Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Universität Konstanz

Rechts-, Wirtschafts- und
Verwaltungswissenschaftliche
Sektion
Fachbereich
Wirtschaftswissenschaften

Diskussionspapiere der DFG-
Forschergruppe (Nr.: 3468269275):

Heterogene Arbeit: Positive und Normative
Aspekte der Qualifikationsstruktur der Arbeit

Andreas Ammermüller

PISA: What makes the Difference ?

Februar 2004

Diskussionspapier Nr. 04/07
http://www.wiwi.uni-konstanz.de/forschergruppewiwi/

Diskussionspapier der Forschergruppe (Nr.: 3468269275) “Heterogene Arbeit: Positive und Normative Aspekte der Qualifikationsstruktur der Arbeit“

Nr. 04/07, Februar 2004

PISA: What Makes the Difference ?

Andreas Ammermüller

Centre for European Economic Research (ZEW)
68161 Mannheim
Germany

mail: ammermueller@zew.de
phone: +49-621-1235-284
fax +49-621-1235-225

Zusammenfassung:

Abstract:

The huge difference in the level and variance of student performance in the 2000 PISA study between Finland and Germany motivates this paper. It analyses why Finnish students performed so much better by estimating educational production functions for both countries. The difference in the reading proficiency scores is assigned to different effects, using Oaxaca-Blinder and Juhn-Murphy-Pierce decomposition techniques. The analysis shows that German students have on average a more favorable background except for the lowest deciles, but experience much lower returns to these background characteristics in terms of test scores than Finnish students. The results imply that early streaming in Germany penalizes students in lower school types and leads to a greater inequality of educational achievement. It remains unclear, however, if this can be attributed to the effect of school types per se or student background and innate ability that determine the allocation process of students into school types. Overall, the variation in test scores can be explained much better by the observable characteristics in Germany than in Finland.

JEL Klassifikation : I21, H52
Schlüsselwörter : Educational production, PISA, student performance
Download/Reference : http://www.wiwi.uni-konstanz.de/forschergruppewiwi/

PISA: What Makes the Difference?

Explaining the Gap in PISA Test Scores Between Finland and Germany*

Andreas Ammermüller

Centre for European Economic Research (ZEW)

Abstract: The huge difference in the level and variance of student performance in the 2000 PISA study between Finland and Germany motivates this paper. It analyses why Finnish students performed so much better by estimating educational production functions for both countries. The difference in the reading proficiency scores is assigned to different effects, using Oaxaca-Blinder and Juhn-Murphy-Pierce decomposition techniques. The analysis shows that German students have on average a more favorable background except for the lowest deciles, but experience much lower returns to these background characteristics in terms of test scores than Finnish students. The results imply that early streaming in Germany penalizes students in lower school types and leads to a greater inequality of educational achievement. It remains unclear, however, if this can be attributed to the effect of school types per se or student background and innate ability that determine the allocation process of students into school types. Overall, the variation in test scores can be explained much better by the observable characteristics in Germany than in Finland.

* Thanks are due to Charlotte Lauer, Denis Beninger, Peter Jacobebbinghaus, François Laisney, Manfred Weiß and Alfred Garloff and Anja Kuckulenz from the DFG research group "Heterogene Arbeit" for helpful comments. Financial support from the European Commission TSER project "Education and Wage Inequality in Europe" is gratefully acknowledged.

1. Introduction

The publication of the PISA outcomes led to a public outcry in Germany and to envious gazes towards Finland. While Finland achieved the top rank in reading proficiency and the third and fourth place in mathematics and science, respectively, Germany was placed well below the OECD average in all three test subjects (Baumert et al., 2001). Other European countries like Italy, Spain and even Switzerland also performed poorly, at least in some subjects (OECD, 2001). An intense political debate began in response to the negative assessment of their students' performance in the key subjects that have been tested. The debate spread over almost all areas of the political and economic life, as the human capital acquired in a nation's schooling system is generally regarded as the most valuable resource of society. Participants in the debate, be it politicians, teachers or else, often took high performing countries as role models for their own schooling systems. They consider specific characteristics of education systems in the highest scoring countries as a potential means to improve schooling quality at home.[1] The favorite role model in Europe is Finland, due to its high average test scores and their little spread. Especially in Germany, Finland was the country most referred to as an example of an efficient and equitable education system. Moreover, these countries are well suited for a comparison of a streamed (Germany) to a single type schooling system (Finland).

Before reforming the schooling systems, the reasons for the different performance of countries in PISA need to be analyzed thoroughly. Differences in PISA performance concern the level of the average test scores but also the dispersion of the score distribution. The question arises whether the student background are more favorable in the high performing country or if the returns to these background characteristics in terms of test scores are more advantageous? For example, the poor German performance could be due to a higher share of students from a lower social background in Germany than in Finland. However, if the assumed negative effect of a poorer social background were smaller in Germany, the overall impact on the average test score could still be comparable to the impact in Finland. Moreover, the resources and the institutional setting of schools might explain the difference in test scores as well.

Previous studies on student performance in Germany mainly focus on the bivariate correlation between inputs and test scores (Baumert et al., 2001). Another study that employs multivariate methods uses unprocessed data that ignore the problem of missing values and

[1] One example is the German debate on the re-introduction of the comprehensive secondary school like in Finland.

include no information on school types (Fertig, 2003). Dropping students with missing information for some variables is likely to lead to a sample selection bias and neglects the use of the entire set of information that is available for the analysis. Ignoring the school types makes an analysis of the diversified German schooling system almost futile. The multivariate analysis conducted in this paper uses a unique dataset with imputed data for missing values and school type information. The latter has been added from additional data sources. The paper examines the differences between the test score distributions in Finland and Germany and decomposes them in order to quantify the different effects. Thus, the analysis aims at disclosing possible sources of the mediocre performance in Germany and thereby gives guidelines where improvements of the schooling system are most feasible.

The remainder of the paper is structured as follows. The second section introduces the PISA study and describes the data for the two countries of interest. The third section discusses the determinants of educational performance. In the fourth section, the Oaxaca-Blinder and Juhn-Murphy-Pierce decompositions are performed. Finally, the fifth section concludes with a summary of the findings and their political implications.

2. PISA Data

The Programme for International Student Assessment (PISA) tested 15 year-old students in the subjects mathematics, science and reading proficiency in the first half of 2000. The goal was not to test only the knowledge of students but rather their understanding of the subject matter and ability to apply the acquired knowledge to different situations. The testing was conducted by the OECD throughout its 28 member countries plus Brazil, Latvia, Liechtenstein and the Russian Federation. Apart from test scores, data from student, school and computer questionnaires were collected. These include information on the student background, the availability and use of resources as well as the institutional setting at schools (Adams and Wu, 2002). For Germany, additional student-level information on the type of school is taken from an extended version of the PISA study. [2] The two data sources were merged on the student level and then the information was extracted. For a detailed description of the German PISA study see Baumert et al. (2001) and for an analysis of the Finnish results see Välijärvi et al. (2002).

The scores are computed according to the item response theory (cf. Hambleton and Swaminathan, 1989). They are the weighted averages of the correct responses to all questions

[2] In Germany, an extended version of the PISA study was conducted on behalf of the states' education ministers. However, the so-called PISA-E data is not well-suited for a comparison to the Finnish data due to the huge difference in sample size and missing information in the publicly available data-files.

belonging to a certain category, where the difficulty assigned to a question is its weight. The scores have then been standardized, to an international mean of 500 and standard deviation of 100, which facilitates the comparison across countries. These weighted likelihood estimates estimate an individual's proficiency in the respective subject. The values given for the population parameters might slightly differ from other publications (i.e. OECD, 2001), which use plausible values instead that are drawn from an estimated ability distribution and provide better estimates at the population level. The weighted means and standard deviations of the scores and the variables used in the analysis are presented in Table A1 in the annex. Table A2 displays statistics of selected variables separately for the different school types in Germany. The means of the variables show how greatly the characteristics of students and schools vary between the school types so that the observance of school types is a necessity. The standard deviations imply that the variation within school types is also high, except for the school being public or not.

In Finland (Germany), over 5,400 (5,600) students in 155 (219) schools participated in PISA 2000 and completed a reading proficiency and mathematics or science test. Together with the immense background information that is provided, the PISA data are the most recent and detailed data on student performance for the two countries and are moreover internationally comparable.

The data are clustered due to the stratified sampling design of the study. The schools that participated have been chosen first, before a random sample of the student target population was drawn. Therefore, the schools are the primary sampling units and not the students.

The main problem of the data are missing values for the over 100 student and school background variables. For Germany and Finland, up to 16 percent of key variables such as parents' education are missing.[3] Commonly, the whole observation (student) is dropped from the regression whenever the value of any explanatory variable is missing. Including many variables in the regression thus leads to a great reduction in the number of observations that can be used for the estimations. The imputation leads to an increase of usable observations of 31% (39%) in Finland (Germany). As these numbers are roughly comparable, the imputation is unlikely to introduce a bias in this cross-country comparison. Apart from losing valuable information, dropping students with incomplete answers to the questionnaires leads to a sample selection bias if the values are not missing randomly. Indeed, given that attentive students are more likely to both complete the questionnaire and to answer the test questions,

[3] Table A3 presents the percentage of missing values.

low performing students have a higher probability of being dropped. Thus, dropping the observations with missing values would lead to an upward bias in the test scores, which can be seen in Table A1.

The approach chosen here to overcome the problem of missing data is to predict missing values on the basis of regressions on those background variables like age, sex and the grade a student is in that are available for all students. Linear models are used for continuous variables and probit and ordered probit models for qualitative variables. Students who did not answer these elementary background questions or did not complete the tests are excluded from the regressions, as well as students with more than 10 missing values.[4] This applies to less than one percent of the sample but leads to a significant increase in mean test scores and a lower standard deviation in Germany. The exclusion of outliers is necessary so that the analysis is not dominated by a small and unrepresentative subsample of the student population. The descriptive statistics and the regression results are also given for the original data without imputed values in Tables A1 and A2, respectively, where all students with at least one missing value are dropped.

The prediction of missing values on the basis of regression results is clearly no impeccable solution. The variation of the variables decreases, as can be seen in the lower standard deviations of the variables including the imputed values as compared to the original data. However, the imputed values vary greatly as well and the information of the non-imputed values of the observation is not lost.

2.1 Distribution of Test Scores

In this part, the distributions of test scores for Finland and Germany will be presented graphically. For each subject, non-parametric kernel density estimates describe the score distribution of the two countries.[5]

Figure 1 displays the test score distributions for the three subjects that have been tested for both Finland (FIN) and Germany (GER). The Finnish scores are on average higher than the German scores, which can be seen in the more right position of the Finnish distribution and the higher weighted average score. The average test scores and their standard deviations as well as the minimum and maximum value for each variable are presented in Table A1 in the Appendix. The peak of the Finnish distribution is also clearly to the right of the German distribution, which reflects the higher mode of the kernel density estimates. Moreover,

[4] Moreover, students with an unrealistically low score of below 200 points (26 students) and students from one school in Finland with identical test scores (5 students) were dropped from the regressions.
[5] For a description of the employed kernel function, see Appendix B.

Finland has not only more good students but especially fewer low and very low performing students than Germany, which has a relatively fat left tail. Despite the higher average scores, Finland has a lower standard deviation of scores. This pattern holds for all three subjects in which the students have been tested.

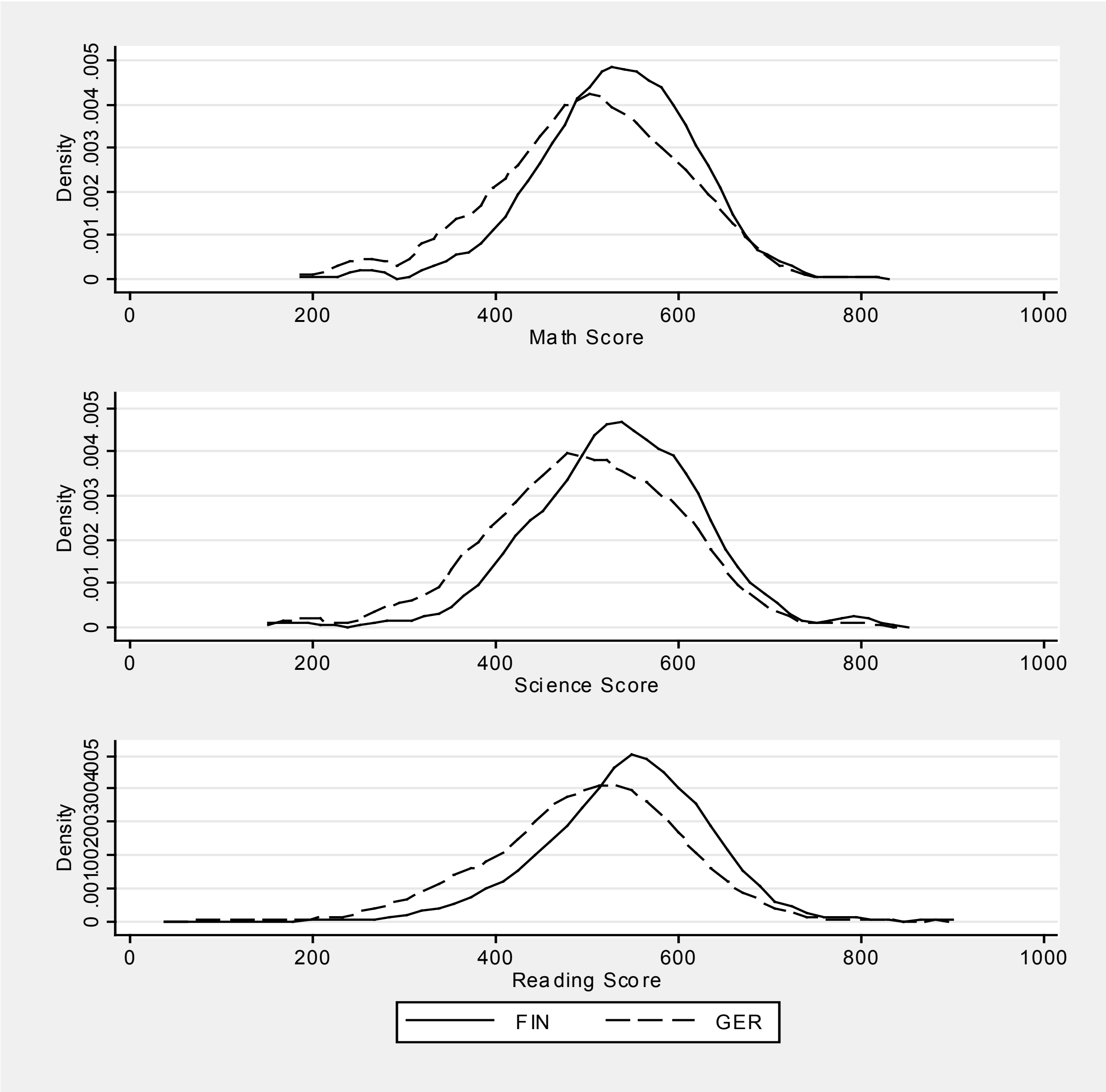

Figure 1: Test Score Distributions

Finland exhibits more desirable characteristics in its test score distributions, namely higher average scores, a higher mode, a lesser spread of test scores and especially fewer very low performing students in all subjects tested. The question therefore arises if the reasons for this great difference in performance to German students can be identified. The subsequent analysis focuses on the reading proficiency of students because the respective test scores are available for all students. The distributions of the other test scores suggest no important difference between the subjects, in which the students were tested. In the next step, the factors that affect the test scores in either country are analyzed.

3. Determinants of Reading Proficiency Scores

3.1 The Production Function Approach

A thorough comparison of student performance in the schooling systems of the two countries presupposes the knowledge of the process by which education is produced. Educational production functions provide a means for understanding the production process by estimating the effects that various inputs have on student achievement. For the production function to yield unbiased estimates of the effects, all current and prior inputs into the education system that are likely to determine educational performance should be included in the production function. The cross-sectional PISA data give information on the background of each student, the current resources including teacher characteristics as well as the institutional setting at the school level. However, no information on prior achievement of students or inputs into educational production at another time are available. Therefore, the estimation of educational production functions is limited by missing information over prior inputs (Todd and Wolpin, 2003). The coefficients of the following model of an educational production function can only be interpreted as causal effects under certain assumptions:

$$(1) \qquad T_{is} = \beta_0 + B_{is}\beta_1 + R_s\beta_2 + I_s\beta_3 + S_S\beta_4 + \upsilon_S + \varepsilon_{is}$$

where T_{is} is the test score of student i in school s, B_{is} is a set of student background variables, R_s comprises the variables on the resources employed at school, I_s represents the institutional variables, S_s the school type variables and ν_s and ε_{is} are the error terms at the school and student level. The groups of parameters β_0 to β_4 are to be estimated and measure the impact of the variables on educational achievement.

Besides innate ability, which cannot be measured, the background of a student has been shown to be the most decisive factor explaining student performance (cf. Hanushek and Luque, 2002; Woessmann, 2000). The background B_{is} includes besides personal characteristics like sex and age also information on the parents' origin and education. These variables are unlikely to change over time and are hence a good proxy also for prior inputs. Its effect on the cognitive achievement of students can therefore be interpreted as a causal relationship. However, the total effect of student background on student performance is underestimated by β_1 if there is an indirect effect via the school type S_s. This is the case when the allocation of students to school types depends not only on innate ability but also on parental background. Therefore, the coefficients of student background should better be interpreted as the lowest boundary of the effects, especially for Germany with its many schooling types.

The current resources R_s that might affect student performance describe parts of the schooling system that depend mainly on the financial investments from the public side. The student/teacher ratio at the school level is used to measure the input of teachers for each student. Instead, actual class size would have the advantage of measuring this input more directly, but the estimate of class size is likely to be biased. For the class size estimate, selection of students within schools adds to the problem of selection between schools (cf. West and Woessmann, 2003). Indeed, weak students might be put in smaller classes in order to foster their learning. School type dummies can only control for selection between schools. Moreover, under the assumption that students do not switch schools and that the student/teacher ratio is roughly constant over time, the current student/teacher ratio is a reasonably good proxy for teacher input per student over the last years.

The institutional setting is the framework within which the different players involved in schools act. It may affect the motivation and incentives, especially of students and teachers.[6] The variables describe the distribution of responsibility and other institutional aspects. As institutional reforms take a long time to be implemented, the current institutional setting should accurately describe the setting over the last years, assuming that students stay in the same school. In Finland, students usually stay in the same comprehensive school over the entire period of compulsory education of nine years while German students commonly change school after four years of elementary school.[7] Given that the tested students have already spent four to five years at their secondary school, the effect of the former elementary school should be negligible. Under the mentioned assumptions, the coefficients for resources and the institutional setting of the school can be interpreted as causal effects, especially since we control for the school type.

Finally, the school type variables indicate the type of schools a student attends, which can be five types in Germany and only one in Finland. German students are allocated to secondary school types after their fourth school year according to their performance in

---

[6] For a theoretical discussion of institutional effects, see Bishop and Woessmann (2002).
[7] Information on the educational systems are taken from Eurybase (2003).

elementary school.[8] Assuming that innate ability of students in the fourth and eighth/ninth school year is not independent, there is a problem of endogeneity between school type and student performance because both are determined by innate ability. Moreover, as educational performance in elementary school, the preference for school types and thus the allocation to a school type are also determined by student background, the school type coefficient might include a part of the student background effect on student achievement. Hence, the school type coefficient consists of the 'true' school type effect, an effect of sorting by innate ability and an additional impact of student background on student performance via school type. The coefficient β_4 can therefore only be interpreted as a partial correlation. All explanatory variables and their descriptive statistics are presented in Table A1 in the Appendix.

3.2 The Estimated Effects

The effect of the characteristics on student performance is estimated in a regression of the explanatory variables on the individual student test score (see equation (1)).[9] Due to the clustered design of the PISA data, survey regressions are used for the estimation. These correct the standard errors for the clustered data design, which implies an interdependence of error terms between students within the same school. As the students of different schools and countries have different sampling probabilities, the sampling weights available in the data are used for the estimation of model (1). The outcomes of the weighted survey regressions with the dependent variable reading proficiency scores are presented in Table A3. Using the data with instead of without the imputed values does not affect the qualitative interpretation of the results but makes them more representative of the student population.

The R^2 of the regressions indicate that more than half of the variation in the German test scores can be explained but only 17 percent of the Finnish variation. The performance of students in Germany depends therefore more highly on conditions that have been controlled for and less on innate ability and other unobserved factors than in Finland. The student background variables are highly significant and have a high impact on student performance.[10] For example, students whose parents do not even have completed secondary education score 37 points lower in Germany, respectively 27 points lower in Finland compared to students whose parents have completed tertiary education, all else equal. The penalty for an unfavorable student background is higher in Germany than in Finland for this example. Girls

[8] Teachers at elementary school write recommendations for each student, then parents have to apply at schools. Only the degree from the higher secondary school (Gymnasium) allows to follow university. The vocational school (Berufsschule) is for students in an apprenticeship.

[9] Characteristics here imply all measurable characteristics, including student background, resources, institutional setting and school types.

perform significantly better than boys in Germany and especially in Finland. Being in the ninth instead of eighth grade raises student performance significantly, especially in Finland. Students who were born abroad or whose parents immigrated score lower than comparable non-immigrated students, especially in Finland where the share of these students is only three percent compared to 20 percent in Germany. The number of books at a student's home has a highly significant and large effect on performance.

The effect of resources is limited and never significant. A high share of low educated teachers leads to a non-significant decline in student test scores in Germany and fewer instruction time decreases the scores in Finland. The student/teacher ratio and lack of material seem not to be significantly related to test scores, either. The variables describing the institutional setting are not significant except for the lacking power of schools to select their students in Germany. The variation of the institutional setting within countries is not very large however, so that inter-country comparisons are more suited for analyzing their effects (cf. Woessmann, 2000). The type of school exhibits highly significant effects in Germany in reference to comprehensive schools, except for vocational schools. Students who attend a low (high) secondary school score 51 (88) points lower (higher) on average than comparable students in a comprehensive school in Germany.

After having shown the determinants of student performance in the two countries, the following section compares the results more systematically by decomposing the score gap between Finland and Germany into different components.

4. Explaining the Test Score Gap

The difference observed between the test score distributions in Finland and Germany may be due to several reasons. First, Finnish students may have a more favorable endowment in characteristics measured by the explanatory variables. Finnish students might for example have better educated parents, who exert a positive influence on the performance of their children. Besides the family background of students, resources at schools and their institutional setting might differ, too. Relatively more and better educated teaching staff and a school's responsibility over the budget could also explain a better performance of Finnish students relative to Germans, if this is shown to have a positive impact on test scores.

Second, the effects of the different characteristics on the performance of students might differ between the two countries. In other words, the same characteristics might be less efficient in producing education in Germany than in Finland. A greater return to family

[10] This is confirmed by the marginal effects.

background characteristics in terms of test scores would imply a higher social differentiation of students. The educational achievement is then predetermined to a higher degree by the family of a student and students from lower social backgrounds find it harder to perform well at school. The effects of the resources and institutions instead display the ability of schools to transform their endowment and their responsibility into improved student test scores.

Third, a part of the test score gap is due to the difference in the residuals of the estimated regressions. Any unobserved factors that affect skills, foremost innate ability of students and their motivation, constitute the residual effect. As the expected value of the residuals is zero, the residual effect is only important when we consider the test score gap at other points of the score distribution than the mean.

These three effects, referred to as the characteristics, the return and residual effect can be quantified by decomposition methods. Two different methods will be employed: The Oaxaca-Blinder (section 4.1) and the Juhn-Murphy-Pierce (section 4.2) decomposition.

4.1 Oaxaca-Blinder Decomposition

This ‘classical’ decomposition technique has been developed by Blinder (1973) and Oaxaca (1973) and splits a gap into two parts. The first part is explained by the differences in the characteristics, the second by the differences in the effects of those characteristics that have been estimated in the regressions. However, the technique considers only the average effects, ignoring differences along the distribution like its dispersion and skewness. The latter aspect will be examined in section 4.2.

The decomposition method used here differs slightly from the classical Oaxaca-Blinder decomposition and follows Lauer (2000). As the aim of the analysis is to explain the low performance of German relative to Finnish students, the different effects that explain the score difference are considered from the point of view of German students.

The total score gap between Finland and Germany at the mean is defined as

$$(4) \qquad \Delta T = \overline{T}^{F} - \overline{T}^{G}$$

where the bars denote averages and the superscripts F and G the countries Finland and Germany, respectively. The total score gap can then be decomposed into a characteristics, a return and a characteristics-return effect.

(5)

$$\Delta T = \hat{\beta}_1^{G}(\overline{B}^F - \overline{B}^G) + \hat{\beta}^{G}{}_2(\overline{R}^F - \overline{R}^G) + \hat{\beta}^{G}{}_3(\overline{I}^F - \overline{I}^G) + \hat{\beta}^{G}{}_4(\overline{S}^F - \overline{S}^G)$$
$$+ (\beta_0^F - \beta_0^G) + (\hat{\beta}_1^F - \hat{\beta}_1^G)\overline{B}^G + (\hat{\beta}_2^F - \hat{\beta}_2^G)\overline{R}^G + (\hat{\beta}_3^F - \hat{\beta}_3^G)\overline{I}^G + (\hat{\beta}_4^F - \hat{\beta}_4^G)\overline{S}^G$$
$$+ (\hat{\beta}_1^F - \hat{\beta}_1^G)(\overline{B}^F - \overline{B}^G) + (\hat{\beta}_2^F - \hat{\beta}_2^G)(\overline{R}^F - \overline{R}^G) + (\hat{\beta}_3^F - \hat{\beta}_3^G)(\overline{I}^F - \overline{I}^G) + (\hat{\beta}_4^F - \hat{\beta}_4^G)(\overline{S}^F - \overline{S}^G)$$

The characteristics effect, which is displayed separately for each category of explanatory variables in the first line of equation (5), measures how much German students would score differently if, given their estimated returns to characteristics in terms of scores, they had the same characteristics as the Finnish students. The second component, the return effect, shows how much German students would hypothetically be better, if they experienced the same production process of schooling, i.e. the same transformation of inputs into educational achievement as the Finnish students, given their own characteristics. The final characteristics-return effect is an interaction between the impact of a possibly better production process and different characteristics in Finland.

The gap between the weighted average reading scores amounts to 54.28 points, as can be seen in Table 1. This difference is substantial since it is more than half of the international standard deviation of the PISA scores and is 45 percent higher than the effect of being in the 9th instead of 8th grade in Germany. The total characteristics effect is negative, implying that the German characteristics are actually more advantageous than the Finnish ones. The overall return effect of 63.45 points is highly positive and seems to explain the score gap. The transformation of given inputs in Finnish schools results in higher student performance than in Germany. The interaction effect is rather small with 16.04 points.

A separation of the effects into the four groups of explanatory variables, student background, resources, institutions and school types, shows a more differentiated picture. While the characteristics effect for student background is negligible, the effect of resources explains about 8 percent of the positive gap. Instead, the negative effects for institutions and school types imply more favorable characteristics for German schools.

The highly positive return effect is driven by the resource variables and the difference in the intercepts. Resources are transformed more efficiently into student performance in Finland than in Germany. This effect depends mainly on the effect of the share of low educated teachers, which is negative in Germany and positive in Finland. Instead, the transformation of the personal and family characteristics is more beneficial for German than for Finnish students and almost offsets the positive effect for resources and the difference in the intercepts.

The interaction effect shows that the interaction between better characteristics and a better production process benefits Finnish students relative to German students for all categories of variables except for resources.

	sum	St. Backgr.	Resour.	Institut.	Schools	Interc.
Total gap	54.28					
Charact. Effect	-25.20	-0.07	4.25	-8.73	-20.66	
Ratios	-0.46	0.00	0.08	-0.16	-0.38	
Return Effect	63.45	-224.17	60.00	-1.34	-20.66	249.61
Ratios	1.17	-4.13	1.11	-0.02	-0.38	4.60
Interaction effect	16.04	5.61	-13.34	3.10	20.66	
Ratios	0.30	0.10	-0.25	0.06	0.38	

Absolute effects in reading scores. Ratios are effects divided by total score gap.

Table 1: Decomposition for Reading Scores for all coefficients

Table 1 considers all coefficients for the decomposition, even when the difference between coefficients in the two countries is not statistically significant at a reasonable level. Table A4 presents the decomposition results when the coefficients that do not differ at the 10 percent-significance-level are restricted to be equal.[11] Therefore, only effects that significantly differ between countries are taken into account.[12] The effects in Table A4 differ only greatly for the resource and institutional variables, for which no variables differ significantly between the two countries. The sum of the effects hardly changes.

When the average of the distribution is considered, the difference in the characteristics cannot explain the better performance of Finnish students. According to the decomposition, the poor transformation of the available resources and the difference in the unobservables in the intercepts is causing the relatively low scores of German students.

4.2 Juhn-Murphy-Pierce Decomposition

Until now, only the mean of the distribution has been considered. However, the distribution of scores differs between the two countries, as has been shown in section 2. Therefore, the decomposition will be performed along the entire score distribution as well.

The following decomposition technique was first employed by Juhn et al. (1993) for a decomposition of change across time. It is also applicable to cross-section data (e.g. Blau and Kahn, 1992), like the PISA data. The method has the distinct advantage of considering not

[11] The effects for the two countries have been estimated simultaneously using interaction terms to see if the coefficients for the countries differ significantly. The interaction terms that are not significant have then been dropped. Reducing the significance level to five percent leads to a further reduction of considered coefficients.

only the mean for the decomposition but the whole distribution. Moreover, it deals explicitly with the residuals from the estimation of the production function, which are equal to zero at the mean but not at specific quantiles. Following a slightly different approach, it allows one to decompose the score gap into a characteristics, return, characteristics-return and residual effect.

The residual ε_i of country y can be thought of consisting of two components: the percentile of an individual i in the residual distribution θ_i, and the distribution function of the residuals, F_i. The inverse cumulative residual distribution function then gives us:

(6) $$\varepsilon_i^y = F^{y-1}(\theta_i^y \mid X_i^y) \ ,$$

where X comprises the four sets of explanatory variables B, R, I and S of country y. Using the estimates from unweighted survey regressions of model (1), the actual and two hypothetical test score distributions for German students can be constructed:

(7) $$GER_i = \widehat{\beta}^G X_i^G + F^{G^{-1}}(\theta_i^G \mid X_i^G)$$

(8) $$GER(1)_i = \widehat{\beta}^F X_i^G + F^{F^{-1}}(\theta_i^G \mid X_i^G)$$

(9) $$GER(2)_i = \widehat{\beta}^G X_i^G + F^{F^{-1}}(\theta_i^G \mid X_i^G)$$

The first hypothetical distribution GER(1) shows what scores German students would attain if they experienced the Finnish production process and the corresponding residuals from the Finnish residual distribution. Equation (9) presents the second hypothetical distribution GER(2), which assumes that the characteristics of German students are transformed into test scores by the German returns, but that the residual distribution is the same as for Finnish students. The two hypothetical Finnish distributions are created likewise.

Following the decomposition as described in equation (5), the characteristics effect is the difference between the test score distributions for FIN(1) and GER. The return effect equals the difference between the two hypothetical test score functions GER(1) and GER(2). . The third effect is due to the different distribution of residuals in the two countries and can be calculated by subtracting GER from GER(2). The interaction effect can be constructed as (FIN-FIN(1))-(GER(1)-GER). Adding up all four effects leads to the total gap (FIN-GER)

[12] These variables are: student age, sex, parents' origin, parents' higher sec. education and all school types except for vocational schools.

that shall be explained here. The resulting score distributions are shown in the following graphical representation.

4.2.1 The Hypothetical Score Distributions

The total reading score gap between Finland and Germany is shown in Figure 2. The gap is declining along the deciles of the score distribution. While it is over 60 points for the lowest performing decile of students, it is 27 points for the best performing 10 percent of students. The relatively bad performance of the lower part of the German student distribution seems to be mainly responsible for the low average score compared to Finland, although also the best German students do not attain the same level as the best Finnish students. The inequality in the test score distribution is hence much higher in Germany than in Finland.

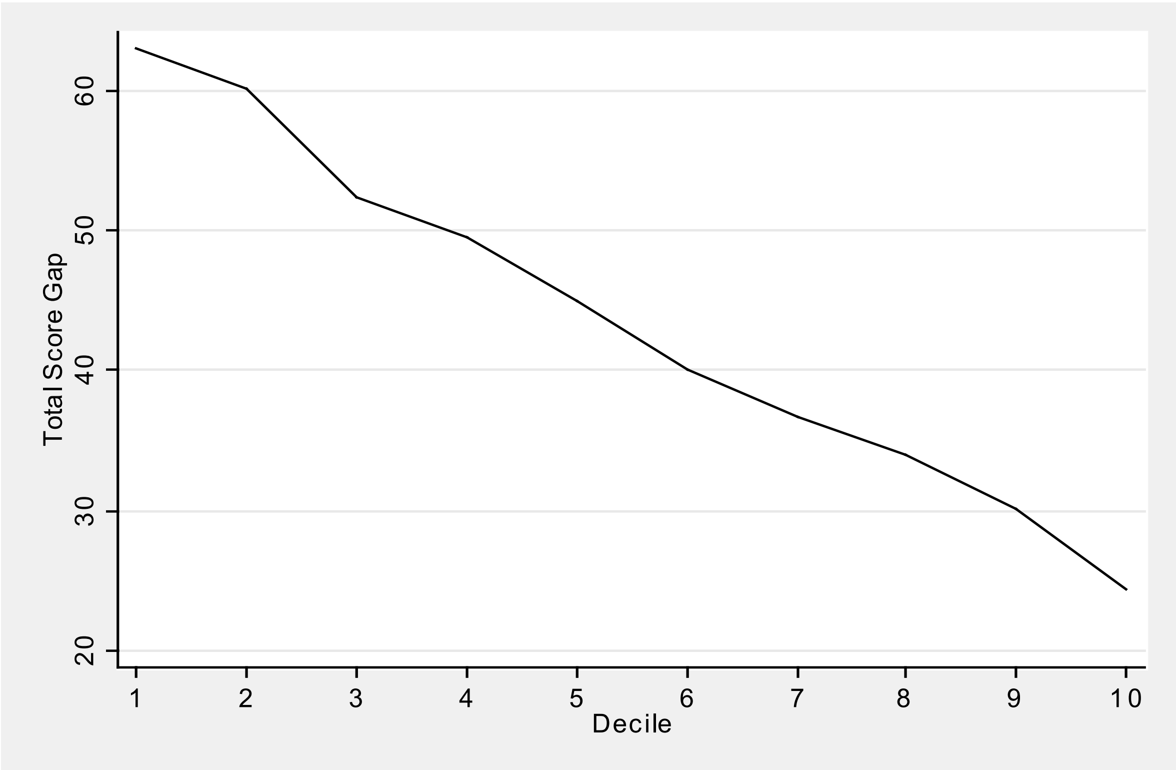

Figure 2: Total Score Gap

In order to show the different effects graphically, Figure A1 in the Appendix displays the real and hypothetical test score distributions. The difference between the reading score distributions as estimated by the kernel density function are only due to a respective effect. The first effect is the characteristics effect. The graph shows the hypothetical Finnish distribution FIN(1) and the actual German distribution GER. The former one displays how the distribution would look when Finnish students with their own characteristics would experience the German returns to these characteristics and the German residuals given their position in the Finnish residual distribution. The difference that remains between the two distributions is only due to the difference in characteristics between the two countries, given the German educational production process. The mode of the hypothetical Finnish distribution is positioned to the left of the German distribution, which has a higher spread and is slightly skewed to the left. This implies that most German students actually have the more favorable (according to the estimation results) and heterogeneous characteristics of the two student samples. The characteristics effect thus implies higher average test scores for German than for Finnish students. However, in the lower part of the distribution the size of the effect decreases and implies higher scores for Finnish than for German students. This is consistent with the slope of the total score gap over the distribution but contradicts the positive sign of the Finnish-German score gap.

The return effect is shown in the next figure. The hypothetical distribution GER(1) shows the predicted scores for German students that experience the Finnish production process including the Finnish residuals. Distribution GER(2) displays how German students in German schools would perform if they had the Finnish residuals. The difference between the distributions is only due to differences in the production of education in the two countries, given the German students' characteristics. The production process in Finnish schools clearly leads to a better performance of students, especially for the lower part of the distribution, than the one in German schools. The return effect can hence explain why German students are performing worse than Finnish students.

The residual effect is depicted in the third graph, where the distributions GER(2) and GER are compared. The Finnish residuals in GER(2) actually lead to a wider distribution than the German residuals, which are quite dense. This is consistent with the earlier results on the determinants of test scores, which showed that the observable factors can explain a higher share of the variation in test scores in Germany than in Finland. Consequently, unobservable factors like the innate ability of students have a greater effect in Finland, which is implied by the residual effect.

The last effect in Figure A1 is the interaction effect, showing the interaction between the possibly better production process and characteristics of Finnish students and schools. The effect is positive but can be explained more clearly in the following section.

The hypothetical distributions showed the predicted test scores for German and Finnish students had they experienced another educational system. In the following section, a closer look is paid to the different effects and the contribution of the different groups of variables to the effects.

4.2.2 The Effects and their Components

The four effects, which all contribute to the total score gap between the countries, can be broken down further and linked directly to the four groups of variables that determine student performance. First, the course of the aggregated effects is shown over the deciles of the test score distribution. Figure 3 displays the total score gap and the absolute effects while the relative effects, which are divided by the total score gap at each decile of the test score distribution, are shown in Figure A2.

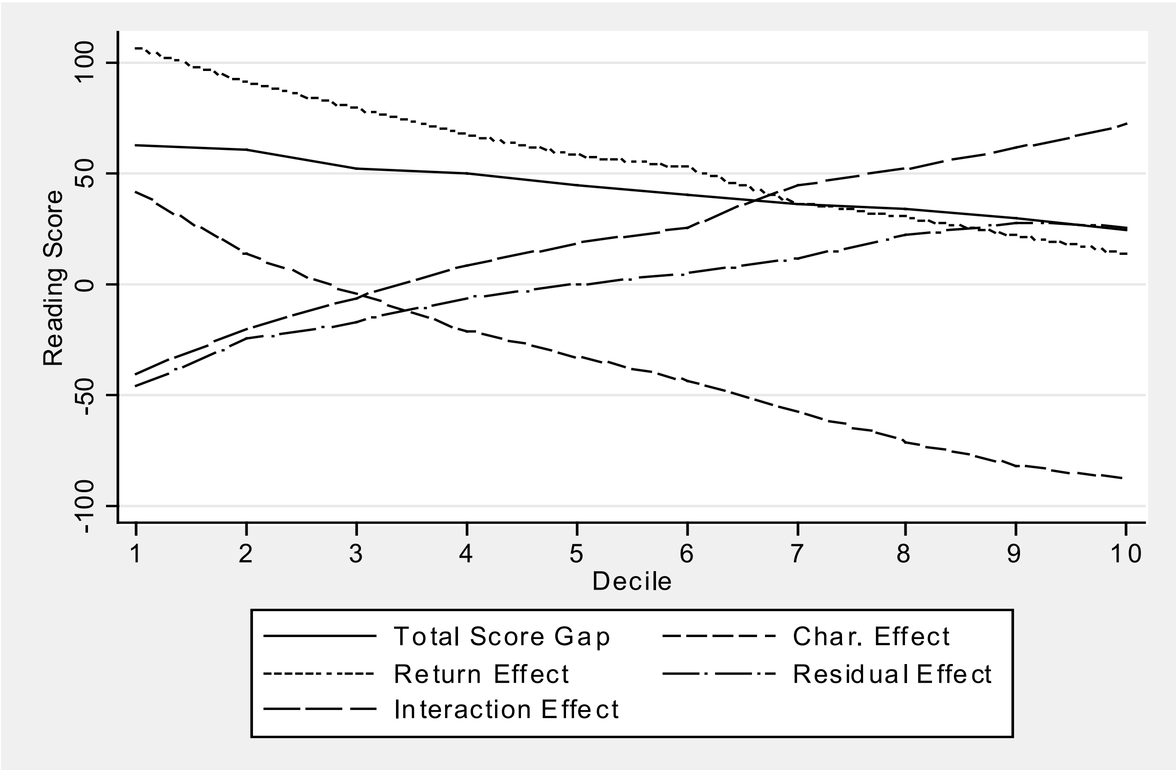

Figure 3: Absolute Effects

The characteristics effect can explain a part of the test score gap only for the lowest three deciles of the test score distribution, where it is positive. It decreases steadily, implying that the characteristics of German students are deteriorating comparatively more when going down the score distribution and thus that the inequality in characteristics is larger in Germany than in Finland. The return effect decreases as well over the distribution, but is always positive and mostly higher than the total score gap, which decreases not as fast. The problem of converting the given endowments in Germany into good performance of students is thus greatest for the weakest students. The residual and interaction effect run almost identically to each other and opposing to the other effects. They increase over the whole distribution and are positive from about the 4th decile upwards. The increase in the residual effect is caused by a steeper rise in the Finnish residuals that are first smaller and then higher than the German residuals. This implies that unobservable factors explain more of the variation in test scores in Finland, which can also be seen in the third part of Figure A1, since students at the bottom of the distribution have lower residuals and students at the top have higher residuals than the corresponding German students. The interaction effect increases because the German characteristics deteriorate more when moving down the score distribution.[13]

Now we turn to the composition of the effects. Figure 4 displays the four components of the characteristics effect. While the positive effect of resources and the negative effect for institutions does not vary greatly over the distribution, the effect of student background decreases and turns from positive for the lowest three deciles to negative.[14] Weaker students in Germany have hence less and high performing students more favorable characteristics than Finnish students at the same decile of the score distribution. The characteristics effect for the variables describing the type of school is also positive for the lower part of the distribution but decreases very strongly up to –70 for the highest decile. The streaming of the schooling system is hence associated with a greater inequality between students in Germany than in Finland because it introduces an additional source of test score variation but cannot explain the score gap between the countries.[15]

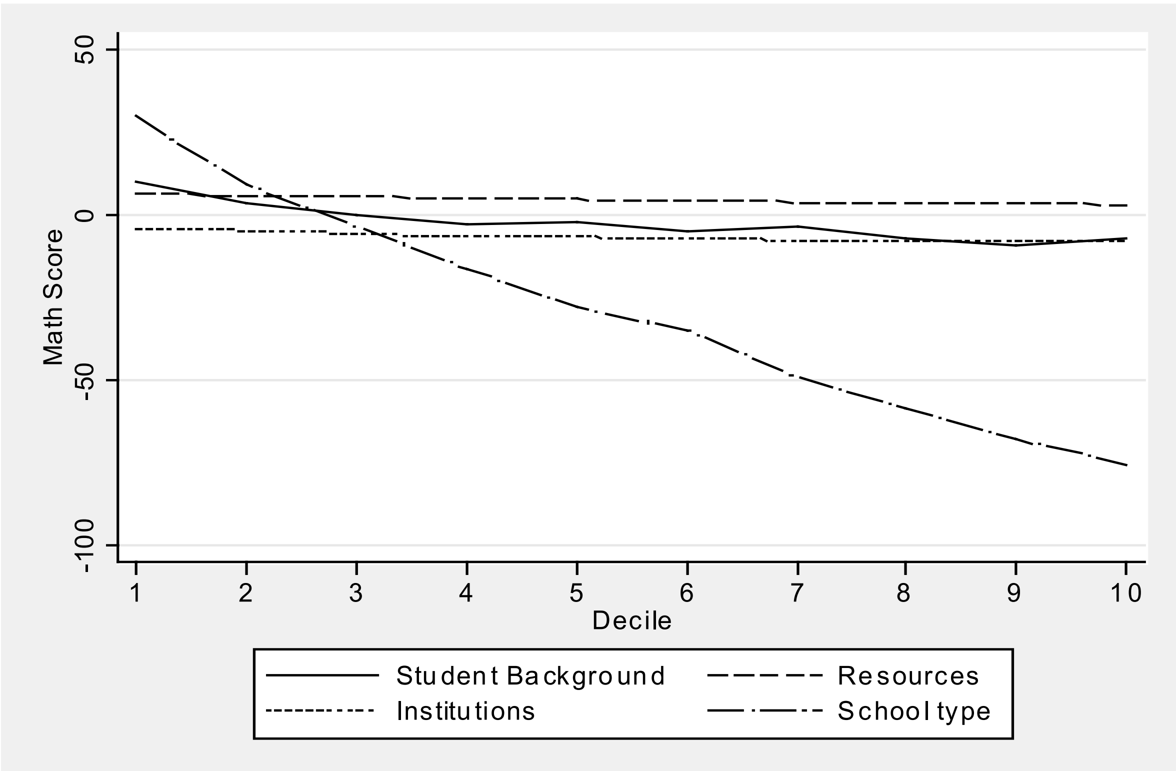

Figure 4: The components of the characteristics effect

[13] Since almost all returns are negative, a greater inequality is represented by faster decreasing endowments along the distribution in Germany than in Finland.
[14] The student background variables that change the most along the score distribution are parents' education and books at home and less the personal characteristics like student's sex and age.
[15] For the difference in test scores between the school types in Germany see Table A2.

Figure 5 shows the return effect separately for each group of variables. The highly negative effect for student background variables implies that these characteristics are transformed into higher test scores in Germany than in Finland. This is mostly due to the larger negative coefficients for student's sex and parent's origin in Finland. Moreover, it remains unclear how much of the effect of student background on performance is hidden in the school type variables in Germany (see section 3.1). The resources are used more efficiently in Finland than in Germany along the whole score distribution and can hence explain the score gap partly. Institutional variables contribute slightly negatively to the return effect. The return effect for school type variables decreases along the score distribution and is mostly negative. Since Finnish students are all in comprehensive schools, the effect reflects only that students in the lowest three deciles are more likely to attend low secondary schools (Hauptschule) in Germany which have a negative effect compared to the reference group of comprehensive schools, while higher performing students are more likely to be in medium (Realschule) or higher (Gymnasium) secondary schools. The difference in the intercept is highly positive, implying that the level of test scores is higher in Finland than in Germany due to unobserved factors.

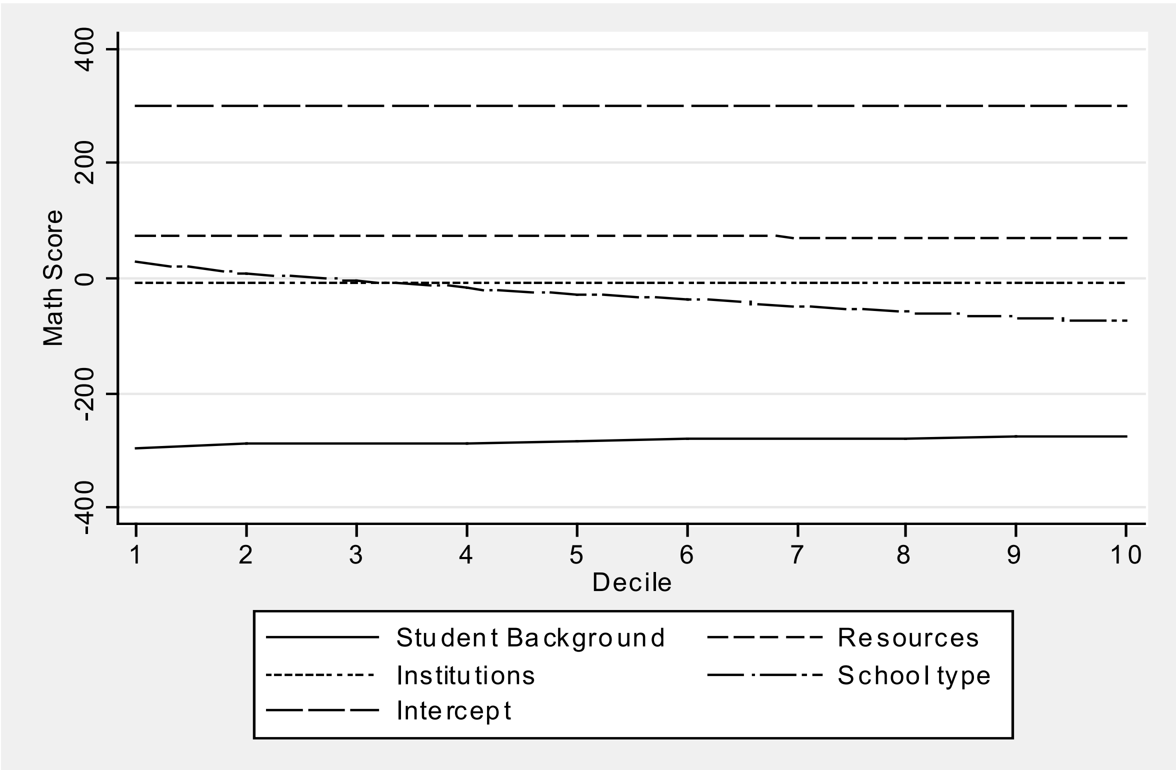

Figure 5: The components of the return effect

Student characteristics are transformed much more favorably in Germany, while resources are used more effectively in Finland. The streamed German schooling system penalizes especially students attending a lower type of school, thus contributing to the large variation in test scores.

A possible further step in the decomposition analysis would be to consider the estimated coefficients along the conditional distribution, not only at the mean of the distribution. However, the coefficients estimated by quantile regressions do not differ significantly from OLS coefficients, only for very few coefficients and some quantiles. Thus, a decomposition using quantile regressions does not seem to add relevant insights and is therefore not conducted.

5. Conclusion

The decomposition analysis showed that the poor performance of German students compared to Finnish students is not due to a less favorable student background, except for the bottom of the score distribution. German students have on average more favorable characteristics but experience much lower returns to these characteristics in terms of test scores than Finnish students. The background of German students changes much faster along the score distribution, which explains the higher inequality in Germany. The institutional setting seems to be more favorable in Germany while Finland is endowed with slightly more resources. The characteristics of students are transformed into higher test scores in Germany than in Finland once their effect on school choice is neglected. Instead, resources are used more efficiently in Finland, where teachers are more highly educated and a lower education of teachers has no negative effect on student performance. A large part of the overall score gap between the countries is due to unobservable factors. The results also imply that streaming in Germany penalizes students in lower school types and leads to a greater inequality of educational achievement. It remains unclear, however, if this can be attributed to the effect of school types per se or student background and innate ability that determine the allocation process of students into school types. Overall, the variation in test scores can be explained much better by the observable characteristics in Germany than in Finland.

In order to improve the performance of students in Germany, especially the educational achievement of students in the lower part of the test score distribution has to be promoted. These students suffer from a highly disadvantaged student background, whose negative impact upon performance might be magnified by the early streaming in the German schooling system at the age of ten. They are not given the chance to compensate for their background before they are divided into different school types. The measured resources, especially the education of teachers, must be employed more efficiently in order to close the gap to leading countries in student performance. There is no evidence for a beneficial effect of lower student teacher ratios but a higher education of teachers seems to benefit students in Germany.

Further research is needed on the effects of school types in educational production functions, which should try to isolate the ‘true’ effect of school type on educational achievement. Only then the determinants of educational achievement can be precisely estimated for schooling systems that massively use streaming.

References

Adams, R. and Wu, M. (2002), *PISA 2000 Technical Report*, OECD, Paris.

Bishop, J. and L. Woessmann (2002), Institutional Effects in a Simple Model of Educational Production, IZA Discussion Paper No. 484.

Blau, F. and L. Kahn (1992), The Gender Earnings Gap: Learning from International Comparisons, *The American Economic Review* Vol. 82(2), Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association, 533-538.

Blinder, A. (1973), Wage Discrimination: Reduced Form and Structural Estimates, *The Journal of Human Resources* Vol. 8(4), 436-455.

Baumert, J. und Deutsches PISA-Konsortium (2001), *PISA 2000 : Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*.

Eurybase (2003), The Information Network on Education in Europe, http://www.eurydice.org, visited October 24th 2003.

Fertig, M. (2003), Who's to Blame? The Determinants of German Student's Achievement in the PISA 2000 Study, IZA Discussion Paper No. 739.

Hambleton, R. and H. Swaminathan (1989), *Item Response Theory. Principles and Applications*, Boston: Kluwer.

Hanushek, E. and J. Luque (2002), Efficiency and Equity in Schools Around the World, NBER Working Papers 8949.

Juhn, C., K. Murphy and B. Pierce (1993), Wage Inequality and the Rise in Returns to Skill, *The Journal of Political Economy* Vol. 101(3), 410-442.

Lauer, C. (2000), Gender Wage Gap in West Germany: How Far Do Gender Differences in Human Capital Matter?, ZEW Discussion Paper No. 00-07, Mannheim.

Oaxaca, R. (1973), Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review* Vol. 14 (3), 693-709.

OECD (2001), *Knowledge and Skills for Life: First Results from PISA 2000*, OECD, Paris.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

Todd, P. and K. Wolpin (2003), On The Specification and Estimation of The Production Function for Cognitive Achievement, *The Economic Journal* Vol. 113(485).

Välijärvi, J., P. Linnakylä, P. Kupari, P. Reinikainen and I. Arffman (2002), *The Finnish Success in PISA – and some Reasons behind it*, http://www.jyu.fi/ktl/pisa/publicationl.pdf.

West, M. and L. Woessmann (2003), Which School Systems Sort Weaker Students into Smaller Classes? International Evidence, IZA Discussion Papers 744.

Woessmann, L. (2000), Schooling Resources, Educational Institutions, and Student Performance: The International Evidence, Working Papers 983, Kiel Institute of World Economics.

Appendix A

Table A1: Weighted means and (standard deviations) for reading

	With imp. values		Without imp. val.		Min	Max	description
	GER	FIN	GER	FIN			
Reading Score	490.86 (102.85)	545.15 (87.27)	509.81 (94.98)	547.71 (86.44)	206.93	887.31	Warm estimate of reading test score
Student Backgr.							
Student's age	188.44 (3.37)	187.56 (3.42)	188.46 (3.40)	187.55 (3.42)	182	194	Student's age in month
Student's sex	.50 (.50)	.49 (.50)	.50 (.50)	.48 (.50)	0 for female	1 for male	Sex of students
8th grade	.14 (.35)	.11 (.31)	.12 (.32)	.10 (.31)	0 for 9th grade	1 for 8th grade	Grade level of students
Parents‘ Origin	.20 (.40)	.03 (.18)	.14 (.35)	.03 (.16)	0	1 if parent foreign	Parents' place of birth
Parents no sec. Ed.	.01 (.12)	.09 (.29)	.02 (.13)	.10 (.30)	0	1 for less than Sec.	Highest educational level reached by a parent
Secondary Ed. 2	.09 (.29)	.10 (.30)	.06 (.24)	.11 (.31)	0	1 for fin. Lower Second.	
Secondary Ed. 3	.52 (.50)	.41 (.49)	.47 (.50)	.41 (.49)	0	1 for fin.upper Second.	
Tertiary Ed. (Ref.)	.38 (.48)	.40 (.49)	.45 (.50)	.38 (.48)	0	1 for fin. Univers.	
Books Cat. 1	.01 (.11)	.01 (.08)	.01 (.10)	.01 (.08)	0	1	No books at students home
Books Cat. 2	.06 (.25)	.07 (.25)	.04 (.21)	.07 (.25)	0	1	1-10
Books Cat. 3	.20 (.40)	.23 (.42)	.18 (.39)	.23 (.42)	0	1	11-50
Books Cat. 4	.23 (.42)	.24 (.43)	.23 (.42)	.24 (.43)	0	1	51-100
Books Cat. 5	.21 (.41)	.25 (.43)	.22 (.41)	.25 (.43)	0	1	101-250
Books Cat. 6	.15 (.36)	.14 (.35)	.17 (.38)	.14 (.35)	0	1	251-500
Books Cat. 7 (Ref.)	.13 (.33)	.06 (.25)	.15 (.35)	.06 (.24)	0	1	More than 500 books
Resources							
Student/teacher ratio	17.92 (4.44)	11.31 (1.91)	18.04 (4.32)	11.56 (1.73)	5.14	46	Students per teacher at school level
Instruction time	54.55 (4.18)	51.30 (0)	54.65 (4.40)	51.30 (0)	42.12	87.75	Minutes per year/1000 (*-1 in regression
% of low educated math teachers	.22 (.33)	.12 (.19)	.19 (.34)	.12 (.20)	0	1	1-(% of math teachers with highest degree)
Lack of material	.08	.08	.07	.10	0	1	School lacks

	(.28)	(.27)	(.26)	(.30)			material
Institutions							
Public school	.96 (.20)	.97 (.17)	.96 (.19)	.97 (.17)	0 if school is private	1	School type
Standardized tests	.03 (.18)	.26 (.44)	.03 (.17)	.26 (.44)	0	1	Standardized tests more than once a year
No selection	.34 (.47)	.80 (.40)	.41 (.49)	.81 (.40)	0 if school may select	1	School has no right to select its students
Budget (category variable)	1.07 (.39)	1.55 (.52)	1.08 (.39)	1.56 (.53)	0	2	School's right over budget allocation and formulation
School Types							
Vocational school	.05 (.22)	0	.04 (.19)	0	0	1	Vocational school (Berufsschule)
Low sec. school	.19 (.40)	0	.18 (.39)	0	0	1	Low sec. school (Hauptschule)
Medium sec. school	.26 (.44)	0	.28 (.45)	0	0	1	Medium sec. school (Realschule)
Highest sec. school	.29 (.45)	0	.36 (.48)	0	0	1	Highest sec. school (Gymnasium)
Comprehensive school (Ref.)	.17 (.38)	1	.13 (.34)	1	0	1	Comprehensive school (Gesamtschule)
school type n.a.	.03 (.17)	0	0	0	0	1	No information on school type

Table A2: Weighted means (st. dev.) of selected variables by school type for Germany

School type	Students	Reading Score	Parents tert. educ.	% of low educ. teachers	Public school	Student/ teacher ratio
all	4921	490.86 (102.85)	.38 (.48)	.22 (.29)	.96 (.20)	17.92 (4.44)
Vocational	116	476.02 (72.71)	.28 (.45)	.23 (.28)	1 (0)	23.81 (7.56)
Low second.	932	405.76 (80.06)	.18 (.39)	.46 (.31)	1 (0)	17.77 (3.35)
Medium second.	1235	498.19 (73.10)	.34 (.47)	.19 (.27)	.95 (.22)	19.39 (5.18)
High second.	1716	577.56 (68.83)	.61 (.49)	.03 (.10)	.91 (.29)	16.78 (2.15)
Comprehensive	885	467.02 (84.81)	.32 (.47)	.21 (.27)	1 (0)	17.27 (3.16)
School type n.a.	37	287.77 (70.65)	.21 (.41)	.72 (.17)	1 (0)	10.53 (2.82)

Table A3: Coefficients (standard errors) of weighted survey regressions

	With imputed values		Without imp. values		Percentage of missing values	
	GER	FIN	GER	FIN	GER	FIN
Student Background						
Student's age	0.45 (0.35)	-.61* (.33)	0.93** (0.44)	-0.35 (0.43)	0	0
Student's sex	-13.74*** (2.59)	-45.02*** (2.51)	-17.35*** (3.09)	-45.20*** (2.87)	0	0
8th grade	-37.47*** (4.01)	-44.77*** (4.89)	-43.37*** (5.03)	-39.40*** (5.70)	0	0
Parents‘ Origin	-17.35*** (3.66)	-32.59*** (7.40)	-13.58*** (4.36)	-35.02*** (9.57)	2.17	1.42
Parents no sec. Ed.	-37.25*** (10.03)	-26.60*** (4.16)	-39.57*** (12.60)	-29.99*** (4.77)		
Secondary Ed. 2	-22.84*** (4.96)	-30.85*** (4.33)	-26.93*** (6.58)	-33.41*** (4.87)	16.11	10.24
Secondary Ed. 3	-3.14 (2.79)	-10.93*** (3.09)	-4.48 (2.98)	-12.46*** (3.41)		
Books Cat. 1	-65.25*** (11.08)	-67.91*** (17.74)	-66.68*** (15.65)	-64.53*** (14.12)		
Books Cat. 2	-60.56*** (8.29)	-49.70*** (7.56)	-56.51*** (7.17)	-54.87*** (9.18)		
Books Cat. 3	-30.78*** (4.70)	-35.44*** (5.83)	-29.94*** (5.58)	-36.00*** (6.21)	1.87	1.36
Books Cat. 4	-21.73*** (4.48)	-29.36*** (5.60)	-22.25*** (5.34)	-31.55*** (6.20)		
Books Cat. 5	-13.25*** (4.15)	-9.23 (5.72)	-15.68*** (4.96)	-8.93 (6.02)		
Books Cat. 6	-8.27* (4.31)	-0.05 (5.79)	-9.66* (5.05)	-3.53 (6.37)		
Resources						
Student/teacher ratio	-0.24 (0.44)	1.24 (1.12)	-0.47 (0.58)	0.79 (1.05)	14.45	10.61
Neg. Instruction time	0.54 (0.47)	-	-0.46 (0.52)	-	14.79	0
% of low educated math teachers	-9.60 (8.88)	9.31 (9.25)	-13.39* (11.35)	-3.18 (9.90)	8.23	14.56
Lack of material	-6.50 (6.91)	-6.28 (8.70)	1.77 (8.13)	1.80 (6.72)	8.56	1.11
Institutions						
Public school	-2.95 (9.96)	-0.70 (11.48)	10.38 (15.03)	5.31 (12.44)	8.76	0
Standardized tests	-15.58 (12.74)	1.99 (3.82)	-13.72 (12.95)	3.31 (3.70)	10.85	0
No selection	-9.93** (4.18)	-6.98 (4.31)	-17.71*** (4.95)	-1.80 (4.94)	8.78	3.36
Budget (category variable)	-1.31 (4.65)	-6.05 (3.86)	-3.19 (6.23)	-10.32*** (3.83)	8.23	0.66
School Types						
Voc. School	4.81 (8.29)	-	-22.34 (14.86)	-	-	-
Low sec. school	-40.88***	-	-39.59***	-	-	-

	(7.06)		(9.16)			
Medium sec. school	26.69***	-	18.80***	-	-	-
	(6.17)		(7.59)			
Higher sec. school	88.41***	-	76.53***	-	-	-
	(6.02)		(7.36)			
School type n.a.	-153.92***	-	-	-	-	-
	(10.59)					
Intercept	469.12***	718.73***	335.51***	678.59***		
	(69.80)	(66.50)	(83.33)	(81.22)		
Number of observ.	4921	4855	2990	3336		
R-squared	0.5312	0.1720	0.4738	0.1749		
F-Test	187.19	41.93	53.26	30.31		

P-Values: * 1 Percent. ** 5 Percent. * 10 Percent**

Cluster robust standard errors are reported in parentheses.

Table A4: Decomposition for significantly different coefficients at 10-percent-level

	sum	**St. Backgr.**	**Resour.**	**Institut.**	**Schools**	**Interc.**
Total gap	54.28					
Charact. Effect	-27.14	0.16	1.58	-8.03	-20.85	
Ratios	-0.50	0.00	0.03	-0.15	-0.38	
Return Effect	55.91	-169.69	0	0	-20.85	246.45
Ratios	1.03	-3.13	0.00	0.00	-0.38	4.54
Interaction effect	25.52	4.67	0	0	20.85	
Ratios	0.47	0.09	0.00	0.00	0.38	

Absolute effects in reading scores. Ratios are effects divided by total score gap.

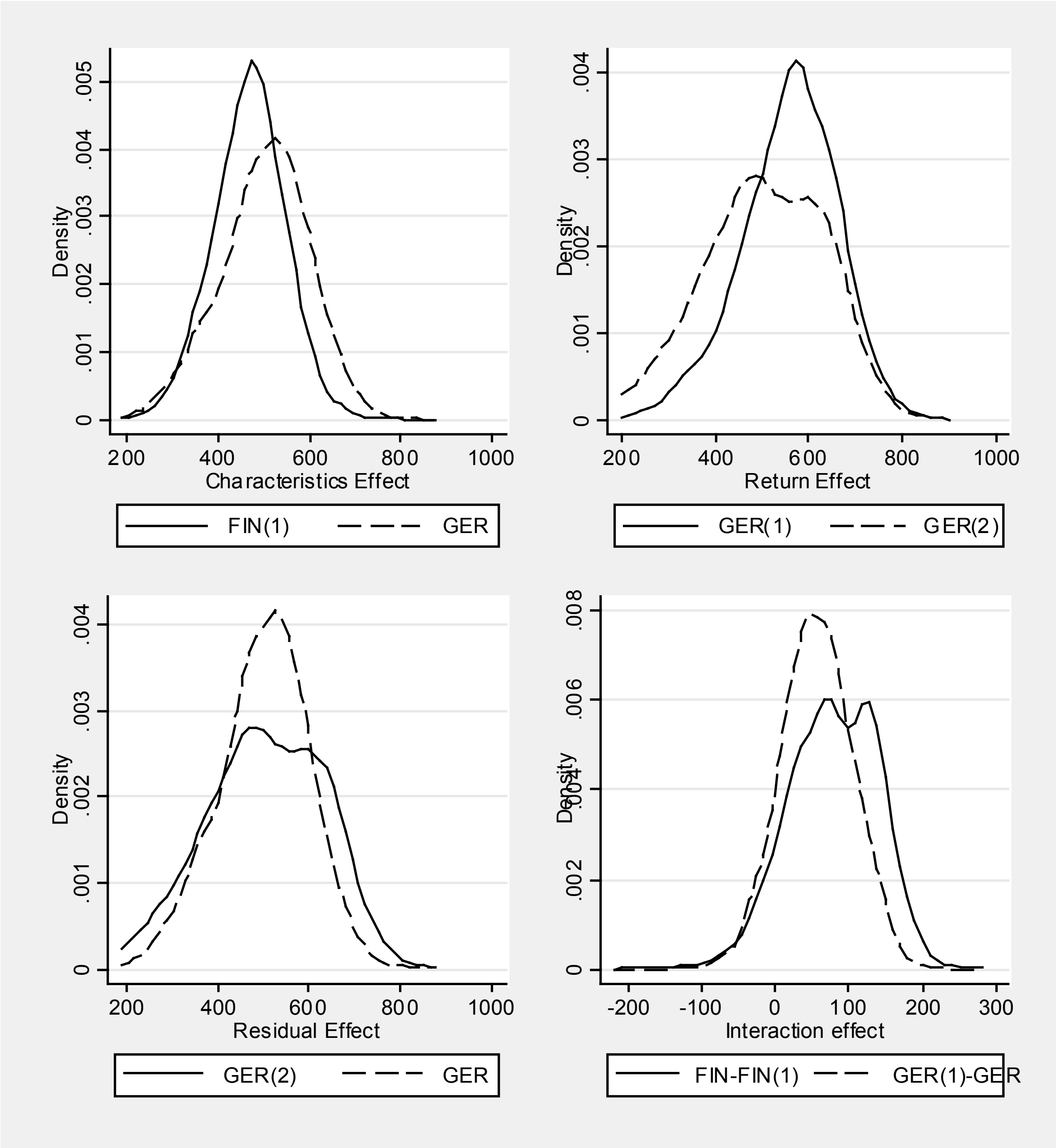

Figure A1: Real and hypothetical test score distributions

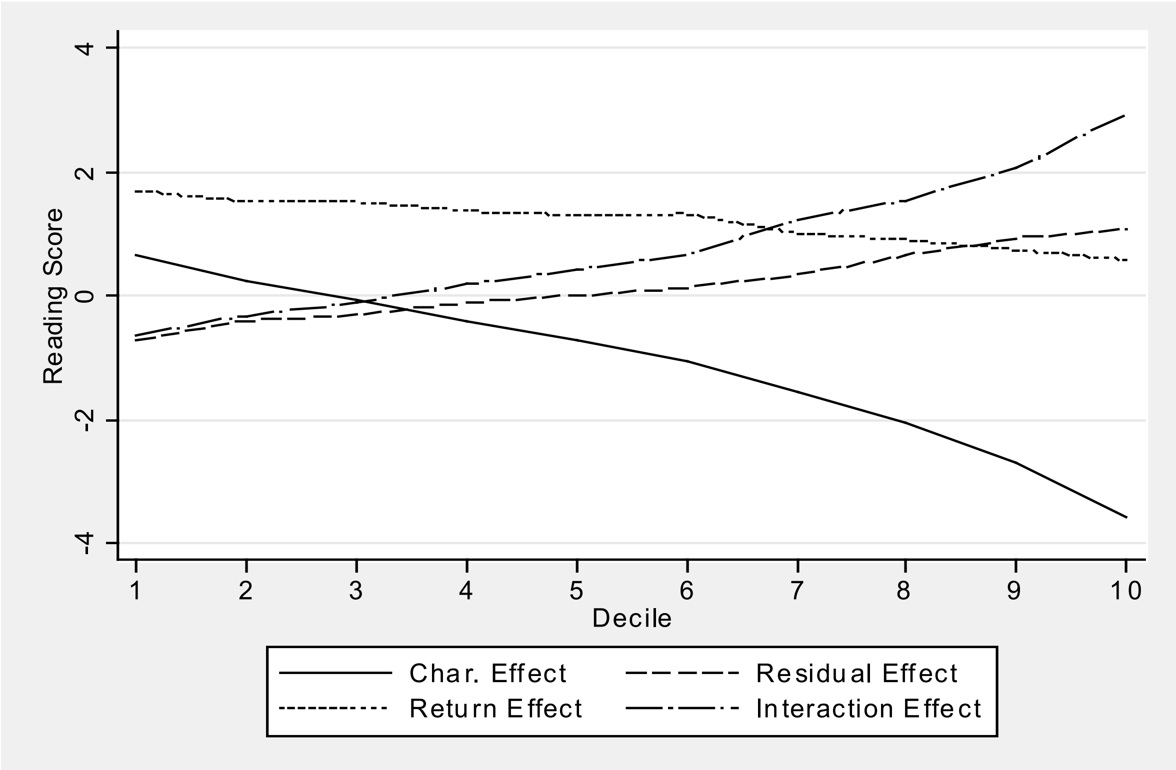

Figure A2: The effects relative to the score gap

Appendix B

A kernel function is a weighting function that produces smoothed estimates of the density at a certain score by basing the density estimate on the frequency of scores in the neighborhood. A large weight is assigned to scores in the near neighborhood and a smaller weight to scores that are further away. The weighted values of the kernel function K are summed in the following function:

$$(1) \qquad F_K = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

The population size is n, x is the score for which we want to estimate the kernel, and h the bandwidth. The variable X_i represents the other scores in the neighborhood of x. The bandwidth h is determined by Silverman's (1986) rule of thumb, which shall obviate the under- or oversmoothing of the data that would allow the variance or the bias to dominate asymptotically, respectively.

The Epanechnikov kernel function K(z), which is the most efficient in minimizing the mean integrated squared error, is used to estimate the density values z = $\left(\frac{x - X_i}{h}\right)$. For |z| < √5,

(2) $$K(z)=\frac{3}{4}(1-\frac{1}{5}z^2)/\sqrt{5},$$

$$K(z)=0$$

otherwise.