K.	*Family Expenditure Survey*	*			*	I	526	74	444	86							25	39.1
							II	416	86									18	39.1
5	Australia	*Income and Housing Survey*	*	*		*	I	1184	81	561	85	882	89					30	40.4
6	Belgium	*Living Conditions of Households*	*	*			I	520	85	308	88	250	92					.	51.4
							II	390	85	289	88	284	92					.	51.4
7	Finland	*Survey of Income Distribution*	*	*			I	923	87	851	91							24	72
							II	645	87	541	91							15	72
8	France	*French Income Survey of Taxes*	*				I	773	79	896	84							29	9.8
							II	275	84									21	9.8
9	W. Germ.	*German Panel Survey: Wave 2*	*	*	*		I	3567	78	192	81	412	84					29	32.9
10	Nether.	*Survey of Income and Program Users*	*	*			I	454	83	367	87	313	91					33	25.5
							II	307	87	365	91							20	25.5
11	Sweden	*Swedish Income Distribution Survey*	*			*	I	728	75	620	81	749	87	871	92			15	82.5
							II	415	87	520	92							16	82.5
12	U.S.A.	*Current Population Survey*	*	*	*		I	416	69	858	74	973	79	684	86	891	91	24	15.6
							II	654	86	1064	91							21	15.6

LIS=* means our source for the data was the Luxembourg Income Survey database; N=sample size; Year=survey year.
All samples are also restricted to: male, not self-employed, and the Full Time criteria A-C as allowed by the data source.
Col A: '30 hours or more of work per week.
Col B: 46 or more weeks per year.
Col C: 'Full time employee
Col PTR: Pupil teacher ratio; Source: UNESCO (1964, 1986, 1987).
Col Union: union density; Source: Traxler (1996).
Row 2: Our Source: Japanese Ministry of Labor (1989); Sample sizes are in thousands
Row 6: Earnings converted to pre-tax using schedules in OECD (1981)
Row 8: The data contains no information on full time status.

Table 2 Wages later in life for the two IME cohorts

	IME country	IME 13-year-olds				Distribution of wages later in life for IME cohort									
		IME	Sample	Median	Score	Obs. 1		Obs. 2		Obs. 3		Obs. 4		Obs. 5	
		coh.	size	score	Gini	A	Gini	A	Gini	A	Gini	A	Gini	A	Gini
1	Brit. Col.	II	2167	17.4	0.107	24	**0.123**	29	0.105						
2	Japan	I	2050	16.0	0.154	22	0.077	27	0.066	32	0.065	37	0.071	42	0.087
		II*	8091	21.0	0.090	27	0.061								
3	Ontario	II	4666	16.0	0.108	24	**0.121**	27	0.106						
4	U. K.	I*	3552	10.4	0.211	27	0.083	39	0.121						
		II*	2483	15.0	0.133	24	0.106								
5	Australia	I*	3078	9.0	0.174	34	0.092	38	0.085	42	0.110				
6	Belgium	I	2645	15.5	0.131	38	0.097	41	0.094	45	0.100				
		II	2724	19.5	0.099	23	0.076	26	0.072	30	0.076				
7	Finland	I*	841	12.9	0.104	40	0.106	44	0.104						
		II	4382	14.0	0.122	25	0.079	29	0.074						
8	France	I	3449	9.4	0.177	32	0.112	37	0.132						
		II*	2483	18.0	0.084	22	0.093								
9	W. Germ.	I	4475	12.6	0.133	33	0.075	34	0.078	37	0.093				
10	Nether.	I	1433	10.0	0.159	36	0.088	40	0.106	44	0.105				
		II	5418	17.7	0.113	25	0.065	29	0.065						
11	Sweden	I	2828	6.8	0.189	28	0.069	34	0.081	40	0.106	45	0.105		
		II	3451	12.4	0.123	25	0.076	30	0.077						
12	U.S.	I*	6544	8.2	0.196	22	0.113	27	0.101	32	0.114	39	0.143	44	0.138
		II*	6648	15.1	0.121	24	**0.136**	29	**0.128**						

Wages from five-year age category centered on age A (see Table 1 for more details).
Bold Gini Coefficient means that the Wage coefficient > Score coefficient;
* means row contains a score or selected wage coefficient that is a mininum or maximum.
Shading indicates the observation chosen to represent the country and the IME cohort.

Table 3 Wage and score percentiles and their interpercentile ranges

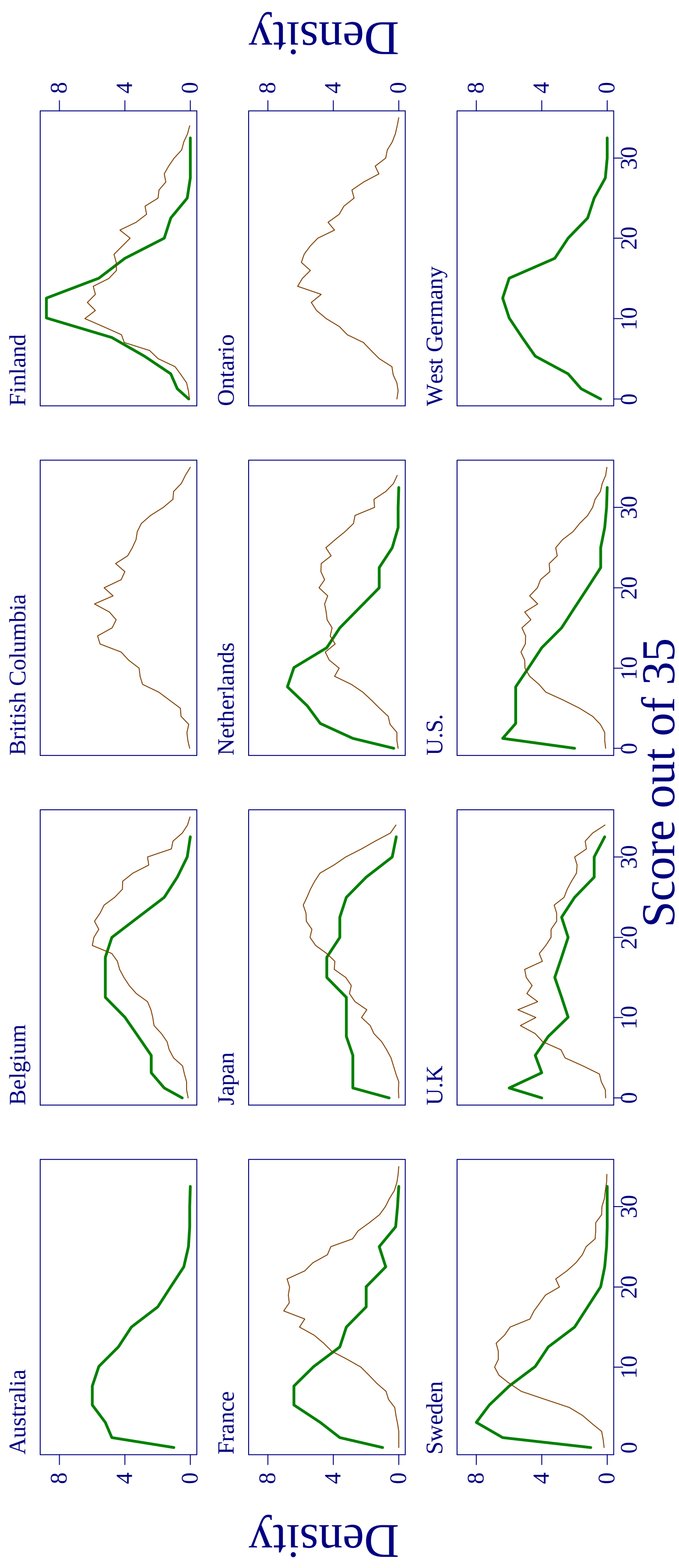
	Percentiles										Interpercentile range / median					
	Test scores					Wages[a]					(50-10)		(90-50)		(90-10)	
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th	Score	Wage	Score	Wage	Score	Wage
Cohort I[b]																
Finland	6.8	10.1	12.9	16.2	19.5	69,000	82,500	103,000	136,000	173,000	47.8	33.0	50.5	68.0	50.5	68.0
W. Germany	4.8	8.2	12.6	16.6	20.7	29,400	35,700	43,200	55,000	65,900	61.6	31.9	64.1	52.5	64.1	52.5
Belgium	4.8	10.1	15.5	20.3	23.9	3,900	4,500	5,100	6,600	9,000	69.0	23.5	54.4	76.5	54.4	76.5
Netherlands	3.1	6.0	10.0	14.5	19.0	33,630	39,986	50,256	64,680	84,080	69.2	33.1	91.3	67.3	91.3	67.3
France	2.5	5.5	9.4	14.7	20.5	50,682	65,791	87,941	114,435	172,990	73.3	42.4	118.7	96.7	118.7	96.7
Japan	3.4	8.6	16.0	22.1	26.4	24,585	30,386	37,444	46,200	56,246	79.0	34.3	64.8	50.2	64.8	50.2
Sweden	1.4	3.5	6.8	11.1	15.0	92,587	115,472	136,665	173,230	225,242	79.3	32.3	120.4	64.8	120.4	64.8
Australia	1.9	4.8	9.0	13.6	17.5	16,705	20,500	25,310	31,856	36,977	79.2	34.0	94.4	46.1	94.4	46.1
U.S.	1.3	3.8	8.2	13.3	18.0	16,000	24,000	34,000	45,094	60,000	84.8	52.9	119.1	76.5	119.1	76.5
U.K	1.1	4.1	10.4	19.0	24.8	6,699	8,639	11,595	15,179	20,124	89.7	42.2	139.7	73.6	139.7	73.6
Cohort II[b]																
France	10.8	14.1	18.0	21.8	25.1	25,800	43,267	53,912	64,180	74,705	39.9	52.1	39.3	38.6	39.3	38.6
Japan	10.8	15.7	21.0	25.4	28.4	18,462	20,707	23,579	27,253	31,876	48.3	21.7	35.5	35.2	35.5	35.2
Finland	6.9	9.8	14.0	19.5	24.1	83,500	93,500	111,000	131,000	155,000	50.6	24.8	72.5	39.6	72.5	39.6
Sweden	6.1	8.7	12.4	16.9	21.2	128,196	154,643	178,044	207,534	248,896	50.6	28.0	71.0	39.8	71.0	39.8
Ontario	8.2	11.5	16.0	20.7	25.0	18,363	25,628	33,192	42,000	50,068	48.8	44.7	56.1	50.8	56.1	50.8
Brit. Columbi	8.7	12.5	17.4	22.8	27.0	18,000	25,000	34,000	42,000	50,645	50.0	47.1	54.8	49.0	54.8	49.0
Belgium	9.4	14.4	19.5	23.9	27.5	3,902	4,502	5,042	6,002	7,203	52.0	22.6	41.2	42.9	41.2	42.9
Netherlands	8.1	11.8	17.7	23.1	26.9	34,555	40,271	46,526	53,801	62,294	54.0	25.7	51.9	33.9	51.9	33.9
U.S.	7.0	10.2	15.1	20.7	25.1	12,480	18,000	25,000	34,000	45,000	53.5	50.1	65.9	80.0	65.9	80.0
U.K	6.5	9.8	15.0	21.3	26.7	5,001	5,997	7,854	10,093	12,970	56.5	36.3	78.4	65.1	78.4	65.1
Correlation between wcore and wage ranges[c]																
Cohort I											-0.23		0.35		0.52	
Cohort II											-0.38		**0.62**		0.31	
Combined											-0.27		**0.41**		**0.56**	

[a] In local currency, current values, except: Japan is Yen/100

[b] Countries are sorted within cohorts by the (50-10) score column.

[c] Values in **bold** are significantly different from 0 at the 10% level.

Figure 1



Australia
Belgium
British Columbia
Finland
France
Japan
Netherlands
Ontario
Sweden
U.K
U.S.
West Germany
Density
Score out of 35

Figure 2

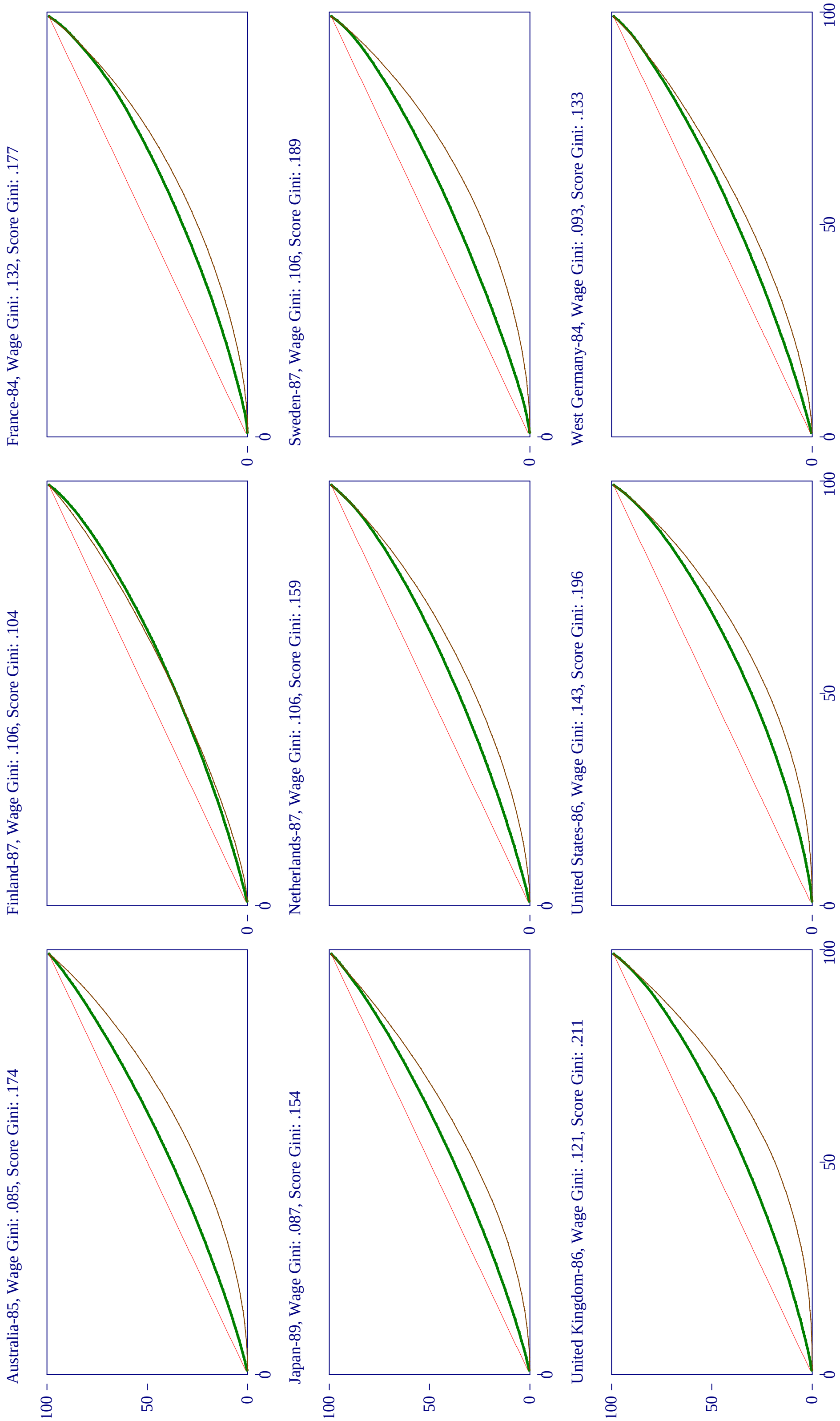
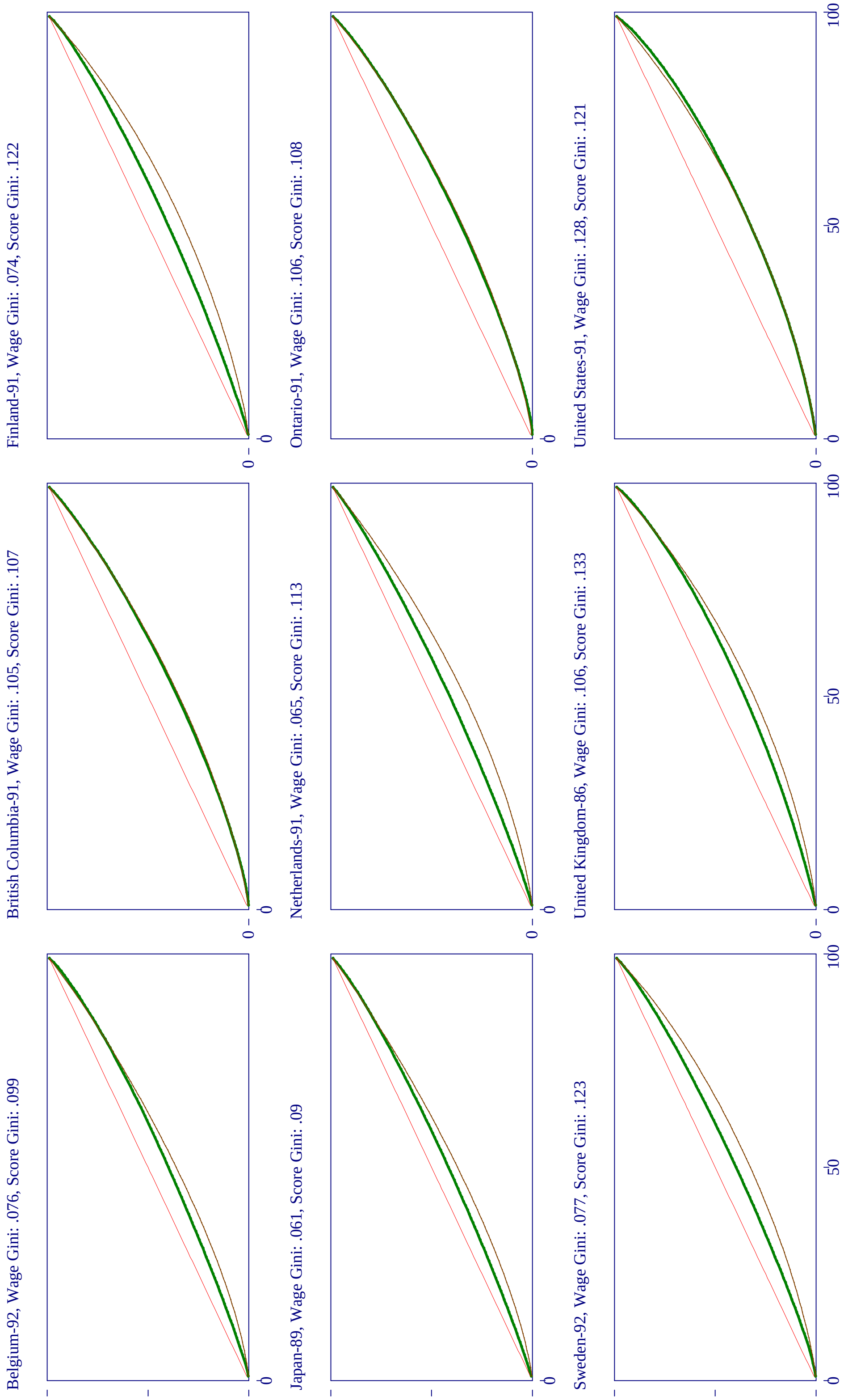

Australia-85, Wage Gini: .085, Score Gini: .174
Finland-87, Wage Gini: .106, Score Gini: .104
France-84, Wage Gini: .132, Score Gini: .177
Japan-89, Wage Gini: .087, Score Gini: .154
Netherlands-87, Wage Gini: .106, Score Gini: .159
Sweden-87, Wage Gini: .106, Score Gini: .189
United Kingdom-86, Wage Gini: .121, Score Gini: .211
United States-86, Wage Gini: .143, Score Gini: .196
West Germany-84, Wage Gini: .093, Score Gini: .133
0
50
100

Figure 3



Belgium-92, Wage Gini: .076, Score Gini: .099
British Columbia-91, Wage Gini: .105, Score Gini: .107
Finland-91, Wage Gini: .074, Score Gini: .122
Japan-89, Wage Gini: .061, Score Gini: .09
Netherlands-91, Wage Gini: .065, Score Gini: .113
Ontario-91, Wage Gini: .106, Score Gini: .108
Sweden-92, Wage Gini: .077, Score Gini: .123
United Kingdom-86, Wage Gini: .106, Score Gini: .133
United States-91, Wage Gini: .128, Score Gini: .121
0
50
100

Figure 4

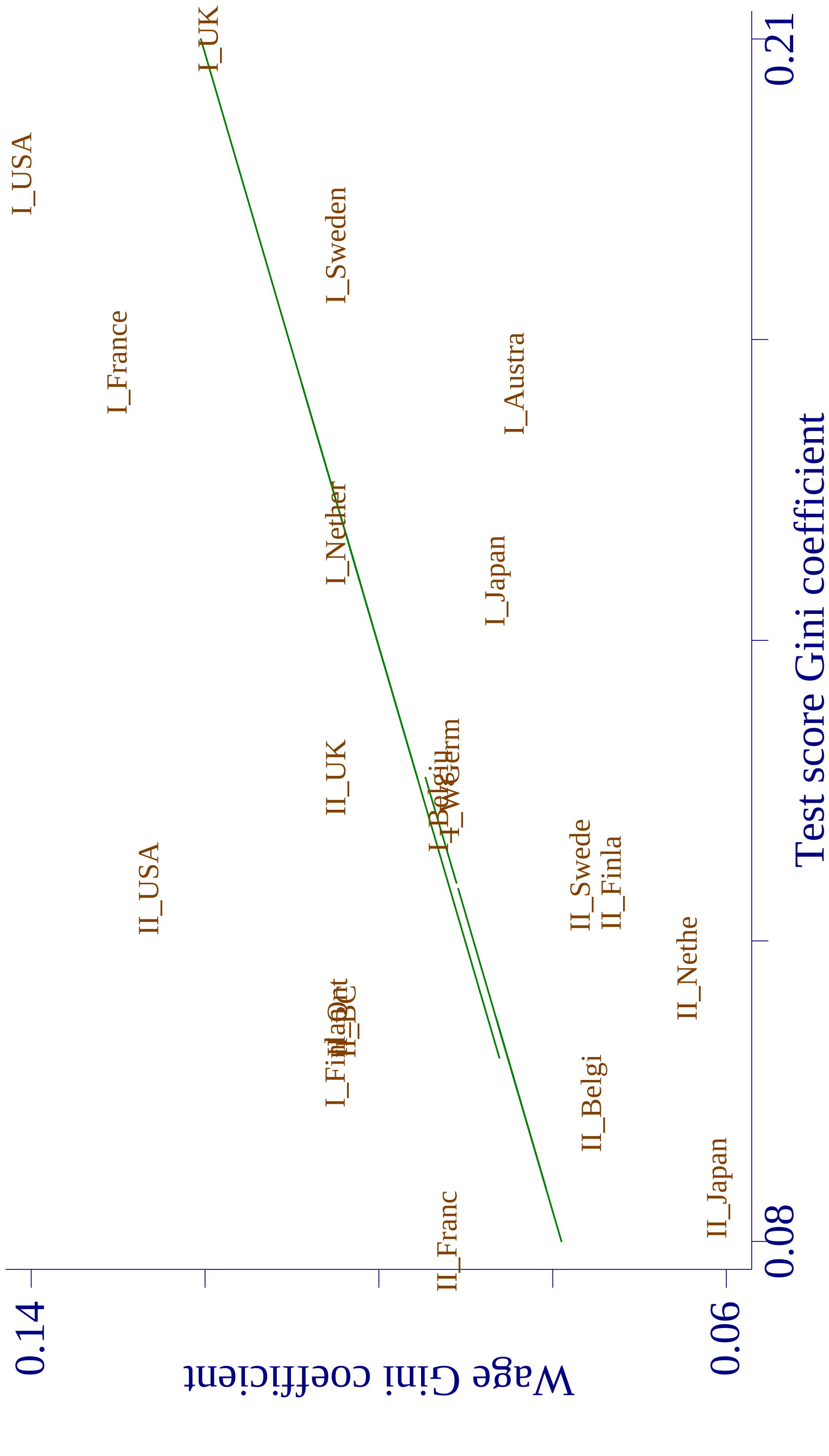
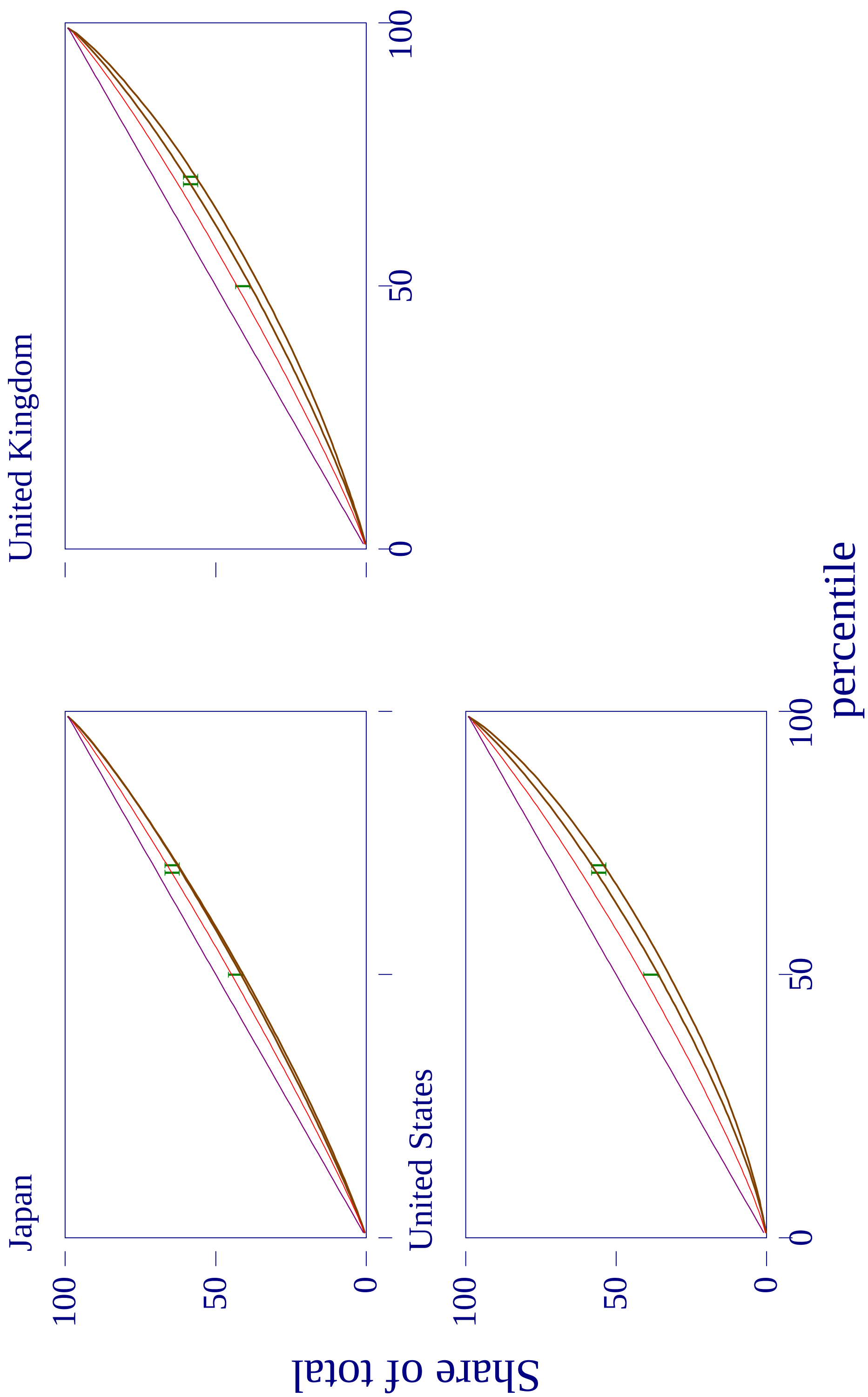

Test score Gini coefficient
0.21
0.08
Wage Gini coefficient
0.14
0.06
I_UK
I_USA
I_Sweden
I_France
I_Austra
I_Nether
I_Japan
II_UK
II_USA
II_Swede
II_Finla
II_Nethe
II_Belgi
II_Japan
II_Franc

Figure 5


Japan
United Kingdom
United States
Share of total
percentile
0
50
100
I
II