

Evaluating the Progress of Germany's 2003–2009 Equal Opportunity Imperative When Outcomes are not Cardinaly Comparable

Gordon Anderson, Thomas Fruehauf*
Maria Grazia Pittau and Roberto Zelli†

October 1, 2016

Abstract

The results of the first Program for International Student Assessment (PISA) administered in Germany shocked the nation and prompted a €4 billion plan to reform the schooling system involving intensified parent and teacher training, revised curricula, increased schooling hours and changes in the way students were taught, tested and tracked. Unfortunately the reforms meant that measures of student performance before and after the implementation were not cardinaly comparable, in particular it rendered recent approaches to the evaluation of the social justice – equality of opportunity imperative difficult to implement. Here, by employing a mixture distribution technology to classify student achievements, new techniques verifying the achievement model and for measuring advances in the social justice imperative are proposed and implemented in such circumstances. Fundamental changes in the structure of the dependency of child outcomes on circumstances were detected with some qualified improvements in equality of opportunity over the period.

Keywords: Fitting mixture distributions, equality of opportunity measurement, PISA program.

JEL classification: I21, C14.

*Department of Economics, University of Toronto,
email: anderson@chass.utoronto.ca, thomasfruehauf@yahoo.ca

†Department of Statistical Sciences, Sapienza University of Rome,
email: grazia.pittau@uniroma1.it, roberto.zelli@uniroma1.it.

1 Introduction

An integral part of the evolution of the human capital endowment of a society is the mechanism by which it is passed on and augmented over a sequence of generations. Within such mechanisms there is a natural tension between the “Private” efficiency aspect of effective generational transmission of human capital in a family context (Becker and Tomes 1986) and the “Public” social justice goal of “equality of opportunity” which seeks independence of a child’s outcomes and their parental circumstances (Roemer, 1998, 2006; Roemer et al., 2003). Parental altruism leads parents to seek advantage for their offspring, privately investing in their children over and above their genetic endowment, elevating the stock of human capital to the benefit of society as a whole. However since genetically and materially wealthy parents can invest more in their children than less well-endowed parents, this can be seen to run counter to a societal social justice imperative.¹

Sen (2009) and Atkinson (2012) have argued that public policy seeks progressive reform rather than transcendental optimality so that in the present context it should be directed toward some combination of progressively equalizing outcomes of individuals with the same effort and choices or progressively equalizing unambiguous inequalities in their circumstances. 21st century educational reforms in Germany², clearly took both directions in targeting low educational status parents (language education for immigrant parents for example) in addition to their child’s educational outcomes (through changes in curricula, the training of teachers and in the way children were taught and tracked). Since reform is progressive, evaluation techniques should be capable of measuring the degree and significance of any advances toward or retreats from the equal opportunity goal. Here new techniques are proposed and implemented in an analysis

¹This prompted development of the notion of qualified equality of opportunity (Anderson et al., 2014; Anderson and Leo, 2015) wherein the policy imperative is to seek independence of outcomes from circumstances without diminishing the outcomes of any circumstance class, i.e. only upward equalizing movements were acceptable. Such a policy would satisfy median voter theorem requirements in a democracy.

²The reforms were prompted by the results of the first Program for International Student Assessment (PISA) which “shocked the German nation” (OECD, 2011, p.208). The country came well below the average overall for all the countries tested in language and it did no better in mathematics and science. Correlations between its family socio-economic status and student achievement was higher than any other OECD country. When matched on actual achievement, 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 parents who had graduated from the lowest school level. In short there was overwhelming evidence of a dependence of a child’s educational outcomes on the circumstances they confronted, an evident lack of equality of opportunity.

of the extent to which the structure of the relationship between student achievements and their parental circumstances has changed in Germany in the 21st century.

Unfortunately much in the nature of individual choice, effort and circumstance is fundamentally multidimensional and unobservable. The vast literature on functionings and capabilities suggests that agents with the same choices and circumstances could have very different achievements (because of variation in effort) that are only partially related to their circumstance differences. Similarly observable circumstances are not deterministically related to their fundamental circumstances but more generally they are drawn from a circumstance class (for example genetic endowments are not an inconsequential component of a child's circumstances as is the level of nurturing of the child which is really a matter of parental effort). This inevitably involves the classification of individual achievements and circumstances into classes so that analytically, it is the extent to which membership of an achievement class is predicated upon membership of a given circumstance class that is of interest.

Conventionally, classifying agents into groups has invariably involved specifying "arbitrary" boundaries (for example income quantiles in the income mobility literature, high school grade levels in the educational literature and the poverty frontier in the poorness literature) which have been a matter of much debate.³ Sen and in addition many many others (e.g. Grusky and Kanbur, 2006; Kakwani and Silber, 2008; Alkire and Foster, 2011; Fituossi et al., 2011; Nussbaum, 2011) have forcibly argued that achievement is a many dimensioned concept and that the fundamentally unobservable or hard to quantify limitations to people's functionings and capabilities are determining or bounding factors. In such circumstances, boundaries become become vague so that agents with the same achievements may come from ostensibly different classes (Anderson et al., 2016), essentially, in the context of an achievement distribution, class boundaries become fuzzy, their number variable and membership only partially determined.

Regrettably when educational outcomes, the object of measurement, lose cardinal comparability many of the econometric techniques for studying the extent to which things have changed lose their effectiveness⁴ and other means of comparison have to

³See for example Sen, (1983), Townsend, (1985), Citro and Michael, (1995), Beach et al. (1997), Foster, (1998), Easterly, (2001), Milanovic and Yitzhaki, (2002), Saez and Veall, (2005), Banerjee and Duflo, (2008), Atkinson and Brandolini, (2013).

⁴For example intertemporal dominance comparisons of outcome distributions of circumstance classes, as proposed in Dardanoni (1993), Lefranc et al. (2009), Dardanoni et al. (2010), are no longer viable.

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 test scores in 2003 were not cardinally comparable with test scores in 2009. The question is how to measure progress in this particular context when test outcomes before and after the reforms cannot be cardinally compared? The primary contribution of this paper is to answer this question by introducing techniques to check if the transition processes from parental circumstances to child outcomes have become more equalizing or polarizing in an equality of opportunity sense.

To avoid arbitrary specification of the number of the classes and their boundaries, student achievement is associated with latent effort/choice variables in a semi-parametric finite mixture model⁵ 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.⁶

Finally new techniques for evaluating the polarizing, converging and directional

⁵Finite mixture models have featured in many fields where heterogeneity of individual types is an issue or where data contamination, misclassification or dynamic regime switching is an issue. See, for example, Eckstein and Wolpin (1990), Keane and Wolpin (1997), Kim and Nelson (1999), Lewbel (2007), Chen et al. (2011). Recently non-parametric identification and estimation of these models has been considered (Henry et al., 2014; Kasahara and Shimotsu, 2014) however here a semi-parametric approach is employed whereby identification is achieved by appealing to central limit theorems to pre-specify the parametric structure of the sub-distributions.

⁶Versions of mixture models have been used in pursuit of the notion of intertemporal transition that has been popular since the seminal work of Quah (1993, 1997) to analyze mobility and convergence in the world size distribution of GDP per capita (see, for example, Paap and Van Dijk, 1998; Bloom et al., 2003; Alfó et al., 2008; Pittau et al., 2010) but to the authors knowledge they have not been used in the context of inter-generational relationships nor in the situation where departure and arrival states differ in nature and number.

nature of the transition matrix are introduced. These techniques are implemented using the PISA data set for Germany for the years 2003 and 2009 i.e. the year before the reforms were introduced and the year sometime after the reforms were introduced. In Section 2 the basic model and the approach to determining outcome and circumstance classes is developed along with the tools for analysis of the extent to which equality of opportunity has progressed over the period. Section 3 provides background context to the reforms in Germany over the 2003–2009 period. Section 4 reports the main results and Section 5 draws some conclusions.

2 The model and tools for evaluating transitional states

PISA data for German 15-year-old⁷ students 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 ($X1$), Language ($X2$), and Science ($X3$). Evaluation of student progress within the German system follows the usual practice of averaging student grades across disciplines (the Durchschnittsnote, usually measured on a 6 point scale) to get an overall grade point average. That practice is followed here but, because the evaluation methodology in the different disciplines within the context of PISA differed both between disciplines and over the two observation periods, the overall achievement index in each year was based upon the average of their maximum mark standardized subject scores i.e. for person i :

$$\text{Achievement Score}(i) = \frac{1}{3} \cdot \left\{ \frac{X1_i}{\text{Max}(X1)} + \frac{X2_i}{\text{Max}(X2)} + \frac{X3_i}{\text{Max}(X3)} \right\}.$$

While this outcome index is cardinally comparable within year, it is not cardinally comparable between years. What can be compared is the anatomy of the transition process from circumstance classes to outcome classes. For this it is necessary to classify the students into achievement groups.

Given that observable student outcomes (aggregate achievement scores) are governed by three unobservable factors, their innate abilities, choices and effort, it is assumed that, in a particular generation, there are a finite number K of student

⁷Actually PISA covered students aged between 15 years 3 months and 16 years 2 months at the time of the assessment who have completed at least six years of formal schooling.

achievement types or classes labeled $k = 1, \dots, K$, where “ k type” students have similar abilities and choices, while effort is randomly distributed across the same type. Following Gibrat’s law (Sutton, 1997) suppose that some monotonic non-decreasing function of the achievement score for a student in achievement class k in period t , for convenience assume it is $\exp(x_{tk})$, follows the law of proportionate effects with v_{tk} its outcome improvement rate in period t , and let T be the elapsed time period over which the student has progressed with $\exp(x_{0k})$ the initial achievement level. Assuming the v ’s to be independent and identically distributed random variables with a small (relative to one) mean δ_k and finite variance σ_k^2 which vary with type, it may be shown that for an elapsed schooling period of T , the log achievement size distribution of such students would be linked systematically from period to period in terms of means and variances in the form:

$$x_{Tk} \sim N(x_{0k} + T(\delta_k + 0.5\sigma_k^2), T\sigma_k^2), \quad k = 1, \dots, K.$$

Note that the distribution is governed by the initial condition x_{0k} and the growth rate δ_k which in turn are dependent on the circumstances, innate abilities as well as efforts of the student. This does imply that the achievement distributions of different classes overlap so that the achievement of a high effort student from a low achievement class could exceed that of a low effort student from a higher achievement class. The size distribution of achievements of a collection of students will be a weighted sum of these achievement class distributions where the weights equal the proportions of the student population in the corresponding classes. The unobservable factors are, to some degree, influenced by a student’s partially observed circumstances, the nature and nurture effects of social and parental background, thus it is assumed that the chance that a student is in a particular achievement class is partially determined by her circumstances. Given the probability that a student with achievement x is in class k and the knowledge that she is from a particular circumstance class facilitates study of the relationship between achievements and circumstances.

The only data available for parental circumstance is family type (one or two parents present) and the educational status of the parents, family income data is not available, further it should be noted 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)). So to develop circumstance classes that reflect a student’s parental environment (circumstances) we construct an index by adding the

educational status (a six point scale) of each parent present in the household and divide 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 circumstance 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.⁸

2.1 Mixture models and class membership probabilities

For generality purposes, suppose K achievement classes emanating from J circumstance classes are contemplated. Achievements are continuously measured on the unit interval, while circumstances are discrete ordered categories. For a given achievement class k the achievement of student i , x_i , may be approximately 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 latent measure of student i 's effort. Thus, the distribution of x is given by:

$$f(x, \Psi) = \sum_{k=1}^K w_k f_k(x, \theta_k), \quad (1)$$

where $f_k(x, \theta_k)$ are $N(\mu_k, \sigma_k)$.⁹ The vector $\Psi = (w_1, \dots, w_{K-1}, \xi')$ contains all the unknown parameters of the mixture model: w_k , $k = 1, \dots, K$ are the mixing proportions summing to 1 ($\sum_{k=1}^K w_k = 1$); the vector ξ contains all the parameters $(\theta_1, \dots, \theta_K)$ 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. The conditional probability of a student i ($i = 1, \dots, n$) with achievement x_i being in achievement class k ($k = 1, \dots, K$) is given by:

⁸It is possible to treat circumstances in a similar semiparametric fashion but the methodology for computing parental circumstances yielded a variable which was effectively discrete with a sample of 2459 parents distributed over less than 20 support points

⁹The 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 (Marron and Wand, 1992; Rossi, 2014).

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

where $A(i)$ indicates the achievement class component to which student i belongs, yielding the probability of achievement for each student i to belong to the mixture component k .

In estimating the parameters, the class weights (the unconditional probabilities), are estimated by using the individual class weights π_{ik}^A as:

$$\hat{\pi}_k^A = \hat{w}_k = \frac{1}{n} \sum_{i=1}^n \pi_{ik}^A, \quad k = 1, \dots, K. \quad (3)$$

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). Starting from a given number of components and an initial parameter $\Psi^{(0)}$, the first stage of the algorithm (E-step) is to assign to each data point its current conditional probabilities. In the second stage (M-step), the maximum likelihood estimates are computed using the conditional probabilities as conditional mixing weights. The estimates of the parameters are used to re-attribute a set of improved probabilities of group membership and the sequence of alternate E and M steps continues until a satisfactory degree of convergence occurs to the ML estimates.¹⁰

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

$$\hat{\pi}_j^c = \frac{1}{n} \sum_{i=1}^n D_{ij} \quad (4)$$

where:

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

¹⁰It 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 Ψ should be the root of the likelihood equation corresponding to the largest of the local maxima located. The solution usually adopted is to apply a range of starting solutions for the iterations. In this paper, randomly selected starts, large sample non-hierarchical (Kaufman and Rousseeuw, 1990) clustering-based starts have been selected for initialization.

Let the $K \times J$ matrix T whose typical element is $t_{kj}(k = 1, \dots, K; j = 1, \dots, J)$ be the transition matrix yielding the conditional probability of being in achievement class k given circumstance class j . The $K \times 1$ 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:

$$\pi^A = T \cdot \pi^C \quad (5)$$

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

$$\pi_i^A = T \cdot D_i + \nu_i, \quad i = 1, \dots, n \quad (6)$$

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 $J \times 1$ vector whose typical element is D_{ij} and ν_i is a $K \times 1$ random vector with zero mean and singular covariance matrix. Thus $\pi^A = E(\pi_i^A) = E(TD_i + \nu_i) = T \cdot E(D_i) = T \cdot \pi^C$, where the columns of T sum to 1. The resulting estimates of conditional probabilities T are given by:¹¹

$$\hat{T}_{kj} = \frac{1}{|C(i) = j|} \sum_{i \in C(i)=j} \pi_{ik}^A.$$

2.2 Choosing the number of achievement classes

It has been argued that 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.

Selection of K for the achievement distribution is performed by minimizing the proximity of the mixture distribution, $f(x, \Psi)$, to a kernel estimate of the distribution, $f_{kern}(x)$, of achievements 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.

¹¹Estimating T can be seen to be a matter of estimating $M \text{diag}(\pi^C)^{-1}$, where M is the joint probability distribution of achievements and circumstances and $\text{diag}(\pi^C)$ is a matrix with circumstance class probabilities π^C on the diagonal, so that if required M can be retrieved by post multiplying T by $\text{diag}(\pi^C)$ or estimated directly by multiplying each of the circumstance dummies by the inverse of the probability of being in that circumstance class.

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.¹² In particular GTR relates to the overlap measure, $\theta = \int_{-\infty}^{\infty} \min \{f(x, \Psi), f_{krn}(x)\} dx$:

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

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

GTRIM is an importance weighted version of GTR:

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

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. It can be seen as weighting the difference of $f(x, \Psi)$ from the “target” $f_{krn}(x)$ by some monotonic function of the “target” function, 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. In essence, the differences are weighted with respect to some reference distribution. Choice of the square root function is very much inspired by, and in the spirit of, entropic measures of variation such as Theil, Pearson and Shannon. For example the continuous version of Theil’s entropic measure (TE) is related to the continuous version of Pearsons Chi squared (PCHI) dissimilarity measure as follows:

$$\begin{aligned} \text{TE} &= \int f(x, \Psi) \ln \left(\frac{f(x, \Psi)}{f_{krn}(x)} \right) dx = \int f(x, \Psi) \ln \left(1 + \frac{f(x, \Psi) - f_{krn}(x)}{f_{krn}(x)} \right) dx \\ &\approx \int f(x, \Psi) \left(\frac{f(x, \Psi) - f_{krn}(x)}{f_{krn}(x)} \right) dx = \int \left(\frac{[f(x, \Psi) - f_{krn}(x)]^2}{f_{krn}(x)} \right) dx = \text{PCHI}. \end{aligned}$$

The argument under the integral sign in GTRIM can be seen to be the square root of the argument under the integral sign of PCHI. The optimal number of components in the mixture is consequently assessed comparing the mixture distribution with the true

¹²Assuming 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)

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.

Once the number of classes has been determined, a measure of the degree of separation between the classes is of interest. In the present context it will indicate the extent to which the changes in *curricula*, testing methods and tracking policies have separated the ability classes. In essence, increased separation reflects increased within class homogeneity or better identified group classification. For this purpose one minus the average overlap of contiguous classes is a useful measure.¹³ When contiguous lower and upper class distributions, $f_l(x)$ and $f_u(x)$ respectively, do not overlap (i.e. $\theta = \int \min(f_l(x), f_u(x)) dx = 0$) they are perfectly segmented in the terminology of Yitzhaki (1994). When they overlap perfectly (i.e. $\theta = \int \min(f_l(x), f_u(x)) dx = 1$) probabilities of class membership cannot be identified. When all classes are perfectly segmented the average of these overlaps will be 0 and one minus the average overlap will be 1. When all classes overlap perfectly and cannot be separately identified it will be 0. In contrast the extent of overlap of the extreme distributions yields a lower bound to the average overlap and one minus this value will reflect the extent to which the extremes of the performance distribution have polarized or separated. Perhaps more importantly it gives an idea of the extent to which there is some convergence or equalization of the highest and lowest outcome distributions. The distributions of various estimates of θ have been found to be normal (Anderson et al., 2012; Anderson et al., 2013) so that averages of them will also be normal. When $f_l(x)$ and $f_u(x)$ intersect once at x^* the overlap is given by $\int_{-\infty}^{x^*} f_u(x) dx + \int_{x^*}^{\infty} f_l(x) dx$. For weighted distributions the overlap is given by $\int_{-\infty}^{x^*} w_u f_u(x) dx + \int_{x^*}^{\infty} w_l f_l(x) dx$ where w_u and w_l are the respective upper and lower distribution weights.

In the present context where the sub distributions are weighted normal with mean and standard deviations equal to $\mu_l, \mu_u, \sigma_l, \sigma_u$ respectively, x^* is the solution to:

$$w_l \frac{e^{-\frac{(x^* - \mu_l)^2}{2\sigma_l^2}}}{\sqrt{2\pi\sigma_l^2}} = w_u \frac{e^{-\frac{(x^* - \mu_u)^2}{2\sigma_u^2}}}{\sqrt{2\pi\sigma_u^2}}. \quad (9)$$

If it exists it can be found as the value of the root residing between μ_l and μ_u in the following equation:

$$-\left(\frac{1}{\sigma_l^2} - \frac{1}{\sigma_u^2}\right) x^{*2} - 2\left(\frac{\mu_u}{\sigma_u^2} - \frac{\mu_l}{\sigma_l^2}\right) x^* + \left(\frac{\mu_u^2}{\sigma_u^2} - \frac{\mu_l^2}{\sigma_l^2}\right) - 2\ln\left(\frac{w_u\sigma_l}{w_l\sigma_u}\right) = 0. \quad (10)$$

¹³Distributional overlap or separation is closely related to the notion of polarization between groups (Anderson, 2004).

2.3 Evaluating generational transition patterns: new indices and methods

Since this framework allows for non-square transition matrices T (the numbers of circumstance J_C and achievement classes K_A are not necessarily the same), the Shorrocks (1978) suggestion of a simple index of mobility (equality of opportunity) as $(K - \text{trace}(T))/(K - 1)$, where K is the number of categories or dimension of T is not viable. When achievements are independent of circumstances, the joint probability matrix $M^* = \pi^A \cdot (\pi^C)'$ and $T^* = M^* \text{diag}(\pi^C)^{-1}$. An index of general equality of opportunity (Anderson and Leo, 2015) is afforded by the degree of overlap between M and M^* :

$$\text{OVLP} = \sum_{k=1}^{K_A} \sum_{j=1}^{J_C} \min(M_{kj}, M_{kj}^*)$$

Instead of estimating M , an index based upon T may be developed. The j th 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 k th row of T as t_k , let $\text{maxr}()$ and $\text{minr}()$ 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:

$$\text{TM}(T) = 1 - \frac{\sum_{k=1}^{K_A} (\text{maxr}(t_k) - \text{minr}(t_k))}{\min(J_C, K_A)}. \quad (11)$$

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 $\text{TM}(T)$ 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 $\text{TM}(T) = 1$.

In the case of perfect immobility, for each row $k \in K_A$ there at least one column j such that $t_{kj} = 0$, therefore the sum of minimums will be always equal to zero. However, we can distinguish three different situations. In all the situations $\text{TM}(T) = 0$,

1. The number of achievement classes is equal to the number of circumstances: $K_A = J_C$. The transition matrix T is squared and complete immobility is char-

acterized 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 and the set of these columns forms an orthonormal basis in \mathbb{R}^{K_A} . The sum of maximums will be equal to K_A , the number of independent conditional distributions, yielding $TM(T) = 0$.
3. The number of achievement classes is greater than the number of circumstances: $K_A > J_C$. In this case the columns of T are orthogonal, the final state outcome distributions do not overlap. The sum of the sum of maximums will be equal to J_C , the number of columns, yielding $TM(T) = 0$.

The index $TM(T)$ satisfies the normalization, immobility and perfect mobility axioms of Shorrocks (1978). It 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 $TM(T) > TM(\check{T})$ when $T_{kj} \geq \check{T}_{kj}$ for all $k \neq j$ with strict inequality holding somewhere, is satisfied.¹⁴ Period consistency requires $TM(T) \geq TM(\check{T})$ implies $TM(T^s) \geq TM(\check{T}^s)$ 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).

It has been suggested that the achievement of equality of opportunity can be assessed by establishing the absence of dominance relationships between the outcome distributions of circumstances (Dardanoni, 1993; Lefranc et al., 2009) but, in the absence of the achievement of the equal opportunity goal, little can be gleaned from this approach, that is to say the approach gives no sense of the progress toward, or away from, an equal opportunity goal. Anderson and Leo (2015) propose an index, which is asymptotically normally distributed with an estimable standard error, based upon the area between the dominance curves at a given order as a measure of distance from an

¹⁴The incremental increase in any off diagonal element requires a concomitant decrease in its corresponding column on diagonal element (to preserve adding up). Such a change can only decrease $\sum(\max r(t) - \min r(t))$ and hence increase $TM(T)$.

equal opportunity goal with reductions in the index indicating progress toward equality of opportunity. In essence this is, at the M th order of dominance, a measure of how far apart are the outcomes of the poorest parentally endowed children and the richest parentally endowed children. Unfortunately since, because of different testing methodologies, outcome distributions in the two observation years are not strictly commensurable only within year comparisons can be made.

2.4 Viewing the circumstance–achievement transition as a process

In the context of a generational model, where this generation’s achievement classes become the foundation for circumstance classes of a subsequent generation, the transition matrix can be viewed as characterizing a process. Anderson (2015) demonstrates how the transition matrix can be used to evaluate whether the transitions are converging or polarizing the outcome or achievement distribution via a “balance of probabilities” measure. In the present context a convergent transition matrix implies equalizing the circumstances for subsequent generations whereas a polarizing process can be seen as moving future circumstance classes further apart, thus reducing the chance of an equality of opportunity outcome.

For expositional simplicity suppose the transition matrix is aggregated into a 3×3 matrix with typical element T_{kj} $k, j = 1, 2, 3$ of transitions to low, middle and high achievements from low, middle, high circumstances, with circumstance probabilities π_j^c $j = 1, 2, 3$.¹⁵ 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. When PT is positive, a convergent process is indicated (polarizing when it is negative).

Following Anderson et al. (2014) and Anderson and Leo (2015), 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, equalizing upward as it were. One interpretation of this imperative is that it is a reconciliation of the private aspirations of parents to preserve, and extend the advantages that their offspring and

¹⁵Given more than 3 outcome and or circumstance groups it is a simple task to aggregate the transition matrix into a 3×3 grouping by aggregating the rows or columns of the matrix respectively. 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.

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. These statistics may be written as:

$$PT = wT_{21} + (1 - w)T_{23} - (T_{12} + T_{32}), \text{ where } w = \frac{\pi_1^c}{\pi_1^c + \pi_3^c};$$

$$PUT = (1 - T_{11})\pi_1^c - (1 - T_{33})\pi_3^c - (T_{12} - T_{32})\pi_2^c.$$

As balance of probability measures, these indices are bounded between -1 and 1 . If it is desired 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 $N(p, (p(1 - p)/n))$ thus facilitating tests of hypotheses of equalizing opportunities and upward mobility.

3 Education reform in Germany

Throughout the latter half of the 20th century Germany had uniform elementary schooling with compulsory Grundschule education for all children aged 6 through 10. On completion students were tracked into one of three types of school, Hauptschule, Realschule or Gymnasium.¹⁶ The vast majority of students of lowest abilities were streamed into Hauptschule where, after a few more years of education, they would receive a qualification entitling them to apply for training generally leading to low skilled

¹⁶School tracking in Germany has been the object of a considerable amount of study, for an excellent survey see Krause and Schuller (2014). Generally tracking is thought to be deleterious to educational attainment in imposing artificial barriers to progress and achievement though, following Hanushek (1986), there are arguments for teaching efficiencies in that it is a more effective use of teaching resources to teach more homogeneous groups of children. It has also long been viewed as an institutional device reinforcing the intergenerational persistence in educational achievements across different social classes (Mare, 1981; Ammermüller, 2005; Schuetz, Ursprung and Wößmann, 2005; Hanushek and Wößmann, 2006) which is in some sense contrary to an equal opportunity goal. Brunello and Checchi (2006) looking at early work time wages find that parental background effects are stronger when tracking starts earlier (although see Malamud and Pop-Eleches, 2011).

blue collar jobs. This training was often obtained after grade 9 between the ages 15 and 17 through vocational training at a Berufsschule, a combination of apprenticeship and classes. More qualified students enrolled into the Realschule generally attending for at least one extra year after which they got vocational training at a Berufsfachschule, tailored to a more specific training of clerks, technicians and lower-level civil servants.¹⁷ The Gymnasium, the highest secondary school, focused on broad preparation in the humanities and the Abitur, which was the sole gateway to the professions, teaching and the upper levels of the civil service. These students completed preparatory classes for university and college and attended Fachoberschule after grade 10, a specialized high school.

The school divisions reflect the social divisions in Germany all very much in concert with the view that tracking reinforces the impact of family background. The Gast Arbeiter program solution to labor shortages in the latter quarter of the 20th century saw a considerable influx of workers from countries with comparatively low levels of education. Many settled and raised families and those successive generations typically had poor German language skills. Increasing demands for high skilled workers led increasing demand for entry into the Gymnasium and Realschule and consequently the Hauptschule was confined to students with limited perspectives, that include immigrants and native German children from lower class families.

In response to the concerns raised by the 2000 PISA results, the “Kultusminister Konferenz (KMK)”, formed by ministers of education in the 16 German States, proposed seven central areas (*sieben zentrale Handlungsfelder*) requiring changes (PISA 2000, 2002). These were: 1) Improve speaking capabilities of preschoolers; 2) Earlier enrollment into Grundschule; 3) Improved curriculum with regard to reading capabilities, mathematical skills and understanding of natural sciences; 4) Direct help with learning difficulties for youths with an immigration background; 5) Implement “consequential” teaching for teachers to promote uniform testing and 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 an all-day school program to provide more extensive education, specifically to students with learning difficulties or special skills (Avenarius et al., 2003; PISA 2000,

¹⁷This secondary schooling system was augmented by what is referred to as the “Dual System” managed at the federal level whereby students who complete a secondary education can be apprenticed to a firm whilst simultaneously attending a vocational “continuation school”. Until recently, the Hauptschule was the primary source of employee’s apprentices with some white collar jobs going to students from the Realschule.

2002). The timing and progression of each reform was left to the state, thus, the full implementation of each of the above-mentioned central fields differed across states.¹⁸

All-day school programs were embarked upon in 2003 extending the school day until 4:00 pm or later.¹⁹ In 2003 and 2004 national educational standards were introduced for children in primary and secondary school in German language, mathematics, a first foreign language (English or French), and science (biology, chemistry and physics). Additional standards were put in place for students at the end of grade 10 in 2007. These performance standards covered subject-specific competencies at a similar level as the PISA tests that students were expected to meet throughout Germany. Prior to this now mandatory agreement, there had never been national standards in Germany.

In 2006, the Council of ministers agreed to develop common assessments based on a national scale for 3rd graders in elementary school, and 8th and 9th graders in secondary schools in all 16 German states. To make the yearly spring tests more comprehensive some states joined forces to develop testing systems that obey the new curriculum standards.

To increase the chance of a student attending a higher secondary school some states delayed tracking into the tripartite system, Hauptschule, Realschule or Gymnasium, until the student was 12 rather than 10 years old. Other states combined the Realschule and Hauptschule into one school, while some allowed students in lower schools to move up the ladder and complete their education with a more prestigious background, allowing for better job-opportunities all of which, following Brunello and Checchi (2006), would have been expected to dilute the influence of parental background. This led to speculations in 2008/2009 that the 2,625 Realschulen and 4,283 Hauptschulen will 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 of the other two school types combined. Implementation of these new policies was helped by the Institute for Educational Progress (Institut zur Qualitätsentwicklung im Bildungswesen (IQB)) based at the Humboldt University in Berlin which monitored the implementation process. These reforms brought fundamental change to the old structure of the schools, which had focused on having few highly educated people, several with medium education and the majority with little education.

¹⁸More details on when they were started and how each state has progressed can be found on the states respective website, or an outline of the very first stages of implementation can be found in Avenarius et al., 2003.

¹⁹Prior to this increase, German students aged 7 and 8 spent an average of 626 hours a year in school, compared to 788 in the OECD countries.

4 Empirical evidence

The Program for International Student Assessment (PISA) results for Germany in the years 2003 and 2009 were employed with only students who attempted all three competencies studied. Table 1 reports summary statistics of the raw data and the constructed achievement and circumstance variables to be used in this study.

Table 1: *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
Family Type	4.000	1.000	1.978	2.000	0.521	3.000	1.000	1.854	2.000	0.372
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 since, as noted in the introduction, changes in national standards, curriculum, teaching and testing methods were implemented in the intervening period making achievement scores not cardinally comparable. Essentially nothing can be deduced from the fact that mean and median scores have gone up in Math and down in Reading and Science over the period. Note however that the math score distributions have switched from being right skewed to left skewed in distribution over the period with the reading score distribution going in the opposite direction. Basically the achievement variable to be used in this study shows a slight decline in the mean and median with a reduction in the spread. To facilitate parsimony the circumstance variable was used to group the parental condition into three ordered categories: Lower, Middle and Upper. The circumstance class cutoffs were set at 4.5 and 6.6 for both observation years. Table 2 reports the parental class sizes that result from the categorization process.

Given identical class boundaries in both periods the circumstance distribution in 2009 stochastically dominates at the first order that of the 2003 period. Membership of the lower circumstance class has reduced significantly and membership of the Middle

Table 2: *Circumstance class sizes*

	2003	2009
Lower	0.430	0.367
Middle	0.343	0.391
Upper	0.227	0.242

and Upper circumstance class has increased. The policy of elevating parental educational status seems to have worked. Turning to the determination of the achievement groups, the results of the various versions of the group number selection criteria and distribution polarization measures are reported in Table 3. Visual representations of kernel²⁰ 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 and distribution polarization measures.*

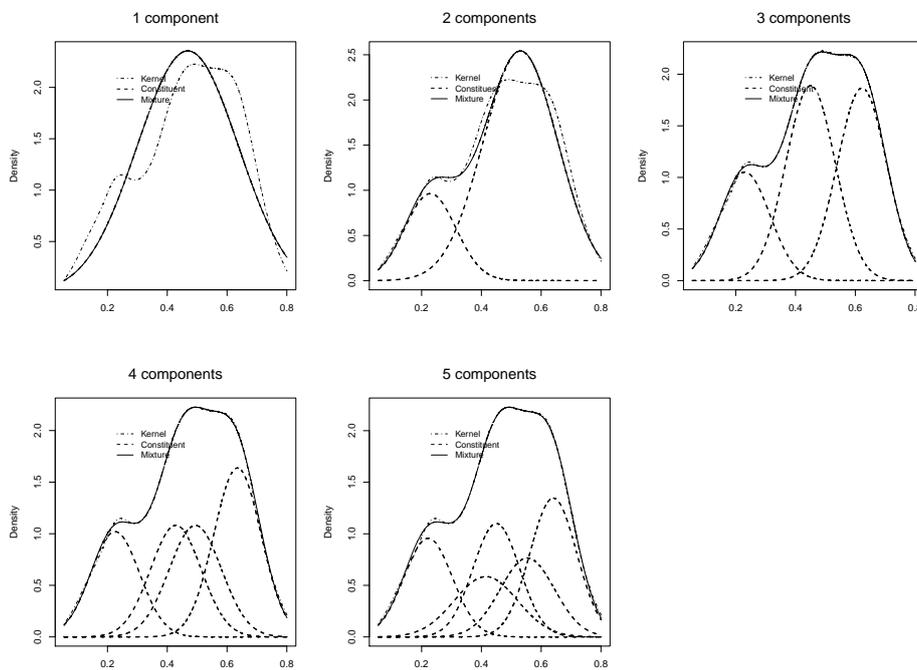
No. classes	Year	GTR	GTR st.error	GTR + penalty	GTRIM	GTRIM+ penalty	Avg Pol.	Avg.Pol. st.error	Pol.Ext	Pol.Ext st.error
2	2003	0.113	0.005	0.127	0.097	0.113	0.945	0.008	0.945	0.008
	2009	0.072	0.002	0.079	0.063	0.073	0.899	0.007	0.899	0.007
3	2003	0.062	0.004	0.084	0.070	0.090	0.932	0.009	0.997	0.002
	2009	0.051	0.002	0.062	0.055	0.067	0.897	0.007	0.982	0.003
4	2003	0.057	0.004	0.086	0.064	0.094	0.904	0.010	0.999	0.001
	2009	0.040	0.002	0.055	0.047	0.065	0.910	0.007	0.997	0.001
5	2003	0.058	0.004	0.094	0.063	0.101	0.906	0.010	0.999	0.001
	2009	0.065	0.002	0.083	0.066	0.082	0.950	0.005	0.999	0.000

Note that for all component comparisons GTR measures are significantly different at conventional levels of significance with the exception of the 4 *versus* 5 components 2003 comparison and all between year comparisons are significantly different. The average polarization statistics all suggest good separation between the subgroups for all choices of number of classes in both years. 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

²⁰A Gaussian kernel density estimator was employed. Following Jones et al., (1996), the bandwidth has been estimated using the plug-in procedure of Sheather and Jones (1991).

for 2009.²¹ Evidently, as anticipated, the raft of policy measures in the intervening period appears to have changed the number of achievement classes between the two observation periods. The polarization of extremes measure suggests little change in the proximity of the highest and lowest classes in either year. However the polarization measure based upon the average degree of separation suggests that there is significantly less separation between contiguous groups in 2009 than there was in 2003 in the three group classification though there is no significant differences in the four group classification.

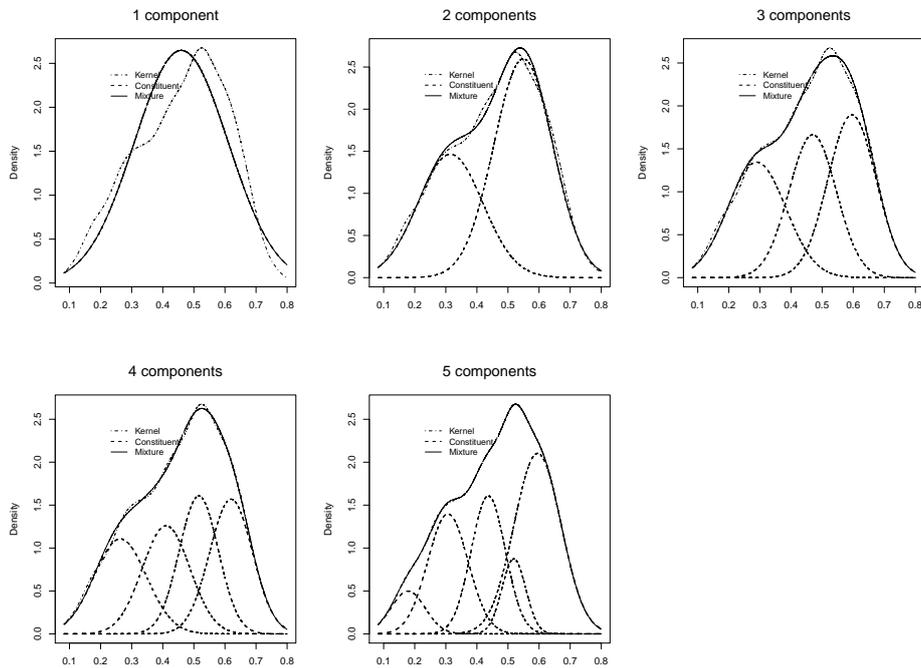
Figure 1: *Kernel and mixture estimation of achievement in Germany: year 2003.*



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. Though, because of changes in standards, methods, teaching technology *etc.* between the observation periods, the achievement scores are not inter-temporally comparable, it is an interesting exercise to make such comparisons. 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 mean has how-

²¹Widely-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.

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



ever improved substantially from 0.231 to 0.263 (the standard normal statistic for the difference is 4.628). 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.

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

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

For the purposes of establishing the presence of equality of opportunity, Lefranc et al. (2009) advocate examining the 2nd order stochastic dominance relationships of achievement distributions conditional on circumstance classes with the absence of dominance providing evidence of equality of opportunity. Because of a lack of cardinal comparability dominance comparisons cannot be made between the years but they can be made between classes within years. 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 achievement distributions of higher class circumstance groups always dominate those of lower class circumstance groups for all pairings in all 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	0.125
Min	0.000	0.000	0.000	0.000	0.000	0.000
Area	0.097	0.062	0.035	0.084	0.048	0.036

Turning to the circumstance to achievement transition matrices, presented in Table 6, we see that the nature of the transition process has clearly changed over the period with the emergence of an additional achievement class. Furthermore, mobility has improved appreciably over the period: The TM mobility index moved from 0.804 in 2003 to 0.852 in 2009. The increase of 0.048 is significant at all conventional levels of significance (see Table 7). Transforming the converging-polarizing balance of probabilities measure to the 0-1 index, ($PT^* < 0.5$ indicating polarizing processes, $PT^* > 0.5$ indicating converging) observe that the transitions were significantly polarizing in 2003 (i.e. preserving and reinforcing class stratification in subsequent generations) and significantly convergent in 2009 (reducing inheritance class differences for future generations). The transformed upward mobility index PUT^* shows that both 2003 and 2009 transition processes were upwardly mobile, favoring the view that movements toward a state of equality of opportunity were qualified by an attempt to preserve the status of well-endowed students. However, the propensity for upward mobility had dimin-

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

	2003 TM(M)=0.804		
	Circumstance 1	Circumstance 2	Circumstance 3
Low achievement group	0.320 (16.528)	0.169 (7.802)	0.108 (4.071)
Middle achievement group	0.416 (21.103)	0.399 (18.065)	0.334 (12.311)
High achievement group	0.265 (12.697)	0.432 (18.496)	0.558 (19.449)
	2009 TM(M)=0.852		
	Circumstance 1	Circumstance 2	Circumstance 3
Low achievement group	0.338 (23.960)	0.191 (14.018)	0.144 (8.274)
Lower-middle achievement group	0.269 (25.806)	0.256 (25.424)	0.180 (14.072)
Upper-middle achievement group	0.174 (12.401)	0.269 (19.827)	0.419 (24.245)
High achievement group	0.219 (21.062)	0.283 (28.148)	0.257 (20.114)

Numbers in brackets indicate asymptotic t -values.

ished significantly in 2009 indicating a reduction in the rate of upward mobility or a diminution in attempts to preserve the status of the well-endowed.

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

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

5 Concluding Remarks

Poor student outcomes in Germany in the initial round of PISA scores and evidence of substantial generational dependencies in those outcomes prompted extensive educational reforms at the federal and state levels over the ensuing years with a view to advancing equality of opportunity. Unfortunately, from a measurement perspective, the substantial changes teaching, testing and tracking practices that took place over the period meant that student outcome distributions were not inter-temporally comparable in a cardinal sense. Since a pure equality of opportunity objective is unlikely ever to be attained, assessing the effectiveness of these reforms is more a question of measuring the progress toward the equality of opportunity goal rather than determining whether or not transcendental optimality has been achieved.

Accordingly new tools which facilitate a comparative examination of parental circumstance - child outcome transitional structures before and after the reforms in the context of an equality of opportunity goal have been proposed. Indices and tests for determining the number of classes in a mixture distribution, for determining whether generational transitions were polarizing or converging and for determining the degree and type of mobility were all implemented in the context of the German reforms using PISA data for the years 2003 (immediately prior to the reforms being implemented) and 2009 (after the reforms had been implemented). Some progress toward an equal opportunity goal was detected which was the target of the reforms. Structural change has taken place with the emergence of an additional achievement class over the period and overall mobility appears to have improved significantly. The nature of the transition process was fundamentally transformed with the emergence of an additional achievement class. It also changed from a polarizing to a convergent process which can be seen to be equalizing the circumstances of future generations, though this was at the expense of some upward mobility (in essence high class inheritors became slightly more downwardly mobile).

References

- Arrow, K. J., Bowles, S., and Durlauf, S. N. (2000). *Meritocracy and Economic Inequality*. Princeton University Press.
- 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.
- Alfö, M., Trovato, G., and Waldmann, R. (2008). “Testing for Country Heterogeneity in Growth Models Using a Finite Mixture Approach”. *Journal of Applied Econometrics*, 23(4), 487–514.
- Alkire, S. and Foster, J.E. (2011). “Understandings and misunderstandings of multi-dimensional poverty measurement”, *Journal of Economic Inequality*, 9, 289–314.
- Ammermüller, A. 2005. “Educational Opportunities and the Role of Institutions”, ZEW Discussion Paper, No. 05–44.
- Anderson, G. (2004). “Toward an Empirical Analysis of Polarization”. *Journal of Econometrics*, 122, 1–26.
- Anderson, G. (2015). “Measuring Polarization and Convergence as Transitional Processes in the Absence of a Cardinal Ordering”. *Working Paper 547*, University of Toronto, Department of Economics.
- Anderson, G., Farcomeni, A., Pittau, M.G., & Zelli, R. (2016). “A new approach to measuring and studying the characteristics of class membership: The progress of poverty, inequality and polarization of income classes in urban China”. *Journal of Econometrics*, 191(2), 348–359.
- Anderson, G., Ge, Y., & Leo, T.W. (2013). “Distributional Overlap: Simple, Multivariate, Parametric and Nonparametric Tests for Alienation, Convergence and General Distributional Difference Issues”. *Econometric Reviews*, 29(3), 247–275.
- Anderson, G., Leo, T. W., and 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. (2015). “Quantifying the Progress of Economic and Social Justice: Charting Changes in Equality of Opportunity in the United States 1960-2000”. *Journal of Human Development and Capabilities*, forthcoming.
- Anderson G., Linton O., and Whang, Y-J. (2012). “Nonparametric estimation and inference about the overlap of two distributions”. *Journal of Econometrics*, 171(1), 1–23.
- Atkinson, A. B. (2012). “Public economics after the idea of justice”. *Journal of Human Development and Capabilities*, 13(4), 521–536.
- Atkinson, A.B., and Brandolini, A. (2013). “On the identification of the middle class”, in *Income Inequality* (eds. J. C. Gornick and M. Jäntti), Stanford University Press,

77–100.

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

Banerjee, A.V., and Duflo, E. (2008). “What is middle class about the middle classes around the world?”, *Journal of Economic Perspectives*, 22(2), 3–28.

Beach, C.M., Chaykowski, R.P. and Slotsve, G.A. (1997). “Inequality and Polarization of Male Earnings in the United States, 1968-1990”, *North American Journal of Economics and Finance*, 8(2), 135–151.

Becker, G.S. and Tomes, N. (1986). “Family capital and the rise and fall of families”, *Journal of Labor Economics*, 4, S1–39.

Bloom, D.E., Canning, D., and Sevilla, J. (2003). “Geography and Poverty Traps”. *Journal of Economic Growth*, 8, 355–78.

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

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

Chen, X., Hong, H., and Nekipelov, D. (2011), “Nonlinear models of measurement errors”. *Journal of Economic Literature*, 49, 901–937.

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

Citro, C.F. and Michael, R.T. (1995). *Measuring Poverty: A New Approach*. National Academy Press, Washington, DC.

Dardanoni, V. (1993). “On Measuring Social Mobility”, *Journal of Economic Theory*, 61,372–394.

Dardanoni, V., Fiorini M., and Forcina, A. (2010). “Stochastic Monotonicity in Inter-generational Mobility Tables”. *Journal of Applied Econometrics*, 27(1), 85–107.

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

Easterly, W. (2001). “The Middle Class Consensus and Economic Development”, *Journal of Economic Growth*, 6(4), 317–335.

- Eckstein, Z. and Wolpin, K. (1990). “Estimating a market equilibrium search model from panel data on individuals.” *Econometrica*, 58, 783–808.
- Fitoussi J-P, Sen, A. and Stiglitz, J.E. (2011). *Mismeasuring Our Lives: Why GDP Doesn't Add Up*, The New Press, New York, London.
- Foster, J.E. (1998). “Absolute versus Relative Poverty”, *American Economic Review*, 88(2), Papers and Proceedings, 335–341.
- Gini, C. (1916). “Il concetto di transvariazione e le sue prime applicazioni”, *Giornale degli Economisti e Rivista di Statistica*. 1–55.
- Greenwood, P.E. and Nikulin, M.S. (1996). *A guide to chi-squared testing*. Wiley, New York.
- Grusky, D. and Kanbur, R. (eds.) (2006). *Poverty and Inequality*. Stanford University Press, Stanford, California.
- Hanushek, E. (1986). “The economics of schooling: production and efficiency in public schools”. *Journal of Economic Literature*, 24, 1141–1177.
- Hanushek, E. and Wößmann L. (2006). “Does Educational Tracking Affect Performance and Inequality? Differences-In-Differences Evidence Across Countries”. *Economic Journal*, 116, C63–C76.
- Henry, M., Kitamura Y., and Salanié, B. (2014). “Partial identification of finite mixtures in econometric models”. *Quantitative Economics*, 5, 123–144.
- Jones, M.C., Marron, J.S. and Sheather, S.J. (1996). “A brief survey of bandwidth selection for density estimation”, *Journal of the American Statistical Association*, 91, 401–407.
- Kakwani, N. and Silber, J. (2008). *Quantitative Approaches to Multidimensional Poverty Measurement*. Palgrave Macmillan, London.
- Kasahara H., and Shimotsu, K. (2014). “Non-parametric identification and estimation of the number of components in multivariate mixtures”. *Journal Royal Statistical Society B*, 76, Part 1, 97–111.
- Kaufman, L. and Rousseeuw P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.
- Keane, M. and Wolpin, K. (1997). “The career decisions of young men”. *Journal of Political Economy*, 105, 473–522.
- Kim, C.J., and Nelson C. (1999). *State-Space Models With Regime-Switching: Classical and Gibbs-Sampling Approaches With Applications*. MIT Press, Cambridge.

- Kobus, M, and Milos P. (2012). “Inequality decomposition by population subgroups for ordinal data”, *Journal of Health Economics*, 31, 15–21.
- Krause, A., and Schüller S. (2014). “Evidence and Persistence of Education Inequality in an Early-Tracking System-The German Case-”. FBK-IRVAPP Working Paper No. 2014-07.
- Lewbel, A. (2007). “Estimation of average treatment effects with misclassification”. *Econometrica*, 75, 537–551.
- Lefranc, A., Pistolesi, N., and Trannoy, A. (2009). “Equality of Opportunity and Luck: Definitions and Testable Conditions, with an Application to Income in France”. *Journal of Public Economics*, 93, 1189–1207.
- Malamud, O. and Pop-Eleches C. (2011). “School tracking and access to higher education among disadvantaged groups”. *Journal of Public Economics*, 95, 1538–1549.
- Mare, R. D. (1981). “Change and Stability in Educational Stratification”. *American Sociological Review*, 46, 72–87.
- Marron, J.S. and Wand, M.P. (1992). “Exact mean integrated squared error”, *The Annals of Statistics*, 20, 712–736.
- McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley, New York.
- Milanovic, B. and Yitzhaki, S. (2002). “Decomposing World Income Distribution: Does The World Have A Middle Class?”. *Review of Income and Wealth*, 48, 155–178.
- Nussbaum, M.C. (2011). *Creating Capabilities: The Human Development Approach*, Harvard University Press.
- OECD (2011). *Strong Performers and Successful Reformers in Education: Lessons from PISA for the United States*. OECD Publishing, Paris.
- Paap, R. and van Dijk, H.K. (1998). “Distribution and mobility of wealth of nations”. *European Economic Review*, 42, 1269–1293.
- 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>.
- Pittau, M.G., Zelli, R. & Johnson, P.A. (2010). Mixture Models, Convergence Clubs and Polarization. *Review of Income and Wealth*, 56, 102–122.
- Quah, D. (1993). Empirical Cross-Section Dynamics in Economic Growth. *European Economic Review*, 37, 426–434.

- Quah, D. (1997). "Empirics for Growth and Distribution: Polarization, Stratification, and Convergence Clubs," *Journal of Economic Growth*, 2, 27–59.
- Roemer, J. E. (1998). *Equality of Opportunity*. Cambridge, Massachusetts: Harvard University Press.
- Roemer, J. E. (2006). *Democracy, Education and Equality: Graz Schumpeter Lectures*. Cambridge University Press. Econometric Society Monographs No. 40.
- 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., and Zubiri, I. (2003). "To What Extent do Fiscal Regimes Equalize Opportunities for Income Acquisition Among Citizens". *Journal of Public Economics*, 87, 539–565.
- Rossi. P.E. (2014). *Bayesian Non and Semi-parametric Methods and Applications*. Princeton University Press.
- Saez, E. and Veall, M.R. (2005). "The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence", *American Economic Review*, 95(3), 831–849.
- Schuetz, G., Ursprung H. and Wößmann L. (2005). "Education policy and equality of opportunity." CESIFO Working Paper No. 1518.
- Sen, A.K. (1983). "Poor Relatively Speaking", *Oxford Economic Papers*, 35, 153–169.
- Sen, A.K. (2009). *The Idea of Justice*. Harvard University Press.
- Sheather, S.J. and 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.
- Townsend, (1985). "A Sociological Approach to the Measurement of Poverty - A Rejoinder to Professor Amartya Sen", *Oxford Economic Papers*, 37, 659–668.
- Yitzhaki, S. (1994). Economic Distance and overlapping of distributions. *Journal of Econometrics*, 61, 147–159.