IARIW 2021

Monday 23 – Friday 27 August

Measuring the Progress of Equality of Educational Opportunity in Absence of Cardinal Comparability

Gordon Anderson

(University of Toronto)

Maria Grazia Pittau

(Sapienza University of Rome)

Roberto Zelli

(Sapienza University of Rome)

Paper prepared for the 36th IARIW Virtual General Conference
August 23-27, 2021
Session 13: Inequalities in Opportunity for Income, Education and Health I
Time: Wednesday, August 25, 2021 [16:30-18:30 CEST]

Measuring the Progress of Equality of Educational Opportunity in Absence of Cardinal Comparability

Gordon Anderson[*], Maria Grazia Pittau[†], Roberto Zelli[†]

April 29, 2020

Abstract

There are many situations in which measures of the full diversity of a collection of distributions is necessary and where simple comparisons of limited numbers of distributional moments are inadequate since they cast a veil of ignorance over the full extent of distributional differences. An example is the equality of opportunity imperative which demands equal chances for diverse circumstance groups. It requires comparison of distributional differences over the full range of their variation since only then can complete equality of chances be guaranteed. Here new techniques in the form of Gini-like coefficients for quantifying multilateral distributional differences in absence of cardinal comparability are introduced and employed to study changes in the German educational system in the first decade of this century.

Keywords: Distributional Gini coefficients, equality of opportunity, PISA program.

[*]Department of Economics, University of Toronto.
[†]Department of Statistical Sciences, Sapienza University of Rome.

1 Introduction

Inequality of educational outcomes is widely recognized as the premise of the inequality of opportunity later in life for children from different family backgrounds and of intergenerational social and economic mobility achieved by society (Nickell, 2004). It is therefore of crucial interest to understand the impact of education policies on schooling achievements of children with various family origins and socio-economic status. How weakly children's educational performance is related to their family background is interpreted as a proxy of the extent of equality of educational opportunity (Coleman et al., 1966; Mosteller and Moynihan, 1972).

The natural tension between the "private" within family goal of effective transmission of human capability (Becker and Tomes, 1986) and the "public" social justice goal of equal opportunity which seeks independence of child outcomes and parental circumstances (Roemer, 1998, 2006; Roemer et al., 2003) presents public policy challenges with respect to equality of opportunity (EO). Since, relative to their less well endowed counterparts, genetically and materially wealthier parents can always more easily acquire advantage for their offspring, there will always be some dependency in the transition from parental circumstance to child capability, rendering the pure EO goal (complete independence of child outcome from parental circumstance) unattainable. In this situation, Sen (2009) and Atkinson (2012) argue that, rather than seeking an unattainable "transcendental optimum", the policy approach should be progressive, which translates to some combination of progressive equalization of child outcomes (given their common efforts and choices) with progressive equalization of inequalities in their circumstances. Accordingly, measurement tools need to be capable of measuring the degree and significance of advances and retreats from or to the EO goal in terms of diminishing or increasing dependencies of achievements on circumstances.

In order to evaluate the progress over time of equality of opportunity in a schooling system, e.g. before and after a structural reform, achievement distributions can be compared using the counter-factual methods for policy impact evaluation. Usually, distributional differences have been examined in terms of relative differences in conditional means (for a survey see Peragine et al., 2014). However, such comparisons cannot reveal the full panoply of distributional differences and "cast a veil of ignorance" over the full extent of distributional variations (Carneiro, Hansen and Heckman, 2003). For example, suppose all circumstance conditioned distributions had identical means but different variances, relative mean difference measures would record equality of opportunity whereas there would be variation across different circumstance groups in the risks of bad (and good) outcomes, i.e. there would not be equality of opportunity with regard to risk factors encountered by circumstance groups.

Regrettably, when students' performances lose cardinal comparability over time many of the econometric techniques for studying the extent to which things have changed lose their effectiveness. For example, regression-based methods that regress the test performance of individual students on proxies of family background and a set of control variables (Schütz et al., 2008) and even intertemporal dominance comparisons of outcome distributions of circumstance classes (Lefranc et al., 2009; Dardanoni et al., 2010), are no longer viable. Therefore, other means of comparison have to be explored. Such was the case over the period of the German educational reforms in the first decade of the 21st century. Substantive changes in core curricula and in the way students were taught, tested and tracked, meant that achievement results obtained by students in 2003 (before the reforms) were not cardinally comparable with achievements in 2009 (after the reforms).

The question is how to measure progress in this particular context when achievement outcomes cannot be cardinally compared over time? The primary contribution of this paper is to answer this question by introducing and employing semi-parametric mixture models, new Gini-like coefficients of distributional differences and transition structure indices to study changes in the educational outcomes of students from different family-background groups. These indices only require that test scores are reflective of some underlying ordinal classification, they are invariant to and independent of scale (Anderson, 2018), and thus circumvent the cardinal comparability problem.

To avoid arbitrary specification of the number of the classes and their boundaries, student achievement is associated with latent effort/choice variables in a semiparametric finite mixture model which facilitates empirical determination of the number and location of achievement classes. Circumstance to achievement class transitions are then estimated using class membership probabilities derived from the mixture model. The number of classes is determined by a new goodness of fit approach based upon adaptations of the transvariation measure of Gini (1916). Class boundaries are only "partially determined" in the sense that what is established is the probability that a child with a given set of grades is in a particular class. However this does not hinder the study of the number of classes and individual class behavior nor the statistical relationship between outcome classes and circumstance classes admitting as it does the possibility that achievement and circumstance classes may differ in number and vary over time. This is important in the present context because the many changes in the school structure, teaching tracking and testing methods, etc., in the intervening period may have altered the number of achievement classes between the two observation periods. Ultimately the extent to which the transition process has become more equalizing or polarizing in an equality of opportunity sense can then be checked. Finally, Gini-like coefficients for measuring differences in collections of distributions, what we called Distributional Gini coefficients (Anderson et al., 2020), are employed to assess the extent of progress toward an equal opportunity state.

The rest of the paper is organized as follows. Section 2 outlines the basic model and approach to determining outcome and circumstance classes and develops the tools for analyzing the extent to which equality of opportunity has progressed. Section 3 provides background to the reforms that took place in Germany over the 2003–2009 period and reports the main evidence of the progress toward an equal of opportunity goal in the German schooling system. Section 4 draws some conclusions.

2 Methodology

2.1 Capability/Achievement classes and circumstance classes

The underlying structure of student capability acquisition behavior is modeled as a mixture distribution of latent capability acquisition classes. It is assumed that individual observable academic achievements (test scores) are a good proxy for an individuals latent capability acquisition and that there are K_A capability acquisition types or classes indexed $k = 1, ..., K_A$. Members of a particular class share some commonality in their capability acquisition abilities but there will be some within class variation because of differences in innate acquisition skills, efforts and choices and inevitable approximation errors, a consequence of using test scores as a proxy for latent capability acquisition. In essence it is assumed that "k type" individuals have similar skill and effort levels and make similar choices, but the agglomeration of these factors results in outcomes that randomly deviate around a norm based upon the common skill set, effort level for the type. There are good statistical rationales (essentially central limit theorems, see for example Gibrat's law of proportional effects (Sutton, 1997)) for believing that the aggregation of all of these factors is normally distributed so that (log) achievement of individual i, x_i, (her capability index) can be assumed normally distributed, hence for achievement class k, with achievement distribution $f_k(x)$, the achievement of individual i may be written as $x_i = \mu_k + \sigma_k \cdot e_i$ where $e_i \sim N(0;1)$, so that $\sigma_k \cdot e_i$ is a measure of i's cumulated capability deviation from the norm μ_k.[1]

2.2 Mixture models, class membership probabilities and number of classes

For generality purposes, suppose K_A capability acquisition classes emanating from J_C circumstance groups are contemplated. The overall distribution of x may be written as a mixture of sub-distributions as follows:

$$f\left(x, \mathbf{\Psi}\right) = \sum_{k=1}^{K_A} w_k f_k\left(x, \theta_k\right), \tag{1}$$

where $f_k\left(x, \theta_k\right)$ are $\mathrm{N}(\mu_k, \sigma_k)$.[2] The vector $\mathbf{\Psi} = (w_1, \cdots, w_{K-1}, \xi')'$ contains all the unknown parameters of the mixture model: w_k, $k = 1, \cdots, K_A$ are the mixing proportions summing to 1 ($\sum_{k=1}^{K} w_k = 1$); the vector ξ contains all the parameters $(\theta_1, \cdots, \theta_K)$ known *a priori* to be distinct. The w_k represent the *a priori* probabilities of a randomly selected student in the population to belong to achievement class k. They are endogenous parameters which determine the relative importance of each component in the mixture and can be interpreted as unconditional probabilities.

Let the conditional probability of a student i ($i = 1, \cdots, n$) with achievement x_i being in achievement class k ($k = 1, \cdots, K_A$) be given by:

$$\pi_{ik}^{A} = \texttt{Prob}\{A(i) = k \mid (x_i; \mathbf{\Psi})\} = \frac{w_k f_k(x_i)}{\sum_{k=1}^{K} w_k f_k(x_i)}, \tag{2}$$

where $A(i)$ indicates the achievement class component to which student i belongs. The class weights (the unconditional probabilities), are estimated by using the individual class weights π_{ik}^{A} as:

$$\widehat{\pi}_k^{A} = \widehat{w}_k = \frac{1}{n} \sum_{i=1}^{n} \pi_{ik}^{A}, \ \ k = 1, \cdots, K_A. \tag{3}$$

[1] This simple formulation nicely illustrates the problem highlighted in the Carniero et. al. (2003) critique in that just seeking equality in means ignores or masks potential variability in variance which would reflect important differences in high and low achievers.

[2] The choice of normal densities depend on the assumption of normality in effort. However this is not an overly strong assumption since, any continuous distribution can be approximated to some desired degree of accuracy by an appropriate finite Gaussian mixture (Rossi, 2014).

Given the number of classes K, the unknown parameters of the mixture (means, variances and proportions of each component) along with the conditional probabilities (π_{ik}^A) are estimated by maximum likelihood (ML) via the expectation-maximization (EM) algorithm (Dempster et al., 1977).[3]

The probability of a randomly selected student belonging to a given circumstance class j ($j = 1, \cdots, J_C$) is π_j^c. Given a sample $i = 1, \cdots, n$, π_j^c is estimated as:

$$\widehat{\pi_j^c} = \frac{1}{n}\sum_{i=1}^{n} D_{ij} \tag{4}$$

where:

$$D_{ij} = \begin{cases} 1 & \text{if student } i \text{ has circumstance } j \\ 0 & \text{otherwise} \end{cases}$$

Let the $K_A \times J_C$ matrix T whose typical element is $t_{kj} (k = 1, \cdots, K_A; j = 1, \cdots, J_C)$ be the transition matrix yielding the conditional probability of being in achievement class k given circumstance class j. The $K_A \times 1$ vector of achievement class probabilities π^A whose typical element is π_k^A is related to the circumstance class probability vector π^C (whose typical element is π_j^C) by the formula:

$$\pi^A = T \cdot \pi^C \tag{5}$$

Given a sample of students, matrix T may be estimated by a simple regression system of the form:

$$\pi_i^A = T \cdot D_i + \nu_i, \;\; i = 1, \cdots, n \tag{6}$$

where π_i^A is the $K \times 1$ vector of the conditional probabilities of student i, whose typical element is π_{ik}^A; D_i is the $J_C \times 1$ vector whose typical element is D_{ij} and ν_i is a $K_A \times 1$ random vector with zero mean and singular covariance matrix. Thus $\pi^A = \mathrm{E}\left(\pi_i^A\right) = \mathrm{E}\left(TD_i + \nu_i\right) = T \cdot \mathrm{E}\left(D_i\right) = T \cdot \pi^C$, where the columns of T sum to 1. The resulting matrix T of estimates of conditional probabilities of being in achievement class k given circimstance class j is given by:

$$\widehat{t}_{kj} = \frac{1}{|C(i) = j|}\sum_{i \in C(i)=j} \pi_{ik}^A, \;\; k = 1, \cdots, K_A; \;\; j = 1, \cdots, J_C,$$

[3]It is well known that the likelihood function of normal mixtures is unbounded and the global maximizer does not exist (McLachlan and Peel, 2000). Therefore, the maximum likelihood estimator of $\mathbf{\Psi}$ should be the root of the likelihood equation corresponding to the largest of the local maxima. The usual solution is to apply a range of starting values for the iterations. In this paper, randomly selected large sample, non-hierarchical clustering-based starts, are employed (Kaufman and Rousseeuw, 1990).

where $|C(i) = j|$ is the cardinality of group j.

The many changes in teaching methodology, curriculum, teacher training, tracking methods could well engender different numbers of achievement classes before and after the change. Hence each year will be investigated separately to determine the number of classes that best fit the data. The number of achievement classes K_A is in fact unknown and has to be estimated. The selection of 'optimal' K_A is performed by minimizing the proximity of the mixture distribution, $f(x, \mathbf{\Psi})$, to a kernel estimate of the achievement distribution, $f_{krn}(x)$, using two versions of Gini's Transvariation Coefficient (Gini, 1916), which measures the dissimilarity of two distributions, modified by a penalty factor. Following arguments in Akaike (1972), the penalty is the number of coefficients in the mixture times $2/n$ where n is the sample size.

The two versions (unweighted and "importance weighted") of the of Gini's Transvariation Coefficient, GTR and GTRIM, relate to the integral of absolute differences between two probability distribution functions.[4] In particular GTR relates to the overlap measure, $\theta = \int_{-\infty}^{\infty} \min\{f(x, \mathbf{\Psi}), f_{krn}(x)\}\, dx$:

$$\mathrm{GTR} = \int_{-\infty}^{\infty} |f(x, \mathbf{\Psi}) - f_{krn}(x)| dx = 2 - 2\theta. \tag{7}$$

Anderson et al. (2012) showed the overlap estimator $\widehat{\theta}$ to be asymptotically normally distributed with mean equal to θ and a certain variance V, and therefore $\mathrm{GTR} \sim \mathrm{N}(2 - 2\theta, 4V)$, thus facilitating inference for GTR.

GTRIM is an importance weighted version of GTR (see Anderson et al., 2017):

$$\mathrm{GTRIM} = \int_{-\infty}^{\infty} |f(x, \mathbf{\Psi}) - f_{krn}(x)| f_{krn}^{-0.5}(x) dx. \tag{8}$$

Gini's transvariation coefficient can be seen as cumulating the absolute difference between the functions over the whole real line, whereas the GTRIM version can be seen as cumulating the "importance" weighted absolute difference, so that a given difference from a small target plays a bigger role in the calculation than the same order of difference in a correspondingly larger target. The optimal number of components in the mixture is consequently assessed comparing the mixture distribution with the true unknown density, consistently estimated by a kernel estimator. Essentially, the value K that minimizes the penalized GTR or GTRIM is the one that is picked.

[4] Assuming for convenience that $f_{krn}(x)$ has positive support over the whole real line, these tests are closely related to Integrated Squared Difference tests (Chwialkowski, et al., 2015), and Pearson goodness of fit tests (Greenwood and Nikulin, 1996).

2.3 Evaluating generational transition patterns: new indices for square and non-square transition matrices

Of primary interest here is evaluating whether there has been progress in equalizing opportunity in acquiring capability in the context of education. Usually this is a matter of evaluating whether or not the dependency of capability acquisition on individual circumstance has diminished. However, this may be overly simplistic since playing fields can be levelled upward or downward and the usual policy intent is to level upward (as for example in the "No Child Left Behind" policies in the US). Here a collection of new measures is outlined and implemented, based upon the idea that the parental circumstance–child achievement transition is a generational stochastic process (Anderson, 2018) in the sense that this generations outcomes are the next generations circumstances.

Turning first to the outcomes of circumstance classes, assessing the extent of (ex-ante) equality of opportunity is about evaluating the degree to which outcome distributions of different circumstance classes differ. Absolute equality of opportunity prevails if circumstance-conditioned outcome distributions are identical. On the other hand, if they are perfectly segmented with no overlap, there will be complete inequality of opportunity in that members of a particular circumstance class never have similar outcomes to members of any other class and achievement class status would be perfectly predictable given a students circumstance class.

Letting $f_j(x) = f(x|C = j)$ be the conditional probability distribution of the $j - th$ circumstance class outcome, with s_j the relative size of the population with circumstance j, these notions can be reflected in a Gini based relative inequality of distribution index "DisGini" (Anderson et al., 2020) having the form:

$$\text{DisGini} = \frac{1}{1 - \sum_i s_i^2} \sum_{i=1}^{J_C} \sum_{j=1}^{J_C} \int_0^{\infty} s_i s_j |f_i(x) - f_j(x)| dx =$$

$$= \frac{1}{1 - \sum_i s_i^2} \sum_{i=1}^{J_C} \sum_{j=1}^{J_C} \int_0^{\infty} s_i s_j (1 - OV_{ij}) dx \quad (9)$$

where OV_{ij}, the overlap measure $\int_0^{\infty} min(f_i(x), f_j(x))dx$ (Anderson, Linton and Whang, 2012) can be computed from kernel estimates of conditional outcome distributions or the parametric estimates of the mixture distributions. An unweighted version of the index is made possible by simply replacing s_i with $\frac{1}{J_C}$ for all i. This is pertinent in the present context since it is preferable to compare the circumstance conditioned distributions directly rather than importance weight them. Another attractive feature of the Distributional Gini coefficient is that it is readily computed in multivariate environments when x is a vector and f is the corresponding multivariate distribution. In the trivariate form for variables x,y,z, DisGini may be written as:

$$\text{DisGini} = \frac{1}{1-\sum_{k=1}^{K} s_k^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \iint \int_0^{\infty} \frac{s_i s_j |f_i(x,y,z) - f_j(x,y,z)|}{2} dxdydz. \tag{10}$$

DisGini can be shown to take the value 0 when all circumstance class achievement distributions are identical and when those distributions are perfectly segmented it will take the value 1, so the hypothesis of equality of opportunity can be examined by examining the proximity of Disgini to 0. It can readily be shown to be invariant to, and independent of, scaling of x and has a computable standard error and, as a linear function of normal random variables, it will also be normal.

Since this framework specifically allows for different numbers of classes in different eras and the possibility of non-square transition matrices T, the usual equal opportunity–mobility matrices (Shorrocks, 1978) are no longer viable. Instead an index based upon the columns of T is developed. The jth column of T corresponds to the probability distribution over the final state outcome space for agents emerging from initial state j. As such, perfect mobility (where the final state is uninfluenced by or independent of the initial state) is characterized by T having common columns which all sum to 1. Writing the kth row of T as t_k, let maxr$(\cdot)$ and minr$(\cdot)$ be operators which return the maximum and minimum value in a row vector respectively, a transition matrix based index of mobility, (TM), which immediately suggests itself, is:

$$\text{TM}(T) = 1 - \frac{\sum_{k=1}^{K_A} (\text{maxr}(t_k) - \text{minr}(t_k))}{K_A} \tag{11}$$

The index TM can be viewed as a multivariate scaled version of Gini's two distribution dissimilarity "transvariation" index (Gini, 1916) or a "many distribution" overlap index calculating the degree of overlap or similarity in many distributions. The index $TM(T)$ ranges from 0 to 1. In case of perfect mobility (equality of opportunity) each column of T will be identical, the final state outcome distributions emerging from the J_C initial states will overlap perfectly and therefore the sum of maximums will equal the sum of minimums, yielding $TM(T) = 1$.

The case of perfect immobility occurs when individuals coming from one circumstance class end up in only one achievement class. In this case, for each row $k \in K_A$

there is at least one column j such that $t_{kj} = 0$, therefore the sum of minimums will be always equal to zero. We can distinguish three different situations.

1. The number of achievement classes is equal to the number of circumstances: $K_A = J_C$. Thir renders T a square matrix and complete immobility is characterized by the identity matrix ($T = I$), or by any permutation matrix. The sum of maximums will be equal to the order of the matrix T, yielding $TM(T) = 0$.

2. The number of achievement classes is less than the number of circumstances: $K_A < J_C$. Only K_A columns of T are orthogonal and unit vectors. The sum of maximums will be equal to K_A, the number of achievements classes, yielding $TM(T) = 0$.

3. The number of achievement classes is greater than the number of circumstances: $K_A > J_C$. In case of perfect immobility, from each circumstance class j there are transitions only towards one class of achievement, implying that $K_A - J_C$ rows are zero vectors. Consequently, this case can be traced back to the situation of square matrices.

The index $TM(T)$ and DisGini satisfy the normalization, immobility and perfect mobility axioms of Shorrocks (1978). $TM(T)$ also satisfies the strong perfect mobility axiom since $TM(T) = 1$ if and only if T has common columns. However, it does not satisfy strong perfect immobility axiom (that is $TM(T) = 0$ if and only if $T = I$) since $TM(T) = 0$ for any column rearrangement of the identity matrix. The monotonicity axiom that requires $TM(T) > TM(\tilde{T})$ when $t_{kj} \geq \tilde{t}_{kj}$ for all $k \neq j$ with strict inequality holding somewhere, is satisfied if we restrict the analysis to the matrices with quasi-maximal diagonals.[5] Period consistency requires $TM(T) \geq TM(\tilde{T})$ implies $TM(T^s) \geq TM(\tilde{T^s})$ for positive integer $s > 0$. Finally, when the outcome and circumstance variables only have an ordinal ranking (i.e. they cannot be cardinally compared) the index can be shown to have the property of scale invariance and scale independence (see for example Kobus and Milos, 2012).

Most indexes of equality of opportunity focus on the extent to which outcomes are independent of circumstances, measuring the extent to which the playing field has been

[5] A matrix T has a quasi-maximal diagonal when there exists positive $\alpha_1, \cdots, \alpha_k, \cdots, \alpha_K$ such that $\alpha_k t_{kk} \geq \alpha_j t_{kj}$, for all k, j. Restricting the analysis to the subset of quasi-maximal diagonal matrices reconciles the axioms of Perfect Mobility and Monotonicity as shown for the Shorrocks's mobility index (Shorrocks, 1978).

leveled (Roemer, 2006). However, the playing field can be leveled in many different ways, leveling down, leveling up or depolarizing (leveling to the middle) and, when transitions are viewed as a process, the extent to which transitions are upward or downward "leveling" or depolarizing is a matter of interest. In the context of a generational model where this generation's achievements become the next generations circumstances, Anderson (2018) demonstrates how the transition matrix can be used to evaluate whether the transitions are converging (equalizing subsequent generations circumstances) or polarizing (dis-equalizing subsequent generations circumstances) processes via a balance of probabilities measure.

For simplicity of exposition, and coherency with the empirical evidence in Section 3.2, suppose the transition matrix is aggregated into a 4×3 matrix with typical element t_{kj} $k = 1, 2, 3, 4,$ $j = 1, 2, 3$ of transitions to low, lower-middle, upper-middle $(A_1, A_2, A_3, A_4,$ respectively) and high achievements from low, middle, high circumstances $(C_1, C_2, C_3,$ respectively), with circumstance probabilities π_j^c. Processes that promote more transitions from middle classes to the peripheries than transitions from the peripheries to middle classes are polarizing.[6] The probability that a student would move out of the achievement middle class given that she is inside the circumstance middle class initially (the divergent component) is:

$$\text{Prob}\left(i \notin A_2 \,\text{or}\, A_3 \,| i \in C_2\right) = \frac{\text{Prob}\left(i \notin A_2 \,\text{or}\, A_3 \,\text{and}\, i \in C_2\right)}{\text{Prob}\left(i \in C_2\right)} = t_{12} + t_{42}.$$

The probability that a student would be in the achievement middle classes given she is poorly or richly endowed in circumstances (the convergent component) is instead:

$$\begin{aligned} \text{Prob}\left(i \in A_2 \,\text{or}\, A_3 \,| i \notin C_2\right) = \frac{\text{Prob}\left(i \in A_2 \,\text{or}\, A_3 \,\text{and}\, i \notin C_2\right)}{\text{Prob}\left(i \notin C_2\right)} = \\ = \frac{\pi_1^c t_{21} + \pi_1^c t_{31} + \pi_3^c t_{23} + \pi_3^c t_{33}}{\pi_1^c + \pi_3^c} = \\ = w(t_{21} + t_{31}) + (1 - w)(t_{23} + t_{33}), \end{aligned}$$

where $w = \frac{\pi_1^c}{\pi_1^c + \pi_3^c}$.

An overall index PT that summarizes a convergence/polarization process can be intuitively defined as the proportion of students transferring from the peripheries of circumstances to the middle classes of achievements net of the proportion of students

[6]When classes are large in number it is possible to study and assess polarization and convergence to multiple poles in a similar fashion but it is not necessary in this case.

transferring from the middle classes of circumstances to the peripheries of achievements. More formally, index PT is defined as the probability of a student with non-middle class circumstances achieving middle class outcomes less the probability of a student with middle class circumstances achieving non middle class outcomes, that is:

$$PT = w(t_{21} + t_{31}) + (1 - w)(t_{23} + t_{33}) - (t_{12} + t_{42}), \text{ where } w = \frac{\pi_1^c}{\pi_1^c + \pi_3^c}.$$

When PT is positive, a convergent process is detected, while negative values of PT indicate a polarization process.

Following Anderson et al. (2014) and Anderson and Leo (2017), there is interest in seeing whether or not progress toward an equality of opportunity goal is being achieved by elevating the outcomes of those poorly endowed in circumstance rather than diminishing the outcomes of those richly endowed in circumstance, reconciling the private aspirations of parents to preserve the advantages that their offspring with the public aspiration of equality of opportunity. This may be examined with a similar balance of probability measure, PUT, which can assess mobility as upward or downward transiting. From an equality of opportunity perspective convergent processes are to be preferred to polarizing processes since they may be seen as equalizing the circumstance classes of subsequent generations and upward transiting rather than downward transiting processes are to be preferred since they are elevating the circumstance classes of future generations. This index, in the situation in which there are four achievement classes and three circumstances classes, can be written as:

$$PUT = (1 - t_{11})\pi_1^c + (t_{42} - t_{12})\pi_2^c - (1 - t_{43})\pi_3^c.$$

As balance of probability measures, these indices are bounded between -1 and 1. To satisfy the Shorrocks (1978) normalization axiom, apply the transformations $PT^* = 0.5PT + 0.5$ (with values > 0.5 implying convergence *etc.*) and $PUT^* = 0.5PUT + 0.5$ (with values > 0.5 implying increased upward mobility) respectively. As such these indices are probability measures and, for inference purposes, under standard assumptions estimators of probability p are asymptotically $\mathrm{N}(p, (p(1 - p)/n))$ thus facilitating tests of hypotheses of equalizing opportunities and upward mobility.

3 Evaluating the German educational reforms of the first decade of the 21st Century

3.1 Education reforms in Germany

Throughout the latter half of the 20th century Germany had compulsory elementary Grundschule education for all children aged 6 through 10. After, students were tracked into a tripartite school system: Hauptschule, Realschule and Gymnasium. Hauptschule, catered to the majority of lower ability students who, after grade 9, could apply for training at a Berufsschule generally leading to low skilled blue collar jobs. Realschule, enrolled more qualified students who got specific clerical, technical and lower-level civil servant vocational training. The Gymnasium, the highest secondary school, focused on broad preparation in the humanities and the Abitur, the primary gateway to the professions, teaching and the upper levels of the civil service. These arrangements reflect social divisions in Germany much in concert with the view that tracking reinforces the impact of family background.[7]

The results from the first Program for International Student Assessment (PISA) called into question the extent of equality of opportunity in education and raised concerns regarding the extent to which a child's parental circumstances limited their academic achievement in Germany (OECD, 2011, p.208). Germany came well below the overall language, mathematics and science averages for all countries tested, family socio-economic status and student achievement correlations were higher than in any other OECD country. Elementary school children, whose parents had attended the highest school level, were three times as likely to be sent to that same highest level of school as children of equivalent ability whose parents had graduated from the lowest school level. In short there was overwhelming evidence of a dependence of a child's educational outcomes on the parental circumstances they confronted, an evident lack of equality of opportunity.

This prompted a five year €4 billion plan to reform the German schooling system involving substantive changes in curricula, in the way students were taught, tested and tracked in addition to changes affecting their parental circumstance. The "Kultusminister Konferenz", formed by education ministers in the sixteen German States, proposed seven central areas requiring change (PISA, 2002). These were (Avenarius

[7]German school tracking, long viewed as in institutional device reinforcing the intergenerational persistence in educational achievements across different social classes has been the object of considerable study (for an excellent survey see Krause and Schuller, 2014).

et al., 2003): 1) Improve preschooler speaking capabilities; 2) Earlier enrollment into Grundschule; 3) Improved reading, mathematics and natural science *curricula*; 4) Direct help with learning difficulties for immigrant youths; 5) Promote uniform testing, compliance with international standards and result focused evaluation of new testing methods; 6) Improved teacher training and diagnosis and support of students with learning difficulties; 7) Introduce and expand all-day school programs to provide more extensive education for students with learning difficulties or special skills.

Full implementation of the above differed across states. All-day school programs were embarked upon in 2003 extending the school day until 4:00 pm or later. In 2003 national educational standards were introduced for children in primary and secondary school in German, Mathematics, a First Foreign Language and Sciences. 2007 saw additional standards put in place throughout Germany for students at the end of grade 10 covering subject-specific competencies at a similar level as the PISA tests, prior to this there had never been national standards. To increase the probability of a higher secondary school attendance, some states delayed tracking into the tripartite system until age 12 rather than age 10. Other states combined the Realschule and Hauptschule into one, while some allowed students in lower schools to move up the tripartite ladder and complete their education with a more prestigious background, which Brunello and Checchi (2006) suggest would dilute the influence of parental background. This led to speculation that the 2,625 Realschulen and 4,283 Hauptschulen would no longer co-exist within 10 years and will merge into one type of school (There were 3,070 Gymnasien during that time, less than half the other two school types combined). These reforms brought fundamental change to the old school structure, which had focused on having few highly educated people, several with medium education and the majority with little education.

3.2 Empirical evidence

The Program for International Student Assessment (PISA) results for Germany in the years 2003 and 2009 were employed. PISA data for German students aged between 15 years 3 months and 16 years 2 months who have completed at least six years of formal schooling in 2003 (before the reforms had been implemented) and 2009 (after the reforms had taken effect) is used to construct an achievement index for students who have completed exams in Math ($X1$), Reading ($X2$), and Science ($X3$). Since EO pertains to the development of a students' overall capabilities and gender equity is paramount, all disciplines and genders are considered jointly.[8] Overall evaluation follows the usual practice of averaging student grades across disciplines. Initially that practice is followed here but, because the evaluation methodology in the different PISA disciplines differed between disciplines and over time, the overall achievement index in each year was based upon the average of their maximum mark standardized subject scores i.e. for person i:

$$\mathrm{X}_{(i)} = \frac{1}{3} \cdot \left\{ \frac{X1_i}{\mathrm{Max}(X1)} + \frac{X2_i}{\mathrm{Max}(X2)} + \frac{X3_i}{\mathrm{Max}(X3)} \right\}.$$

Since students in a given year will have confronted common assessment schemes, curricula and teaching methods, this outcome index is cardinally comparable within that year. However it is not comparable within the country between years since, due to the many and various interventions mentioned above, curricula, examination, grading, teaching and learning methodologies and standards, all will have changed between comparison years. Subsequently the circumstance conditioned joint distributions of the math reading and science triple will be considered in the context of equation (10). This avoids the arbitrary weighting process above and is good for all possible monotonic non-decreasing weighting functions of the discipline marks.

To develop circumstance classes that reflect a student's family environment, several indicators can be selected.[9] To simplify matters, the only data employed to describe parental circumstance is family type (one or two parents present) and the educational status of those parents. Coherent family income data is not available in both periods, in any event educational status will be a good proxy for family income. Further note that in these studies generally only the corresponding income/education status of the correspondingly gendered parent is used (see for example Arrow et al., 2000) whereas here the circumstance variate is a family effect for children irrespective of gender. To develop circumstance classes that reflect a student's family environment (circum-

[8]It would be also interesting, and left for future research, to investigate the progress in EO separately for gender and other students' characteristics. Moreover, in this paper we are interested in an overall evaluation of the skills mastered by students at the age of 15, but a more comprehensive understanding could emerge from the analysis of each discipline since similar achievement index may reflect different scores for each field. Actually, correlation between test scores in each field is not strong. Correlation between test scores in 2003 shows a moderate positive correlation between reading and mathematical literacy (ρ=0.534), a moderate correlation between reading and science (ρ=0.367), while correlation between science and math is negligible (ρ=0.067). Interestingly, in 2009 correlation between science and math becomes positive and significant (ρ=0.460), the correlation between reading and science becomes negative (ρ=-0.277), while correlation between reading and math remains moderately positive (ρ=0.405).

[9]An extensive discussion of context indicators is, e.g., in Duru-Bellat and Souchat (2005).

stances) an index is constructed by adding the educational status (a six point scale) of each parent present in the household and dividing by the square root of the number of parents present. This is akin to using the square root rule for parental circumstance support common in consumer equivalence scaling (Brady and Barber, 1948) wherein there is an advantage to the presence of more than one parent but it is an advantage with diminishing returns to scale (0.5 elasticity). This index is then used to define three ordered categories: Lower, Middle and Upper of roughly equal sizes in the initial year by exploiting gaps in the index scale so that, unlike the "smoothly distributed" achievement variable, circumstance class membership is definitively discrete.[10] Table 1 reports the observed parental class sizes.

Table 1: *Circumstance class sizes*

		2003	2009
C1	Lower	0.430	0.367
C2	Middle	0.343	0.391
C3	Upper	0.227	0.242

Table 2 reports summary statistics of the raw data and the constructed achievement and circumstance variables to be used in this study.

Table 2: *Summary Statistics*

	2003 (n=832)					2009 (n=1627)				
	max	min	mean	median	std.dev.	max	min	mean	median	std.dev.
Math Score	0.261	0.011	0.096	0.087	0.056	0.375	0.025	0.170	0.175	0.075
Reading Score	0.471	0.029	0.302	0.324	0.113	0.230	0.007	0.106	0.086	0.055
Science Score	0.944	0.028	0.373	0.333	0.201	0.661	0.018	0.308	0.268	0.146
Fathers Educ	6.000	0.000	3.901	4.000	1.613	6.000	0.000	4.157	4.000	1.465
Mothers Educ	6.000	0.000	3.630	4.000	1.557	6.000	0.000	3.848	4.000	1.445
Achievement	0.800	0.054	0.469	0.485	0.165	0.799	0.082	0.460	0.479	0.148
Circumstance	8.485	0.000	5.110	4.975	2.030	8.485	0.000	5.386	5.657	1.854

The raw data reveals improvements in parental circumstances over the period, though an increase in the prevalence of single parent families is evident, this is all reflected in the circumstance variable which shows increases in the mean and median and a reduction in the spread over the period. Coherent changes in the raw achievement variables are more difficult to discern because of the changes in national standards, curriculum, teaching and testing methods implemented in the intervening period. Essentially nothing can be deduced from the fact that mean and median scores have gone

[10]Circumstances could be treated in a similar semiparametric fashion but the technique yielded a variable which was effectively discrete with just 20 points of support.

up in Math and down in Reading and Science over the period (again a consequence of the lack of inter-temporal cardinal comparability). Note that the math score distributions have switched from being right skewed to left skewed over the period with the reading score distribution going in the opposite direction. The capability acquisition variable shows a slight decline in the mean and median with a reduction in the spread.

Turning to the determination of the achievement groups, the results of the various versions of the group number selection criteria are reported in Table 3. Visual representations of kernel[11] and semi-parametric versions of the distributions are provided in Figures 1 and 2.

Table 3: *Kernel* vs *semi-parametric mixture 2003 and 2009: transvariation measures.*

No. classes	Year	GTR	GTR st.error	GTR + penalty	GTRIM	GTRIM+ penalty
2	2003	0.113	0.005	0.127	0.097	0.113
	2009	0.072	0.002	0.079	0.063	0.073
3	2003	0.062	0.004	**0.084**	0.070	**0.090**
	2009	0.051	0.002	0.062	0.055	0.067
4	2003	0.057	0.004	0.086	0.064	0.094
	2009	0.040	0.002	**0.055**	0.047	**0.065**
5	2003	0.058	0.004	0.094	0.063	0.101
	2009	0.065	0.002	0.083	0.066	0.082

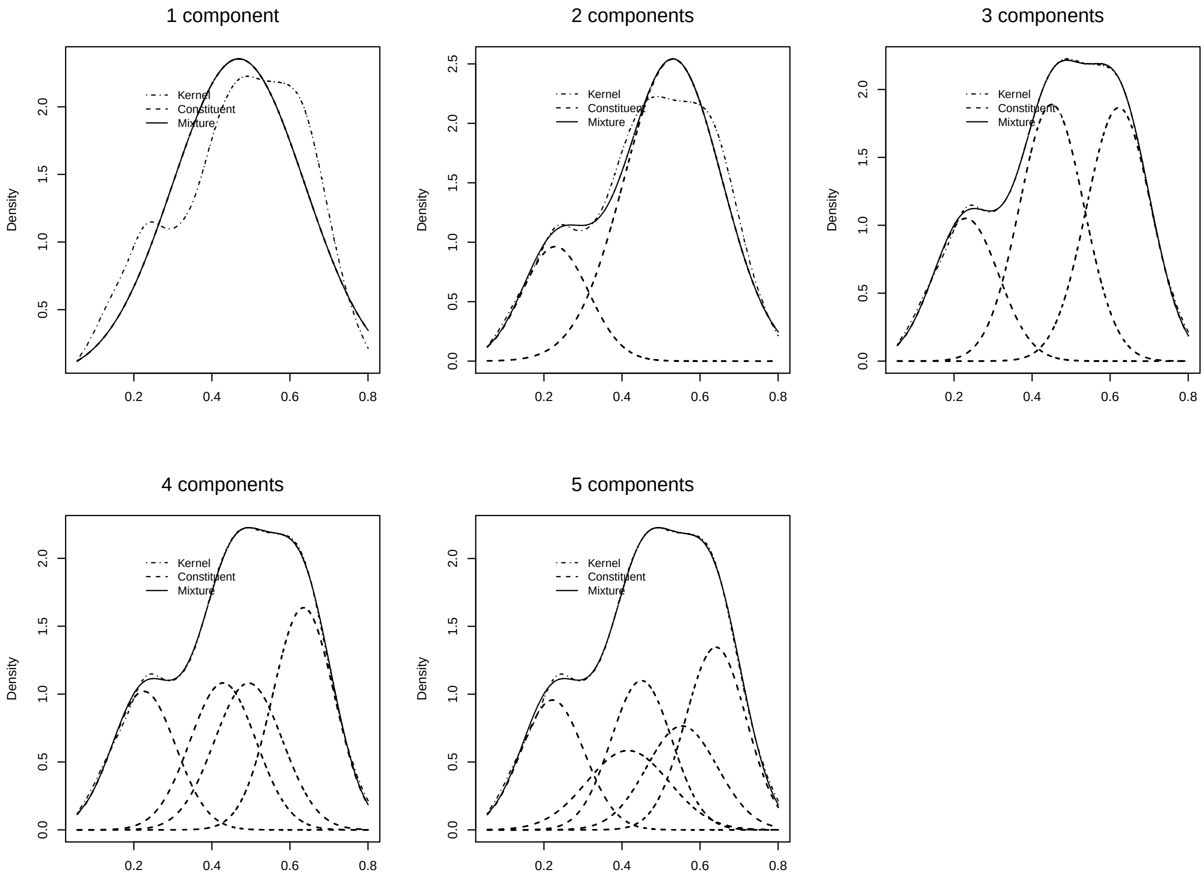
Note that for all component comparisons GTR measures are significantly different at conventional levels of significance with the exception of the 4 *versus* 5 component 2003 comparison. Both unweighted and importance weighted transvariation measures yield the same conclusions when there is no penalization factor, 5 components in 2003 and 4 components in 2009. Similarly they yield the same conclusions under parsimony penalization, this time 3 components in 2003 and 4 components in 2009. Another way of viewing this result is that parsimony penalization only has an impact on the choice of the number of components in 2003 (which was the smaller sample year). Thus, following the parsimony penalized criterion, the 3 achievement group model was selected for 2003 and the 4 achievement group model was selected for 2009.[12] This is also consistent with the figures in which the fitting with 3 components in 2003 and 4 components in 2009 is almost perfect. Evidently, as anticipated, the raft of policy

[11] A Gaussian kernel density estimator was employed. The bandwidth has been estimated using the plug-in procedure of Sheather and Jones (1991).

[12] Widely-used parsimony-based criteria (AIC, AIC3, CAIC, BIC) confirm that in 2003 the prevalent choice is three components and that in 2009 is definitely four components.

measures in the intervening period appears to have changed the number of achievement classes between observation periods.

Figure 1: *Kernel and mixture estimation of achievement in Germany: year 2003.*



The resultant achievement sub-group distributions for the two years are reported in Table 4. In 2003 all achievement groups have the same standard deviation suggesting that the effort distribution is common to all groups. As may be observed the lowest achievement group has a similar population share in both 2003 and 2009 with a similar standard deviation (suggesting no change in the effort distribution) in both periods. The Middle and Upper achievement groups of 2003 seem to have re-oriented themselves by 2009 into three equally sized groups identified as the Lower-Middle achievement group, the Upper-Middle achievement group and the High achievement group so that 2009 sees four roughly equal sized achievement groups. The effort distribution of the Lower-Middle achievement group has remained the same (an insignificant reduction in the standard deviation) whereas the effort distribution of the Upper-Middle and High achievement groups has tightened significantly in 2009.

Figure 2: *Kernel and mixture estimation of achievement in Germany: year 2009.*

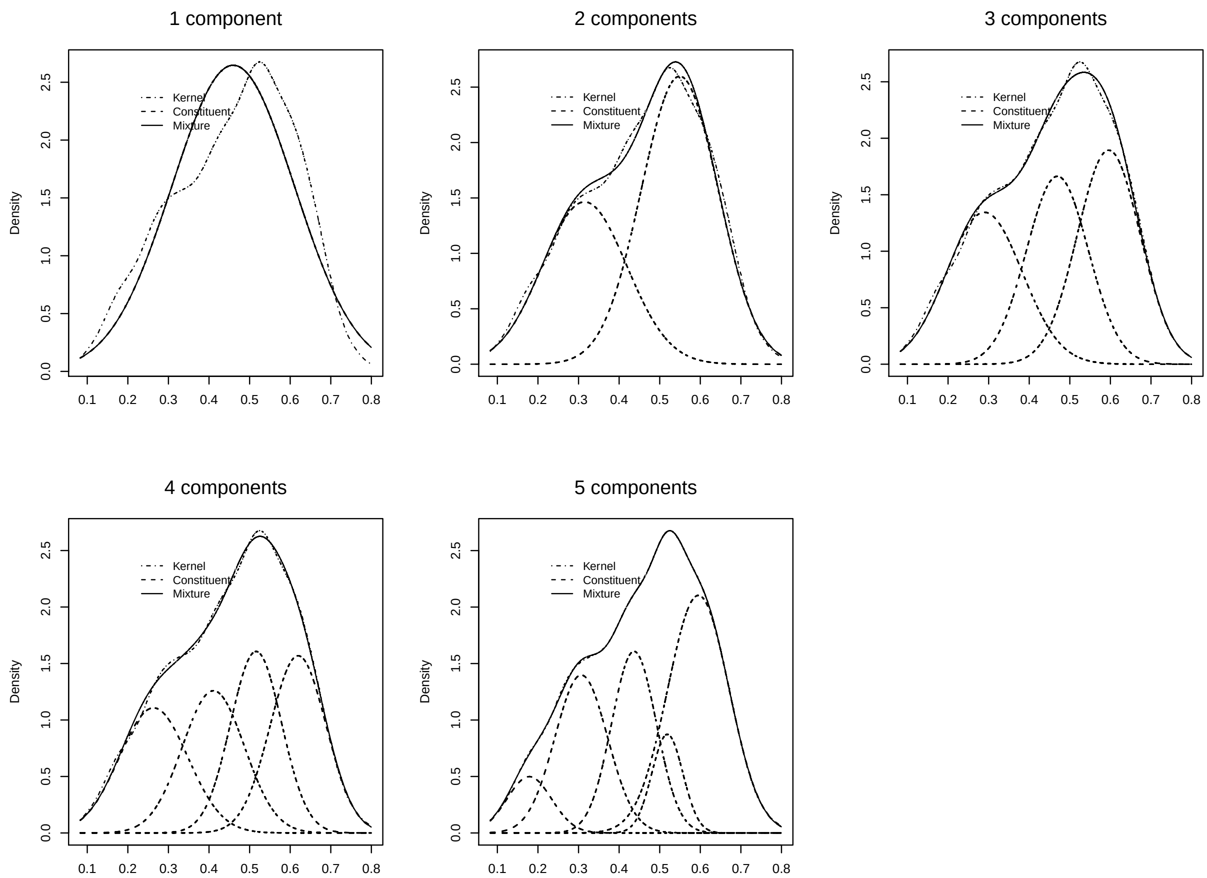


Table 4: *Achievement sub-group distributions: the components of the mixture models: years 2003 and 2009.*

		mean	std dev	weights
	2003			
A1	Low achievement group	0.231	0.075	0.220
A2	Middle achievement group	0.450	0.074	0.391
A3	High achievement group	0.622	0.074	0.389
	2009			
A1	Low achievement group	0.263	0.078	0.234
A2	Lower-Middle achievement group	0.411	0.070	0.242
A3	Upper-Middle achievement group	0.516	0.055	0.253
A4	High achievement group	0.620	0.062	0.271

Because of a lack of cardinal comparability, dominance comparisons cannot be made between the years but they can be made between classes within years. With the exception of circumstance classes 2 and 3 in 2009, Table 5 presents evidence of 1st order dominance relationships of circumstance class conditional achievement distributions for all groups in a given year for each of the years. That is to say, with the exception of the circumstance groups 2 and 3 in 2009, achievement distributions of higher class circumstance groups always dominate those of lower class circumstance groups in both years. Since first order dominance always prevails so will second order dominance prevail, thus demonstrating that the equality of opportunity imperative has not been achieved in either year establishing that the "transcendentally optimal" equal opportunity state has not been achieved in either year.

Table 5: *Differences in achievement cumulative densities conditional on circumstance class: years 2003 and 2009.*

	2003			2009		
	$F(x\|C1)$ $-F(x\|C3)$	$F(x\|C1)$ $-F(x\|C2)$	$F(x\|C2)$ $-F(x\|C3)$	$F(x\|C1)$ $-F(x\|C3)$	$F(x\|C1)$ $-F(x\|C2)$	$F(x\|C2)$ $-F(x\|C3)$
Max	0.258	0.157	0.104	0.255	0.148	0.123
Min	0.000	0.000	0.000	0.000	0.000	-0.027

Table 6: *Circumstance to achievement transitions: years 2003 and 2009.*

	2003 TM(M)=0.804		
	Circumstance 1	Circumstance 2	Circumstance 3
A1 Low achievement group	0.320	0.169	0.108
	(16.528)	(7.802)	(4.071)
A2 Middle achievement group	0.415	0.399	0.334
	(21.103)	(18.065)	(12.311)
A3 High achievement group	0.265	0.432	0.558
	(12.697)	(18.496)	(19.449)
	2009 TM(M)=0.852		
	Circumstance 1	Circumstance 2	Circumstance 3
A1 Low achievement group	0.338	0.191	0.144
	(23.960)	(14.018)	(8.274)
A2 Lower-middle achievement group	0.269	0.256	0.180
	(25.806)	(25.424)	(14.072)
A3 Upper-middle achievement group	0.174	0.70	0.419
	(12.401)	(19.827)	(24.245)
A4 High achievement group	0.219	0.283	0.257
	(21.062)	(28.148)	(20.114)

Numbers in brackets indicate asymptotic t-values.

In analysing the circumstance to achievement transition matrices, note from Table 6 that the transition process has clearly changed in nature with the emergence of an additional achievement class in 2009. From Table 7 mobility improved appreciably with a TM index value of 0.804 in 2003 and 0.852 in 2009, a statistically significant change at all conventional levels of significance. The transformed converging - polarizing balance of probabilities measure ($PT^* < 0.5$ indicating polarizing, $PT^* > 0.5$ indicating converging processes) indicates significant polarizing patterns in 2003 (preserving and reinforcing class stratifications in subsequent generations) and significant converging patterns in 2009 (reducing class differences in future generations). The upward mobility index PUT^* indicates both 2003 and 2009 processes were upwardly mobile, though the propensity for upward mobility had diminished significantly in 2009.

Table 7: *Measures of mobility and polarization of transition matrices, 2003 and 2009.*

	2003	2009
Mobility Index (TM) (Standard Error)	0.804 (0.014)	0.852 (0.009)
2009–2003 difference (Standard Error)	0.048 (0.016)	
Polarization/Convergence Index (PT*) (Standard Error)	0.390 (0.017)	0.516 (0.012)
Polarization Test H_0 : PT* < 0.5; H_1 : PT* ≥ 0.5	Fail to reject (5%)	Reject (5%)
2009–2003 difference (Standard Error)	0.226 (0.021)	
Upward Mobility Index (PUT*) (Standard Error)	0.642 (0.017)	0.553 (0.012)
Downward Mobility Test H_0 : PUT* < 0.5; H_1 : PUT* ≥ 0.5	Reject (5%)	Reject (5%)
2009–2003 difference (Standard Error)	-0.089 (0.020)	

The Distributional Gini coefficient is an index of the extent of differentness in the attainment distributions. Using Gaussian kernel estimates of the three circumstance conditioned univariate achievement distributions and multivariate achievement distributions, Table 8 reports the weighted and unweighted Distributional Gini coefficients for the univariate (average achievement over the three disciplines) and multivariate (wherein the three disciplines are considered jointly rather than aggregated).

Both sets of results indicate a significant drop in distributional inequality over the period (asymptotic t statistics of 3.125 and more) indicating some significant movement toward equality of opportunity, more so in the multivariate analysis which avoids the restrictive marks aggregation process involved in the univariate analysis. There the similarity ends, generally the circumstance conditioned univariate distributions overlap much less than the corresponding multivariate distributions and thus yield higher inequality coefficients. Under marks aggregation the middle and upper circumstance groups exhibit the most commonality in both periods whereas in the multivariate anal-

Table 8: *Weighted and unweighted overlap measures and Distributional Gini coefficients, years 2003 and 2009.*

2003 (weights: L 0.3425; M 0.3534; H 0.3041)	Unidimensional comparison		Multidimensional comparison	
OV_{LM} (Standard Error)	0.4388	(0.0156)	0.9111	(0.0090)
OV_{LH} (Standard Error)	0.4166	(0.0155)	0.8840	(0.0101)
OV_{MH} (Standard Error)	0.6150	(0.0153)	0.2324	(0.0133)
Weighted DisGini (Standard Error)	0.3401	(0.0084)	0.2067	(0.0070)
Unweighted DisGini (Standard Error)	0.3398	(0.0084)	0.2160	(0.0070)
2009 (weights: L 0.2698; M 0.3817; H 0.3485)	Unidimensional comparison		Multidimensional comparison	
OV_{LM} (Standard Error)	0.4905	(0.0158)	0.9032	(0.0093)
OV_{LH} (Standard Error)	0.3508	(0.0150)	0.9125	(0.0089)
OV_{MH} (Standard Error)	0.7145	(0.0142)	0.9830	(0.0040)
Weighted DisGini (Standard Error)	0.3029	(0.0083)	0.0419	(0.0056)
Unweighted DisGini (Standard Error)	0.3209	(0.0083)	0.0447	(0.0058)

ysis they appear to have little in common in 2003 but a great deal in common in 2009, a difference which is largely responsible for the considerable decline in inequality. Weighting/not weighting does not appear to make much difference but this is not surprising since circumstance group sizes are fairly similar.

4 Concluding Remarks

Poor student outcomes in the 2000 round of PISA scores prompted extensive changes teaching, testing and tracking practices in Germany over the ensuing years in order to advance equality of opportunity in capability acquisition. Though a complete equality of opportunity objective will never likely be attained it is still possible to measure progress toward this goal. Unfortunately, from a measurement perspective, the changes that took place meant that student capability distributions were not inter-temporally comparable in a cardinal sense. Accordingly new tools which circumvent the cardinal comparability problem and facilitate comparative examination of parental circumstance - child outcome transitional structures before and after the reforms were introduced. Using PISA data for the years 2003 - immediately prior to - and 2009 - sometime after introduction of the reforms, indices and tests for determining the number of classes in a mixture distribution of capability acquisition (the grade distribution), the polarizing or converging nature of transitional structures and the degree and type of mobility were all proposed and implemented. Considerable change in the circumstance to capability transitional structure was detected with the emergence of an additional capability class. A significant increase in mobility indices over the period indicated progress toward an equal opportunity goal which was the objective of the reforms. The inherent transitional processes also changed from polarizing to converging processes which can be seen to be equalizing the circumstances of future generations.

Acknowledgments

We acknowledge Thomas Fruehauf for his research assistance and for having raised the issue of evaluating the German education system when only ordinal comparability over time is allowed. We would like to thank two anonymous referees, and participants of the Equal Chances International Conference (Bari), the World Bank Conference on "Equity and Development: Ten Years On" (Washington) and economics seminars at Cambridge University and the University of Toronto for their useful comments and suggestions. Obviously we are the solely responsible of any further errors and omissions.

References

Akaike, H. (1972). “Information theory and an extension of the maximum likelihood principle”. *Proceedings of the 2nd International Symposium on Information Theory to Problems of Control and Information Theory*, 267–281.

Anderson, G. (2018). “Measuring Aspects of Mobility, Polarization and Convergence in the Absence of Cardinality: Indices Based Upon Transitional Typology”. *Social Indicators Research*, 139, 887–907.

Anderson, G., Leo, T.W., Muelhaupt, R. (2014). “Measuring Advances in Equality of Opportunity: The Changing Gender Gap in Educational Attainment in Canada in the Last Half Century”, *Social Indicators Research*, 119, 73–99.

Anderson, G. and Leo T.W. (2017). “Quantifying the Progress of Economic and Social Justice: Charting Changes in Equality of Opportunity in the USA, 1960-2000”. *Journal of Human Development and Capabilities*, 18(1), 17–45.

Anderson, G., Linton O., Whang, Y-J. (2012). “Nonparametric estimation and inference about the overlap of two distributions”. *Journal of Econometrics*, 171(1), 1–23.

Anderson, G., Linton, O., Pittau, M.G., Whang, Y-J, Zelli, R. (2020). “Are Eurozone Household Income Distributions Converging? Introducing MGT and DisGini, New Tools for Multilateral Distributional Comparisons”. Department of Economics Working Papers-657, University of Toronto.

Arrow, K.J., Bowles, S., Durlauf, S. N. (2000). *Meritocracy and Economic Inequality.* Princeton University Press.

Atkinson, A.B. (2012). “Public economics after the idea of justice”. *Journal of Human Development and Capabilities*, 13(4), 521–536.

Avenarius, H., Ditton H., Döbert H., Klemm K., Klieme E., Rürup M., Tenorth H.E., Weishaupt H., Weiß M. (2003). “Bildungsbericht Für Deutschland”. Opladen: Leske Budrich. 258–282.

Becker, G.S., Tomes, N. (1986). “Family capital and the rise and fall of families”, *Journal of Labor Economics*, 4, 31–39.

Brady, D.S., Barber, H.A. (1948). “The Pattern of Food Expenditures”, *Review of Economics and Statistics*, 30, 198–206.

Brunello, G., Checchi, D. (2006). “Does School Tracking Affect Equality of Opportunity? New International Evidence”. IZA Discussion Paper No. 2348.

Carneiro, P., Hansen, K.T., Heckman, J.J. (2003). “2001 Lawrence R. Klein Lecture

Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice". *International Economic Review*, 44, 361–422.

Chwialkowski, K., Ramdas, A., Sejdinovic, D., Gretton, A. (2015), "Fast Two-Sample Testing with Analytic Representations of Probability Measures", *Advances in Neural Information Processing Systems (NIPS)*, 28, 1981–1989.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., York, R.L. (1966). *Equality of educational opportunity.* Office of Education, US Department of Health, Education and Welfare, Washington, D.C.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). "Maximum likelihood from incomplete data via EM algorithm", *Journal of the Royal Statistical Society*, B, 69, 1–38.

Duru-Bellat, M., Suchaut, B. (2005). "Organisation and Context, Efficiency and equity of Educational Systems: what PISA tells us", *European Educational Research Journal*, 4(3), 181–194.

Gini, C. (1916). "Il concetto di transvariazione e le sue prime applicazioni", *Giornale degli Economisti e Rivista di Statistica*. 1–55.

Greenwood, P.E., Nikulin, M.S. (1996). *A guide to chi-squared testing.* Wiley, New York.

Kaufman, L., Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.

Kobus, M., Milos P. (2012). "Inequality decomposition by population subgroups for ordinal data", *Journal of Health Economics*, 31, 15–21.

Krause, A., Schüller, S. (2014). "Evidence and Persistence of Education Inequality in an Early-Tracking System-The German Case-". FBK-IRVAPP Working Paper No. 2014-07.

McLachlan, G., Peel, D. (2000). *Finite Mixture Models.* Wiley, New York.

Mosteller, F., Moynihan, D.P. (eds.) (1972). *On Equality of Educational Opportunity.* New York: Random House.

Nickell, S. (2004). "Poverty and Worklessness in Britain", *Economic Journal*, 114, C1-C25.

OECD (2011). Strong Performers and Successful Reformers in Education: Lessons from PISA. OECD Publishing, Paris.

Peragine, V., Palmisano, F., Brunori, P. (2014). "Economic Growth and Equality of Opportunity", *The World Bank Economic Review*, 28, 247–281.

PISA 2000–Zentrale Handlungsfelder. (2002). "Zusammenfassende Darstellung der laufenden und geplanten Maßnahmen in den Ländern". Kultusminister Konferenz. 6-7. http://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2002/massnahmen.pdf.

Roemer, J.E. (1998). *Equality of Opportunity*. Cambridge, Massachusetts: Harvard University Press.

Roemer, J.E., Aaberge, R., Colombino, U., Fritzell, J., Jenkins, S., Lefranc, A., Marx, I., Page, M., Pommer, E., Ruiz-Castillo, J., Segundo, M.J.S., Tranaes, T., Trannoy, A., Wagner, G. G., Zubiri, I. (2003). "To What Extent do Fiscal Regimes Equalize Opportunities for Income Acquisition Among Citizens". *Journal of Public Economics*, 87, 539–565.

Roemer, J.E. (2006). *Democracy, Education and Equality: Graz Schumpeter Lectures*. Cambridge University Press. Econometric Society Monographs No. 40.

Rossi. P.E. (2014). *Bayesian Non and Semi-parametric Methods and Applications*. Princeton University Press.

Schütz, G., Ursprung H., Wößmann L. (2008). "Education policy and equality of opportunity", *Kyklos*, 61, 279–308.

Sen, A.K. (2009). *The Idea of Justice*. Harvard University Press.

Sheather, S.J., Jones, M.C. (1991). "A reliable data-based bandwidth selection method for kernel density estimation", *Journal of Royal Statistical Society (B)* 53, 683–690.

Shorrocks, A.F. (1978). "The Measurement of Mobility" *Econometrica*, 46, 1013–1024.

Sutton, J. (1997). "Gibrat's Legacy", *Journal of Economic Literature*, 35, 40–59.