# Evaluating Education Systems

**Nicolas Gravel**
**Edward Levavasseur**
**Patrick Moyes**

# Evaluating Education Systems

Nicolas Gravel,[*] Edward Levavasseur[†]and Patrick Moyes[‡]

September 06th 2019

## Abstract

This paper proposes *two* dominance criteria for evaluating *education systems* described as joint distributions of the pupils' cognitive skill achievements and family backgrounds. The first criterion is shown to be the smallest transitive ranking of education systems compatible with *three* elementary principles. The *first* principle requires any improvement in the cognitive skill of a child with a given family background to be recorded favorably. The *second* principle demands that any child's cognitive skill be all the more favorably appraised as the child is coming from an unfavorable background. The *third* principle states that when two different skills and family backgrounds are allocated between two children, it is preferable that the high skill be given to the low background child than the other way around. The criterion considers system $A$ to be better than system $B$ when, for every pair of reference background and skill, the fraction of children with both a lower background *and* a better skill than the reference is larger in $A$ than in $B$. Our second criterion completes the first by adding to the three principles the *elitist* requirement that a mean-preserving spread in the skills of two children with the same background be recorded favorably. We apply our criteria to the ranking of education systems of 43 countries, taking the PISA score in mathematics as the measure of cognitive skills and the largest of the two parents International Socio Economic Index as the indicator of background. We show that, albeit incomplete, our criteria enables conclusive comparisons of about 19% of all the possible pairs of countries. Education systems of fast-growing Asian economies - in particular Vietnam - appear at the top of our rankings while those of relatively wealthy arabic countries such as Lebanon, United Arab Emirates and Jordan are at the bottom. The fraction of countries that can be ranked successfully happens to be only mildly increased as a result of adding elitism to the three other principles.

*JEL Classification Numbers*: D63, I21, I23, O57.

*Keywords*: Education, inequality, family background, opportunities, dominance, math scores, international comparisons.

"Schools are remarkably similar in the effect they have on the achievement of their pupils when the socioeconomic background of the students is taken into account."
James Coleman, 1967

---

[*]Centre de Sciences Humaines & Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE, 2, Dr. APJ. Abdul Kalam Road, 11 0011 Delhi, India, Nicolas.gravel@csh-delhi.com.

[†]Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE, 2, Boul. Maurice Bourdet, 13001 Marseille Cedex France, edward.levavasseur@yahoo.fr

[‡]CNRS, Université de Bordeaux IV, Patrick.Moyes@orange.fr

# 1 Introduction

All countries have put into place more or less compulsory *education systems.* These education systems, usually made of a mixture of public and private schools that follow specific learning curricula, take children at the age of five and enroll them in learning programs for about 10 to 12 years, depending upon the country. The result of this enrollment is the acquisition, by the children, of various *cognitive skills* that are of obvious importance for their future welfare. For one thing, the cognitive skills are major contributors of children' future earnings and employment opportunities (see e.g. Hanushek and Woessmann (2008), Hanushek, Schwerdt, and Woessmann (2015) or Nickell (2004)). But the acquisition of cognitive skills in mathematics, literacy, etc. may also impact individuals' well-being in a way that is not reducible to their pecuniary consequences, however important these may be. As noticed by many (for example Oreopoulos and Salvanes (2011)) cognitive skills may indeed foster future information acquisition, and help individuals to make better decisions about health, spouse partnership, parental choices, etc.

There are by now a few internationally standardized procedures for gathering data on the measurement of these cognitive skills on suitably chosen samples of children and for comparing schools and countries based on the distribution of these skills (see e.g. Hanushek and Woessmann (2011) for a survey of those). One of the most largely commented and discussed such data set is the Programme for International Student Assessment (PISA), which tests math, science, and reading performance of 15-year-olds children on a three-year cycle since 2000.

It is also widely acknowledged, notably by Mayer (1997), Black, Devereux, and Salvanes (2005), Schutz, Ursprung, and Woessmann (2008) or Dahl and Lochner (2012), that family background plays a determinant role in the child's cognitive skills acquisition process. As one sociologist put it to the scholar-politician Daniel Patrick Moynihan in reaction to the (quoted above) Coleman report on the educational opportunities offered by American schools in the sixties: "Have you heard what Coleman is finding? It's all family." The precise channels through which the family background affect the children' skill acquisition process is still subject to discussion. One channel may be genetic. Another channel may be the time and energy spent by the parents in helping the children to acquire those skills. But whatever the channel is, the family influence on the children' cognitive skill acquisition process must be accounted for when evaluating the performance of education systems. Two education systems who produce the same distribution of cognitive skills can *not* be considered as equally performing if the distribution of the children family backgrounds differ between the two. Moreover, there is a widely held view, often developed under the heading of "equality of opportunity" (see Schutz, Ursprung, and Woessmann (2008)), that good education systems are those that succeed somehow in breaking the dependency of the children skill acquisition process upon the family circumstances.

This paper proposes a *robust methodology* for evaluating education systems on the basis of a few explicit *elementary principles* that capture these general ideas. The principles apply to data on education systems in which every pupil is described by two numbers: one measuring his/her cognitive skill, and the other measuring his/her family background. Viewed in this way, the issue of comparing education systems amounts to comparing distributions of pairs of numbers, just as in the traditional *multi* - actually *two* - dimensional normative evaluation developed along the lines of Atkinson and Bourguignon (1982) (see e.g. Atkinson and Bourguignon (1987), Bourguignon (1989), Jenkins and Lambert (1993), Gravel and Moyes (2012), Moyes (2012) for theoretical contributions and Duclos, Sahn, and Younger (2006), Gravel, Moyes, and Tarroux (2009), Gravel and Mukhopadhyay (2010) and Hussain, Jorgensen, and Osterdal (2016) for empirical applications). However, the particular nature of the two numbers that describe education systems suggests principles for comparing them that differ from those considered in the two-dimensional normative evaluation literature.

The *first* principle that we consider is the favorable recording of any improvement in cognitive skill *ceteris paribus*. Most popular discussions about the relative performance of different national education systems, notably around the releases of PISA studies, clearly agree with this principle. The *second* principle requires that skills be appraised more favorably when observed in children with low backgrounds. Such a principle clearly underlies many national public schools systems that provide additional funding to schools located in underprivileged neighborhoods. A good example of such scheme is the French *Zone d'Education Prioritaire* program analyzed by Bénabou, Kramarz, and Prost (2009). The *third* principle reflects a preference - alluded to above - for education systems who succeed in reducing the family's influence on the child cognitive skill. Consider indeed an education system in which one child from a favorable background achieves high cognitive skills while another child from a less favorable background achieves a lower skill level. Consider another education system that differs from the previous one only by the fact that the high-skill child is now coming from the low family background while the low-skill child is coming from the high background. Aversion to correlation would suggest that the second system performs better than the first. At least many empirical studies - such as Schutz, Ursprung, and Woessmann (2008) - who regress the skill variable over a set of explanatory variables - including of course some that measure family background - and who compare school systems based on the value of the regression coefficient of the family variable would agree with this principle.

This paper proposes a dominance criterion for evaluating education systems that is shown to be the unanimity of all transitive rankings who agree with these three principles. The criterion says that one education system dominates another if, for any pair of reference levels of background and skill, the fraction of children with *both* a *lower* background *and* a *better* skill than that reference is larger in the dominating than in the dominated system. This criterion shares with one of the first order criteria of Atkinson and Bourguignon (1982) - when applied to education systems - the agreement with the first and the third principles. However, it differs from the Atkinson and Bourguignon (1982) criterion in considering the second attribute - family background - as having a *negative* impact on the performance of a education system. Atkinson himself (see especially Atkinson (1981a) and Atkinson (1981b)) has applied one of the first order dominance criteria of Atkinson and Bourguignon (1982) to the issue of measuring intergenerational income mobility (see e.g. Fields and Oke (1999) for a survey on income mobility measurement). By so doing, he endorsed the view that improving the distribution of parental status *ceteris paribus* improves intergenerational mobility. While this view may be defensible for evaluating intergenerational income mobility - at least if one is adopting for that purpose the perspective of Shorrocks (1978) - it is less so when appraisal of education system performance is at stake.

The three principles just sketched, and the dominance criterion that they characterize, form the core of the analysis of this paper. However, the fact that the criterion stands on very consensual principles makes it fairly incomplete. Additional principles need to be called for if one wants to increase the number of conclusive comparisons of education systems. One such principle concerns attitude toward inequalities in cognitive skills, given the background. Making such an attitude precise requires, when developed in the conventional framework of inequality measurement, that cardinal significance be attached to the measurement of cognitive skills. Provided that this is the case, are inequalities in cognitive skills - *ceteris paribus* of course - a good or a bad thing ? While spontaneous intuition - such as that underlying the empirical analysis of Goussé and LeDonné (2015) - seems to favour the second rather than the first answer to this question, a second thought may make one more hesitant. This is at least so if one recognizes, in line with much of the empirical literature (see e.g. Green and Riddell (2003), Heckman, J. Stixrud, and Uzrua (2006) or Barrett (2012)), that income is a *convex* function of cognitive skills. If this convexity is strong enough, it is possible that the function that converts cognitive skill into well-being (or some other normatively relevant measure of "advantage") be itself convex. If the

individual well-being (or advantage) is a convex function of cognitive skills, then a utilitarian ethical observer could favour increasing inequalities - as defined by mean-preserving spread - in cognitive skills, everything else being the same. Following Bazen and Moyes (2012), we call *elitism* such a favorable appreciation of mean-preserving spreads in cognitive skills performed between children with the same background. As it happens, when we add elitism to the three other principles, we characterize an additional dominance criteria that is compatible with the previous one while being more discriminatory. Unfortunately, and for reasons that we believe to be deep (and beyond the scope of this paper), we are not capable of characterizing a dominance criterion that respects the three first principles but that replaces elitism by the converse *egalitarian* view which dislikes mean preserving spreads in cognitive skills.

We then put our dominance criteria to work by comparing the national education systems of 43 countries based on the 2015 wave of the PISA survey. We specifically compare, across those countries, the joint distributions of the children's scores in mathematics - as measured by PISA tests - and their parents social status (defined as the highest *International Socioeconomic Index* of the two parents). The most discriminatory criterion - which adds elitism to the three first principles discussed above - is capable of conclusively ranking nearly 19% of all the possible pairs of countries. The percentage of clear-cut comparisons obtained from the three core principles alone - without any elitist principle - is in itself 17.5%. Hence, it happens that the added elitist value judgement does not contribute much to the ability of the dominance criterion to rank conclusively education systems. While the fractions of conclusive comparisons may be considered small, the robustness of the obtained comparisons is worth emphasizing. Among the noteworthy robust comparisons, one finds that Vietnam has one of the most performing education system in the world. In effect, the Vietnamese education system dominates 17 out of the 42 other countries, and is dominated by none. To some extent, this reflects the fact that Vietnamese children do very well in their PISA test even though they come from parents with relatively low status. Among the developed countries, Japan and Poland appear to stand the best against the others. Their education systems dominate that of respectively 9 and 12 other countries, while being dominated by none of them. Finland, usually described as a top performer insofar as education goes, does not perform outstandingly according to our criteria. While it is never dominated, it only dominates five other countries. At the bottom of our rankings, one finds countries such as Jordan, Lebanon and the United Arab Emirates (UAE). The education system of any of these three countries dominates that of no other. However the education system of the UAE is dominated by that of 29 countries. The performance of Lebanon (dominated by 25 countries) and Jordan (dominated by 18 countries) is also low when appraised by our criteria.

The rest of the paper is organized as follows. The next section presents the criteria and principles used to compare education systems, and establishes the equivalence among them. Section 3 discusses the data and the empirical methodology. Section 4 shows and discusses the empirical results and section 5 concludes.

## 2  Criteria for comparing education systems

### 2.1  Framework and notation

We are interested in comparing alternative education systems by means of some *anonymous* criterion. Every such system educates a set of $n$ children (with $n \geqq 3$).[1] Every child $i \in \{1, ..., n\}$

---

[1] We assume that education systems all educate the same number of children for pedagogical convenience. This assumption can be dispensed with if one adheres to the *Dalton principle* according to which replicating finitely many times a given population of children with a given distribution of skills and backgrounds is a matter of indifference.

4

has a family **b**ackground $b_i$, taken from some finite set $\mathcal{B} = \{1,...,\overline{b}\}$ of positive integers. At the end of the education process, every child $i$ acquires a cognitive **s**kill $s_i$ that is also taken from some finite set $\mathcal{S} = \{1,...,\overline{s}\}$ where $\overline{s}$ is some positive integer. Hence an **e**ducation system **e** is an ordered list $\{b_i^{\mathbf{e}}, s_i^{\mathbf{e}}\}_{i=1}^n$ of such pairs of numbers. This discrete setting does not entail any significant loss of generality since one can always take the numbers $\overline{b}$ and $\overline{s}$ to be very large, and interpret the units as being suitably small fraction of some scale. The anonymity of the criterion is the requirement that no importance be given to the children names or other irrelevant characteristics (beside skills and backgrounds). We make this requirement precise by summarizing the relevant information associated to an education system **e** by its $\overline{b} \times \overline{s}$ integer density matrix $\mathbf{d}^e$ :

$$\mathbf{d^e} := \begin{bmatrix} \mathbf{d_{11}^e} & ... & \mathbf{d_{1\overline{s}}^e} \\ \vdots & \vdots & \vdots \\ \mathbf{d_{\overline{b}1}^e} & ... & \mathbf{d_{\overline{b}\overline{s}}^e} \end{bmatrix}, \tag{1}$$

where, for $b \in \mathcal{B}$ and $s \in \mathcal{S}$, the number $\mathbf{d}_{bs}^{\mathbf{e}} \in \{0,...,n\}$ denotes the (possibly null) number of children with background $b$ and skill $s$ observed in education system **e**. We of course require that $\sum_{(b,s)\in\mathcal{B}\times\mathcal{S}} \mathbf{d}_{bs}^{\mathbf{e}} = n$. We let $\mathcal{E} = \{\mathbf{d} \in \{0,...,n\}^{\overline{bs}} : \sum_{(b,s)\in\mathcal{B}\times\mathcal{S}} \mathbf{d}_{bs} = n\}$ denote the (finite) set of all such possible education systems.

Education systems are compared by means of a reflexive and transitive binary relation $\succsim$ on $\mathcal{E}$. For any two education systems $\mathbf{d}^{\mathbf{e}^*}$ and $\mathbf{d}^{\mathbf{e}^\circ}$ in $\mathcal{E}$, we interpret the statement $\mathbf{d}^{\mathbf{e}^*} \succsim \mathbf{d}^{\mathbf{e}^\circ}$ as meaning that education system $\mathbf{d}^{\mathbf{e}^*}$ performs at least as well as $\mathbf{d}^{\mathbf{e}^\circ}$. An analogous interpretation is given to $\mathbf{d}^{\mathbf{e}^*} \succ \mathbf{d}^{\mathbf{e}^\circ}$ (strictly better than) and $\mathbf{d}^{\mathbf{e}^*} \sim \mathbf{d}^{\mathbf{e}^\circ}$ (equally well as). The requirement of reflexivity and transitivity of the ranking strikes us as quite natural. Notice that we do not require $\succsim$ to be complete. As discussed earlier, the dominance criteria used in the paper will very often fail in providing clear-cut comparison of education systems.

For any density matrix $\mathbf{d^e}$ associated to an education system **e** and any target $b$ of background and $s$ of skill, we denote by $S^{\mathbf{e}}(b,s)$ the **s**uccess relative to $s$ of children with background $b$ in the education system **e** defined by:

$$S^{\mathbf{e}}(b,s) = \sum_{\sigma=s}^{\overline{s}} \mathbf{d}_{b\sigma}^{\mathbf{e}}(\sigma - s) \tag{2}$$

This success is the sum, taken over all pupils of background $b$ who have a better skill than $s$, of their *skill excess* over $s$. This expression plays a key role in the definition of the dominance criterion based on elitism discussed below.

## 2.2   The dominance approach

We now formulate elementary principles that could plausibly underlie the comparisons of education systems. Three main such principles drive our attention herein. The first of them is the favorable appraisal of an improvement in the cognitive skills of a child, everything else the same. Specifically, consider the following definition of an *improvement in a child's skill.*

**Definition 1** Improvement in a child's skill. *We say that education system $\mathbf{d}^{\mathbf{e}^*}$ is obtained from education system $\mathbf{d}^{\mathbf{e}^\circ}$ by means of an improvement in a child's skill if there exists some $b \in \mathcal{B}$*

*and $s \in \{1, ..., \overline{s} - 1\}$ such that:*

$$\mathbf{d}^{\mathbf{e}^*}_{\beta\sigma} = \mathbf{d}^{\mathbf{e}^\circ}_{\beta\sigma} \textit{ for all } (\beta, \sigma) \in (\mathcal{B} \times \mathcal{S}) \backslash \{(b,s), (b, s+1)\}$$
$$\mathbf{d}^{\mathbf{e}^*}_{bs} = \mathbf{d}^{\mathbf{e}^\circ}_{bs} - 1 \textit{ and,}$$
$$\mathbf{d}^{\mathbf{e}^*}_{bs+1} = \mathbf{d}^{\mathbf{e}^\circ}_{bs} + 1.$$

In words, an improvement in a child's skill describes the process by which a child with background $b$ sees his/her cognitive skill improving by one unit, everything else remaining the same. Any such improvement would naturally be considered favorably by an evaluation of an education system. This formal definition is illustrated below in the (stylized) case where there are four children and two categories of skill and background who are, in the initial distribution $\mathbf{d}^{\mathbf{e}^\circ}$, uniformly distributed in the four cells of the matrix.

$\mathbf{d}^{\mathbf{e}^*} =$

|  | low skill | high skill |
|---|---|---|
| low background | 0 | 2 |
| high background | 1 | 1 |

$\mathbf{d}^{\mathbf{e}^\circ} =$

|  | low skill | high skill |
|---|---|---|
| low background | 1 | 1 |
| high background | 1 | 1 |

The second principle considers the dual situation of two education systems that differ only by the fact that two children with the *same skill* are coming from *different backgrounds*. In which of the two education systems is this cognitive skill achievement the most remarkable? It would seem that it is when the child is coming from a low, rather than from a high, background. We define in this spirit as follows the notion of a *deterioration in a child's family background*.

**Definition 2** Deterioration in a child's background. *We say that education system $\mathbf{d}^{\mathbf{e}^*}$ is obtained from education system $\mathbf{d}^{\mathbf{e}^\circ}$ by means of a deterioration in a child's background if there is a background $b \in \{2, ..., \overline{b}\}$ and a skill $s \in \mathcal{S}$ such that:*

$$\mathbf{d}^{\mathbf{e}^*}_{\beta\sigma} = \mathbf{d}^{\mathbf{e}^\circ}_{\beta\sigma} \textit{ for all } (\beta, \sigma) \in (\mathcal{B} \times \mathcal{S}) \backslash \{(b,s), (b-1, s)\}$$
$$\mathbf{d}^{\mathbf{e}^*}_{bs} = \mathbf{d}^{\mathbf{e}^\circ}_{bs} - 1 \textit{ and,}$$
$$\mathbf{d}^{\mathbf{e}^*}_{(b-1)s} = \mathbf{d}^{\mathbf{e}^\circ}_{(b-1)s} + 1.$$

An example of an education system $\mathbf{d}^{\mathbf{e}^*}$ obtained from $\mathbf{d}^{\mathbf{e}^\circ}$ by means of a deterioration in a child's background is depicted below.

$\mathbf{d}^{\mathbf{e}^*} =$

|  | low skill | high skill |
|---|---|---|
| low background | 1 | 2 |
| high background | 1 | 0 |

$\mathbf{d}^{\mathbf{e}^\circ} =$

|  | low skill | high skill |
|---|---|---|
| low background | 1 | 1 |
| high background | 1 | 1 |

The third considered principle concerns the extent to which an education system *reduces the correlation* between the child's cognitive skill and the child's family background. To use the terminology of Daniel Patrick Moynihan mentioned earlier, a good education system is one in which "it is not all family". And a better education system than another is one in which it "less family" than in the other. We formulate this reduction in correlation between skill and background in the following fashion.

6

**Definition 3** Reduction in correlation between skill and background. *We say that education system* $\mathbf{d^{e^*}}$ *is obtained from education system* $\mathbf{d^{e^\circ}}$ *by means of a reduction in correlation between skill and background if there are* $b$ *and* $b'$ *in* $\mathcal{B}$ *and* $s$ *and* $s'$ *in* $\mathcal{S}$ *satisfying* $b < b'$ *and* $s < s'$ *such that:*

$$
\begin{aligned}
\mathbf{d}^{\mathbf{e}^*}_{\beta\sigma} &= \mathbf{d}^{\mathbf{e}^\circ}_{\beta\sigma} \ \textit{for all } (\beta, \sigma) \in (\mathcal{B} \times \mathcal{S}) \backslash \{(b,s), (b',s), (b,s'), (b',s')\} \\
\mathbf{d}^{\mathbf{e}^*}_{bs} &= \mathbf{d}^{\mathbf{e}^\circ}_{bs} - 1 \ \textit{and } \mathbf{d}^{\mathbf{e}^*}_{b's'} = \mathbf{d}^{\mathbf{e}^\circ}_{b's'} - 1 \ , \\
\mathbf{d}^{\mathbf{e}^*}_{bs'} &= \mathbf{d}^{\mathbf{e}^\circ}_{bs'} + 1 \ \textit{and } \mathbf{d}^{\mathbf{e}^*}_{b's} = \mathbf{d}^{\mathbf{e}^\circ}_{b's} + 1
\end{aligned}
$$

Hence, a reduction in correlation between skills and background is just a switch of cognitive skills between two children with differing skills *and* backgrounds and who are ordered with respect to both skill and background. In order to be correlation reducing, the high background-high skill child must switch his/her skill with that of the low-skill-low-background child. An illustration of such a reduction in correlation between skills and background is, again, provided below in the four-children-two levels of skill and background case.

$\mathbf{d^{e^*}} =$

|  | low skill | high skill |
|---|---|---|
| low background | 0 | 2 |
| high background | 2 | 0 |

$\mathbf{d^{e^\circ}} =$

|  | low skill | high skill |
|---|---|---|
| low background | 1 | 1 |
| high background | 1 | 1 |

In the empirical literature, education systems are often compared on the basis of the value assigned to them by some *Index I*. With such an approach, the statement $I(\mathbf{e}) \geq I(\mathbf{e}')$ means that education system $\mathbf{e}$ performs better than education system $\mathbf{e}'$. A widely used such index is the linear regression coefficient of the cognitive skills on family background, as used for example in Schutz, Ursprung, and Woessmann (2008). This index is based on the estimation, for every education system $\mathbf{e}$, of the following regression model (abstracting from the additional "control" variables often considered in the regressions):

$$
s_i^{\mathbf{e}} = \alpha^{\mathbf{e}} + \beta^{\mathbf{e}} b_i^{\mathbf{e}} + \varepsilon_i^{\mathbf{e}}
$$

where $\varepsilon_i^{\mathbf{e}}$ is the regression error term observed on child $i$ of system $\mathbf{e}$ and $\alpha^{\mathbf{e}}$ and $\beta^{\mathbf{e}}$ are the (theoretical) constant and regression coefficient (respectively) of this linear model applied to system $\mathbf{e}$. It is well-known from elementary econometrics that the least-square estimate of $\beta^{\mathbf{e}}$, denoted $\widehat{\beta}(\mathbf{e})$, is defined by:

$$
\begin{aligned}
\widehat{\beta}(\mathbf{e}) &= \frac{cov(s_i^{\mathbf{e}}, b_i^{\mathbf{e}})}{var(b_i)} \\
&= \frac{\displaystyle\sum_{i=1}^{n} s_i^{\mathbf{e}} b_i^{\mathbf{e}} - n\overline{s}(\mathbf{e})\overline{b}(\mathbf{e})}{\displaystyle\sum_{i=1}^{n} (b_i^{\mathbf{e}})^2 - n\overline{s}\overline{b}(\mathbf{e})}
\end{aligned}
\tag{3}
$$

where $\overline{s}(\mathbf{e})$ and $\overline{b}(\mathbf{e})$ denote the average skill and family background (respectively) observed in the education system $\mathbf{e}$. The index $\widehat{\beta}(\mathbf{e})$ is usually used as a (negative) measure of the performance of an education system (the lower the index, the better the system). Beside the convenience of

obtaining the index out of a simple estimation of a linear regression, there is not much justification for the ranking of education systems that it performs. For one thing, the ranking of education systems induced by $\widehat{\beta}$ does not always record favorably a child's improvement in cognitive skill. It will actually record it unfavorably if the improvement concerns a child whose family background is above the mean. The ranking induced by $\widehat{\beta}$ does not either record favorably a child's deterioration of family background. However a reduction in the correlation between skill and background will clearly reduce the value of $\widehat{\beta}$.

In what follows, we propose to compare education systems by the (unique) criterion that is agreed upon by *all* reflexive and transitive rankings of education systems in $\mathcal{E}$ who consider favorably child's improvements in cognitive skill, child's deteriorations in the family background, and reductions in correlation between skill and background as just defined. As it turns out, this criterion is extremely easy to apply. It amounts to verifying, for any reference pair of skill and background, if the number (fraction) of children with *both* a *worse* background *and* a *better* skill than the reference is larger in one education system than another.

The formal statement of this result is the object of Theorem 1 below. The proof of the result is somewhat technical. It has been for this reason relegated in the Appendix. Some of it uses abstract results from duality theory and follows a proof strategy adopted by Muller and Scarsini (2012), but adapted to the discrete nature of our setting using the approach of Magdalou (2018).

**Theorem 1** *Let* $\mathbf{d^{e^*}}$ *and* $\mathbf{d^{e^\circ}}$ *be two education systems in* $\mathcal{E}$. *Then, the following two statements are equivalent.*

(a) $\mathbf{d^{e^*}} \succsim \mathbf{d^{e^\circ}}$ *for any reflexive and transitive ranking* $\succsim$ *that records favorably improvements in a child's cognitive skill, deteriorations in a child's family background and reductions in the correlation between skill and background as per Definitions 1-3.*

(b) $\displaystyle\sum_{\beta=1}^{b} \sum_{\sigma=s}^{\overline{s}} \mathbf{d}^{e^*}_{\beta\sigma} \geq \sum_{\beta=1}^{b} \sum_{\sigma=s}^{\overline{s}} \mathbf{d}^{e^\circ}_{\beta\sigma}$ *for all* $(b, s) \in \mathcal{B} \times \mathcal{S}$.

## 2.3 Attitudes toward inequalities in cognitive skills

Theorem 1 provides a simple test for checking whether an education system is better than another for any transitive ranking that records favorably improvements in cognitive skills, deteriorations of family backgrounds and reductions in correlation between background and skills. The test consists in verifying if, for any pair of skill and background levels, the fraction of children with *both* a better skill *and* a lower family background than those levels is larger in one system than another. However the large consensus over which the test stands is likely to make the ranking highly incomplete. As usual in dominance analysis, any gain in discriminatory power comes at the cost of requiring the criterion to satisfy additional principles. What could these be ?

One concerns attitude toward inequalities in cognitive skills between pupils with the same background. Consider indeed two alternative hypothetical situations involving two differently skilled children with the same family background. Assume that the average cognitive skill - calculated over the two children - is the same in the two situations but that the two skills are more spread out in one situation than in the other. By "more spread out", we mean, as usual in inequality analysis, "resulting from a mean-preserving spread" or, equivalently, a mean-preserving "regressive transfer". Of course the very notion of a mean-preserving spread of skill rests on the belief that the skill variable is measured on a cardinal scale so that the very notion of a "preserved mean" makes sense.[2] Suppose we have this belief. Is a more spread-out distribution

---

[2] See Allison and Foster (2004) for a discussion of the difficulty of applying conventional concepts of inequality

8

of skill better or worse than a less spread-out one ? Intuition coming from conventional attitude toward income inequality could suggest that a less spread out - or a more equal - distribution of skills is better than a more spread-out one. Yet, such an intuition may be misleading for an attribute such a skill. This is at least so if one adopts the welfarist perspective according to which the "social goodness" of alternative states of affairs depends only upon the distribution of individual well-being that they generate. As documented in the literature, there seems to be two channels by which cognitive skills affects well-being. One of them is through the income that the cognitive skills enable the individual to earn. If one restricts attention to cognitive skill as the unique determinant of income, one can denote by $y(s)$ the income that an individual with skill $s$ can earn on the labour market. The empirical evidence (see e.g. Green and Riddell (2003), Heckman, J. Stixrud, and Uzrua (2006) or Barrett (2012)) on the functional relation $y$ connecting skill to income is that it is *increasing* and *convex*. That is, the *gain* in earning capacity brought about by an increase in skill is it self *increasing* with skill. The other channel through which cognitive skills affect well-being is a direct one. Cognitive skills help individuals to make better decisions (in the choice of his/her partner, career profile, medical treatments, etc.) and better use of the information irrespective of their impact on income. This suggests that the individual well-being, $u$ say, is an *intrinsic* function of two variables: skills ($s$) and income ($y$):

$$u = U(s, y)$$

Yet, since the income $y$ is itself a function of the skills, one can view the individual well-being $u$ as a function $\Psi$ of the skills only with $\Psi$ defined by:

$$u = \Psi(s) = U(s, y(s))$$

As commonly assumed in economics, the function $U$ that associates well-being to every combination of skill and income would be increasing in both variables and concave with respect to income (marginal utility of income is decreasing with respect to income at any skill level). Yet, economic theory, empirical evidence, and introspection do not provide clear evidence about the concavity or convexity of $U$ with respect to skill. Convexity - e.g. the fact that, given income, the marginal utility of an increase in skill is increasing with skill - is not implausible. A similar lack of *a priori* intuition concerns the relation between the (positive) marginal utility of income and the skill level. Are skill and income complement, or substitute, for the achievement of a given level of well-being ? It is not implausible to believe that they are complement, so that the marginal utility of income is increasing (at least weakly) with respect to skill. To sum up, $U$ could plausibly satisfy (assuming differentiability):

$$
\begin{align}
U_j(.) &\geq 0 \text{ for } j = s, y \tag{4} \\
U_{yy}(.) &\leq 0 \tag{5} \\
U_{ss}(.) &\geq 0 \tag{6} \\
U_{ys}(.) &\geq 0 \tag{7}
\end{align}
$$

where, for every $i, j \in \{s, y\}$, $U_j(.)$ and $U_{ij}(.)$ denote, respectively, the partial derivative of $U$ with respect to $j$ and the second derivative first with respect to $i$ and second with respect to $j$. Under these assumptions, and assuming again differentiability, one can see that

$$
\begin{align}
\Psi_s(.) &= U_s(.) + U_y(.)y_s(.) \\
&\geq 0
\end{align}
$$

---

measurement to distribution of an ordinal variable and Gravel, Magdalou, and Moyes (2015) for a dominance approach to the issue.

if the function $y$ is increasing in skills. More importantly for our purpose, one can also see that, under assumptions (4)-(7) and the convexity of the function $y$:

$$\Psi_{ss}(.) \quad = \quad \underbrace{U_{ss}(.)}_{+} + \underbrace{2U_{sy}(.)y_s(.)}_{+} + \underbrace{U_{yy}(.)[y_s(.)]^2}_{-} + \underbrace{U_y(.)y_{ss}(.)}_{+}$$

Hence, while the sign of $\Psi_{ss}(.)$ is *a priori* ambiguous, it is not implausible that the sum of the three positive terms outweigh the negative one.

We accordingly consider the possibility for skill to have a positive and *convex* final effect on well-being. If we do so, it becomes possible to defend the view that, for a population of children with the same background, a more unequal - or spread out - distribution of skills of a given mean is better than a less-spread out one. This approach, which can be viewed as the opposite to egalitarianism, has been called elitism by Bazen and Moyes (2012). The main elementary operation that describes the notion of dispersion in skills that would be considered favorably by elitism is the following.

**Definition 4** Dispersion in children skills. *We say that education system $\mathbf{d^{e^*}}$ is obtained from education system $\mathbf{d^{e^\circ}}$ by means of a dispersion in children skills if there is a background level $b \in \mathcal{B}$ and two skill levels $s$ and $s'$ in $\mathcal{S}$ satisfying $s' > s + 1$ such that:*

$$\mathbf{d^{e^*}_{\beta\sigma}} \quad = \quad \mathbf{d^{e^\circ}_{\beta\sigma}} \textit{for all } (\beta,\sigma) \in (\mathcal{B} \times \mathcal{S}) \backslash (\{(b,s'),(b,s'+1)\} \cup \{(b,s-1),(b,s)\})$$
$$\mathbf{d^{e^*}_{bs'}} \quad = \quad \mathbf{d^{e^\circ}_{bs'}} + 1, \, \mathbf{d^{e^*}_{bs'-1}} = \mathbf{d^{e^\circ}_{bs'-1}} - 1 \, , \, \mathbf{d^{e^*}_{bs}} = \mathbf{d^{e^\circ}_{bs}} + 1 \textit{ and } \mathbf{d^{e^*}_{bs+1}} = \mathbf{d^{e^\circ}_{bs+1}} - 1$$

*if $s' - 1 > s + 1$ and:*

$$\mathbf{d^{e^*}_{bs'}} = \mathbf{d^{e^\circ}_{bs'}} + 1, \, \mathbf{d^{e^*}_{bs'-1}} = \mathbf{d^{e^*}_{bs+1}} = \mathbf{d^{e^\circ}_{bs'-1}} - 2 \, , \, \mathbf{d^{e^*}_{bs}} = \mathbf{d^{e^\circ}_{bs}} + 1$$

*if $s' - 1 = s + 1$.*

As defined here, a dispersion in children skills is an operation by which a child of background $b$ and (relatively low) skill level $s + 1$ "falls down" by one category - to skill $s$ - in exchange of having, for the same background $b$, a child with a higher skill $s' - 1$ who "climb up" by one category. This operation clearly disperses skills of children of the concerned category without affecting the average skill of that category (since the falling down by one category of one child is exactly compensated by the climbing up by one category of the other). An illustration of the formal definition of a Dispersion in children skills in a case where there is only one background and three different skill levels is provided below.

$\mathbf{d^{e^*}} =$

|  | low skill (1) | mid skill (2) | high skill (3) |
|---|---|---|---|
| background $b$ | 2 | 0 | 2 |

$\mathbf{d^{e^\circ}} =$

|  | low skill (1) | mid skill (2) | high skill (3) |
|---|---|---|---|
| background $b$ | 1 | 2 | 1 |

We now identify, in the following theorem, an easily implementable dominance criterion that coincides with the unanimity of all reflexive and transitive rankings of education systems in $\mathcal{E}$ that record favorably improvements in cognitive skills, deteriorations of family background, reductions in correlation between skill and background *and* dispersions in children skills. Specifically, we prove in the Appendix the following theorem.

**Theorem 2** *Let $\mathbf{d}^{\mathbf{e}^*}$ and $\mathbf{d}^{\mathbf{e}^\circ}$ be two education systems in $\mathcal{E}$. Then, the following two statements are equivalent.*

(a) $\mathbf{d}^{\mathbf{e}^*} \succsim \mathbf{d}^{\mathbf{e}^\circ}$ *for any reflexive and transitive ranking $\succsim$ that records favorably improvements in a child's skill, deteriorations in a child's background, reductions in the correlation between skill and background and dispersions in children cognitive skills as per Definitions 1-.4*

(b) $\displaystyle\sum_{b\in\beta(\mathbf{e})\cup\boldsymbol{\beta}(\mathbf{e}^\circ)} S^{\mathbf{e}}(b,s(b)) \geq \sum_{b\in\beta(\mathbf{e})\cup\boldsymbol{\beta}(\mathbf{e}^\circ)} S^{\mathbf{e}^\circ}(b,s(b))$ *for all increasing functions $s:\mathcal{B}\to\mathcal{S}$ and*

$\displaystyle\sum_{\beta=1}^{b}\sum_{\sigma=1}^{\overline{s}} \mathbf{d}^{\mathbf{e}^*}_{\beta\sigma} \geq \sum_{\beta=1}^{b}\sum_{\sigma=1}^{\overline{s}} \mathbf{d}^{\mathbf{e}^\circ}_{\beta\sigma}$ *for every background b.*

Statement (b) of Theorem 2 thus provides an exact test for whether or not an education system stands above another by all rankings that record favorably improvements in a child's skill, deteriorations in a child's background, reductions in the correlation between skill and background and dispersions in children skills. The main part of the test works as follows. One first assigns to every background level a specific target of cognitive skill achievement that is *non-decreasing* with respect to the background. Thus, children from high backgrounds are assigned higher targets than those from lower backgrounds. One then sums, over all backgrounds, the success of the children - as defined by Expression (2) - relative to the target assigned to their background. An education system that exhibits a larger sum of such success than another for every assignment of targets that are (weakly) increasing with respect to backgrounds is then said to dominate the other. In addition of this requirement, Statement (b) demands that the dominating system has a weakly larger fraction of children coming from a lower background than any specified $b$ than the dominated one. Put otherwise, Statement (b) requires the "good" education system to have a marginal distribution of backgrounds that is stochastically dominated at the first order by that of the "bad" system. This requirement is of course in tune with the idea that the performance of an education system be considered all the more impressive as the family backgrounds of the children is unfavorable. In statement (b) of Theorem 1, this negative stochastic dominance condition on the marginal distribution of backgrounds was not explicitly mentioned because it was implied by the dominance test associated with this statement. However the ordered success excess criterion expressed in the first part of Statement (b) of Theorem 2 does not imply the requirement for the marginal distribution of backgrounds in the dominating system to be stochastically dominated at the first order by that of the dominating system. One must therefore add this requirement.

One can certainly hesitate in adhering to the elitist value judgement that dispersions in cognitive skills of children with a given family background are, ceteris paribus, a good thing. For reasons that are not completely clear to us (and actually to other users of multidimensional dominance analysis such as Muller and Trannoy (2012) (p. 138-139) who have discussed difficulties of the same nature), there does not seem to be an easily identifiable empirical test that corresponds to the unanimity of all criteria that record favorably improvements in a child's cognitive skill, deteriorations in a child's family background, reductions in the correlation between skill and background but who *dislike* dispersions in children cognitive skills. At least, we have not been capable of identifying such an empirical test.

In the next section, we illustrate empirically the usefulness of our criteria for ranking national education systems..

# 3 Empirical Analysis

## 3.1 Data and variables

We base our empirical analysis on the 2015 wave of the OECD Program for International School Assessment (PISA) survey. This survey assesses the skills of some 540 000 students - aged 15 - across 72 countries, 35 of which belonging to the OECD. The empirical analysis actually focuses on the 43 "full" countries (excluding therefore cities like Hong Kong, Shanghai and Taipei) who had more than 5 millions inhabitants in 2016. Another criterion of selection of the 43 country was the availability of information on the family background of the children for a sufficiently large fraction of them. The PISA survey does indeed provide some detailed information on the children' parents and family environment, but the fraction of the sampled children for which the information is available happens to vary somehow across countries. The sample of children selected by the PISA team in each country is based on a random selection of a sample of schools from which, in a second step, a random selection of the pupils is performed. In the last step, individuals are weighted in such a way as to make the sample representative of the actual population of the country. It is important to notice that despite these corrections, the samples of pupils evaluated in the PISA survey are not totally representative of the population of interest. Excluded from the samples are children who are not enrolled at school, or who are enrolled in very low grades for their age, or who do not go to school because of physical or intellectual deficiencies. While these limitations in the coverage of children are somewhat small in developed countries (where more than 90% of the children population is represented), they become more important for developing or newly developed countries that are not in the OECD (see e.g. Carvalho, Gamboa, and Waltenberg (2012)). These considerations have led us to remove relatively populated countries, such as Argentina and Dominican Republic, from the list.

The cognitive skills evaluated by the PISA survey concern 3 different subjects: Mathematics, Reading and Science. We focus herein on the Mathematics test, which seems to be less culturally biased. The results achieved by children on the test are standardized by the PISA team through a somewhat complex Item Response method described in PISA (2017). The pupils scores are specifically standardized as resulting from what was, in 1998, a Gaussian distribution centered at the OECD mean of 500 with a standard deviation of 100. Over the years, as new countries have been added in the data set, a concern for intertemporal comparability has led to the decision of maintaining the Gaussian distribution of 1998 as the standard. This has led over the year to a world distribution of the scores that has slightly moved away from the Gaussian. The 2015 world distribution of scores in mathematics used in this paper is shown in Figure 1. We emphasize that the standardization procedure underlying the definition of these individual scores should make one hesitant in attaching cardinal significance to the information conveyed by this variable (see e.g. Ferreira and Gignoux (2014) and Jacob and Rothstein (2016)). While this does not cause any problem for the criterion characterized in Theorem 1, it is of some importance for the more discriminatory elitist criterion of Theorem 2 who is sensitive to the cardinal measurement of the score. While the test score provided by PISA is a continuous variable, PISA also provides a discretization of the score into seven "proficiency categories" defined in Table 1.[3]

---

[3] See again the Report PISA (2017) for a detailed description of the procedure used for defining these categories.

| Level | interval of definition | mean value | % of the world sample |
|-------|------------------------|------------|------------------------|
| 6 | ]669.3, 1000] | 702.5 | 1.91 |
| 5 | ]606.99, 669.3] | 632.44 | 6.37 |
| 4 | ]544.68, 669.3] | 572.99 | 13.93 |
| 3 | ]482.38, 544.68] | 512.03 | 20.16 |
| 2 | ]420.07, 482.38] | 450.9 | 21.81 |
| 1 | ]357.77, 420.07] | 390.01 | 18.78 |
| 0 | ]0, 357.77] | 308.04 | 17.05 |

Table 1: Definition of proficiency levels
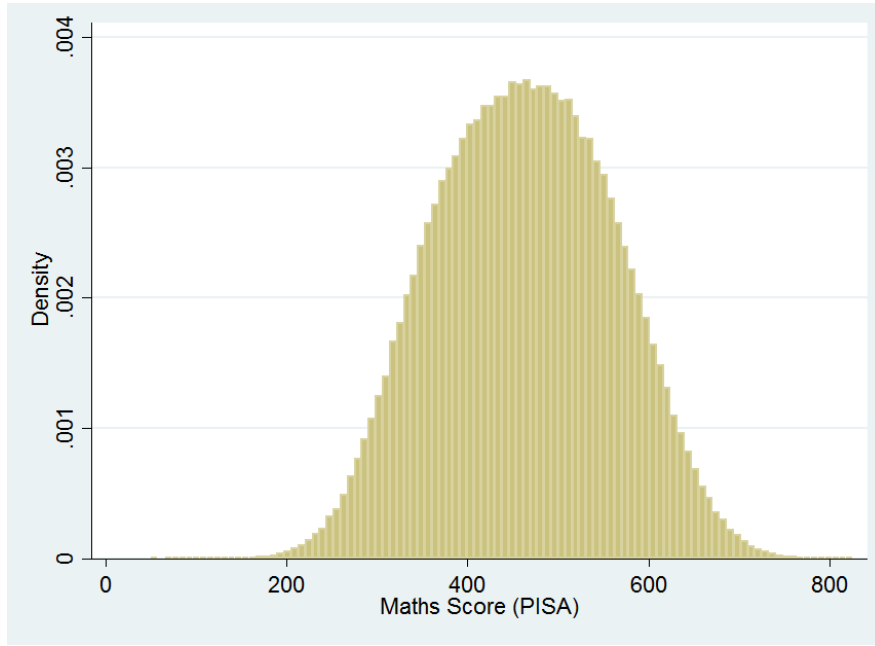


Figure 1: World distribution of scores in mathematics, 2015.

The analysis conducted below is based on this discretization of the PISA scores.

As for the family background, we take it to be **H**ighest of the two parents' **I**nternational **S**ocio**E**conomic **I**ndex (HISEI), as defined in Ganzeboom, Graaf, and Treiman (1992)). The HISEI can be described as a weighted average of the parents' education and profession, the latter being itself ranked by the average level of income associated to that profession within the country. The HISEI is a continuous variable ranging from 0 to 100. We justify the choice of the HISEI, as compared to, say, the largest of the two parents years of completed education (also provided in the PISA survey) by the fact that HISEI is sensitive to both the income level and the education of the children' parents. It is in effect well-documented that both parents education and income contribute to the children human capital. To that extent, we feel that HISEI is the most comprehensive summary indicator of the children family background that is available in the PISA data. Figure 2 below shows the 2015 world distribution of HISEI in the PISA data

base.

As can be seen, the HISEI is not at all distributed according to a Gaussian distribution at the world level.

Table 2 below provides the relevant descriptive statistics for each of the two variables at the country level. Since our criteria are sensitive to the correlation between skill and background, the table also provides the observed correlation between the two variables for each country, using two measures of correlation. The first one is the usual Pearson's correlation coefficient (covariance divided by the product of the standard deviation of each of the variable), and the other is the regression coefficient $\widehat{\beta}(\mathbf{e})$ discussed above. While correlation appears to be important, it does exhibit significant difference between countries. Countries such as Algeria, Japan or Russia exhibit a rather low correlation (less than 0.25) between children math score and family backgrounds. At the other extreme, the correlation gets above 0.42 in countries such as Belgium, France, the Czech Republic, Hungary and Peru. All in all, the world level correlation between cognitive skill in mathematics and family background, sightly above 0.33, is significant, but not outlandish.
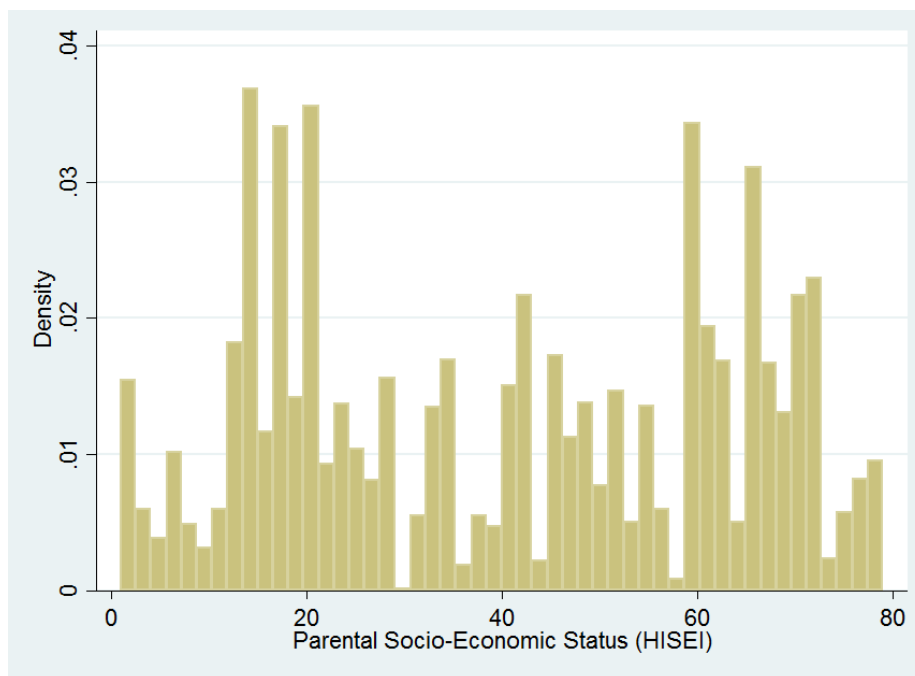


Figure 2: World distribution of the Highest of the parents' International Socio-Economic Status (HISEI), 2015.

Figure 3 shows the estimated linear relation between the two variables measured at the country level. The correlation appears significantly more important than what is observed within each country. This difference is not so surprising. The correlation indicated in Figure 3 captures in part the well-known aggregate relation between economic development - strongly correlated with average HISEI - and average school achievement. The within-country correlations described on Table 2 reflect more the effect of the children's background on their school achievement. It appears therefore that these individual effects are smaller than the aggregate one.
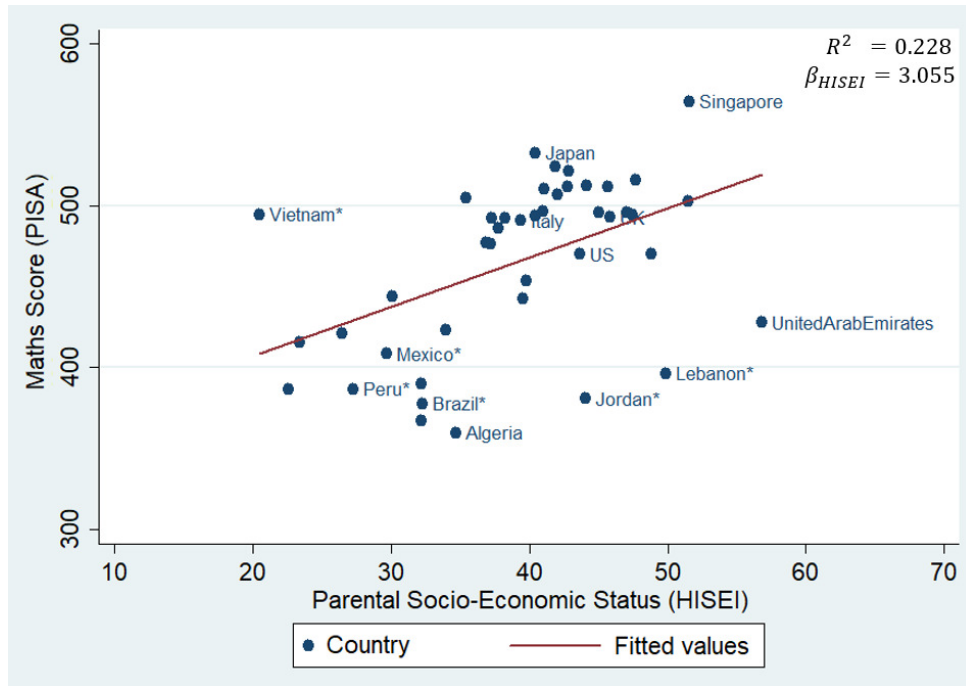
14

Figure 3: Regression of the country's average math score upon the country's average family background

Countries such as Vietnam, Japan or Singapore appear to be top performers in terms of their achieved math scores, as compared to their predicted math scores. Countries like Algeria, Jordan, Lebanon and the United Arab Emirates (UAE) appear on the other hand to be under-performers on that basis.

The analysis conducted below rides on a discretization of the HISEI based on the quintiles of world distribution of this variable depicted on Figure 2. The discretization generates the following 5 categories:

- Category 1 : $HISEI < 18$

- Category 2 : $18 \leq HISEI < 31$

- Category 3 : $31 \leq HISEI < 48$

- Category 4 : $48 \leq HISEI < 63$

- Category 5 : $HISEI \geq 63$

## 3.2   Results

We now provide cross-countries comparisons of the joint distributions of backgrounds (discretized HISEI) and cognitive skills (discretized score in mathematics) based on the criteria characterized in Theorems 1 and 2. Each of these criteria amounts to verifying if a finite set of inequalities connecting the joint distributions hold. Since the verification of these inequalities is done on samples of the children, we perform statistical inference to make our conclusion somewhat applicable to the whole population of interest. We use for this purpose the methodology of Davidson and Duclos (2000) that we adapt to the current setting by following the approach of Gravel, Moyes, and Tarroux (2009). We provide a brief description of this methodology in the Appendix. The results reported and discussed in this section are all significant at the 95% confidence interval, and the joint test of the non-rejection of the hypothesis that all weak inequalities that define dominance are of the required sign is based on so-called *Union-Intersection* criterion proposed in Bishop and Formby (1999).

We start by comparing countries on the basis of first order stochastic-dominance applied to the marginal distributions of the parent's discretized HISEI. The results are depicted on the Hasse diagram of Figure 4. On this diagram, countries are vertically ordered in terms of their average HISEI score. Vertical links are then drawn between them when (and only when) the countries are connected by first order stochastic dominance applied to their distributions of discretized HISEI scores. Countries which appear with a star (*) are those for which the "net secondary school enrolment rate" is either below 80% or not available. Vietnam for instance does not report its secondary enrollment rate. The PISA survey obviously provides an upward biased picture of the countries for which a significant fraction of the children are not enrolled in secondary schools (because it only focuses on the somewhat privileged segment of its children population who is enrolled at those schools). However, despite this possible upward bias, the countries marked with a star tend to be at the bottom of the ranking. The only exception to this is Vietnam, to be discussed further below.

Recall that observing (the inverse of) first order dominance of the marginal distribution of the family backgrounds between two countries is necessary for having the dominating countries to be better than the dominated one by the criterion characterized in Theorem 1. Indeed, it is well-known (see e.g. Lehmann (1955)) that (inverse) first order dominance between two marginal distributions of HISEI is equivalent to the possibility of going from the dominated to the dominating distribution by a finite sequence of deteriorations in a child' background. Hence, the somewhat high position of countries such as the United Arab Emirates (UAE) or Singapore on Figure 4 makes clear that these countries, who rank therefore at the bottom of the inverse first order ranking, can not dominate any other by the criteria characterized in Theorems 1 and 2. All in all, the ranking of countries based on first order stochastic dominance of their

| Country | Average Math score | average HISEI | Correlation | $\widehat{\beta}(e)$ |
|---|---|---|---|---|
| | | | | |
| Singapore | 564 | 51.5 | 0.34 | 1.632 |
| Japan | 533 | 40.5 | 0.24 | 1.019 |
| Korea | 524 | 41.9 | 0.29 | 1.415 |
| Switzerland | 522 | 42.9 | 0.35 | 1.472 |
| Canada | 516 | 47.7 | 0.30 | 1.212 |
| Netherlands | 512 | 44.1 | 0.32 | 1.419 |
| Denmark | 512 | 45.7 | 0.33 | 1.137 |
| Finland | 511 | 42.8 | 0.32 | 1.156 |
| Germany | 510 | 41.1 | 0.37 | 1.558 |
| Belgium | 507 | 42.1 | 0.43 | 1.828 |
| Poland | 505 | 35.5 | 0.29 | 1.090 |
| Norway | 503 | 51.5 | 0.28 | 1.130 |
| Austria | 497 | 41.0 | 0.35 | 1.510 |
| Russia | 496 | 45.0 | 0.21 | 0.802 |
| Australia | 496 | 47.1 | 0.31 | 1.300 |
| Sweden | 495 | 47.4 | 0.36 | 1.506 |
| Vietnam | 495 | 20.6 | 0.30 | 1.110 |
| France | 494 | 40.5 | 0.42 | 1.788 |
| UK | 493 | 45.8 | 0.30 | 1.274 |
| Czech Republic | 493 | 37.3 | 0.42 | 1.900 |
| Portugal | 492 | 38.2 | 0.40 | 1.593 |
| Italy | 491 | 39.4 | 0.33 | 1.382 |
| Spain | 486 | 37.8 | 0.33 | 1.149 |
| Hungary | 477 | 36.9 | 0.48 | 1.989 |
| Slovakia | 476 | 37.2 | 0.35 | 1.528 |
| Israel | 470 | 48.9 | 0.36 | 1.716 |
| US | 470 | 43.7 | 0.32 | 1.248 |
| Greece | 454 | 39.8 | 0.34 | 1.229 |
| Romania | 444 | 30.1 | 0.35 | 1.343 |
| Bulgaria | 442 | 39.5 | 0.40 | 1.667 |
| UAE | 428 | 56.8 | 0.18 | 1.120 |
| Chile | 423 | 34.0 | 0.38 | 1.422 |
| Turkey | 421 | 26.5 | 0.29 | 1.129 |
| Thailand | 416 | 23.4 | 0.27 | 1.044 |
| Mexico | 408 | 29.7 | 0.27 | 0.864 |
| Lebanon | 396 | 49.9 | 0.15 | 0.680 |
| Colombia | 390 | 32.2 | 0.32 | 1.064 |
| Peru | 386 | 27.3 | 0.44 | 1.463 |
| Indonesia | 386 | 22.6 | 0.34 | 1.227 |
| Jordan | 381 | 44.1 | 0.25 | 0.979 |
| Brazil | 378 | 32.3 | 0.35 | 1.313 |
| Tunisia | 367 | 32.2 | 0.33 | 1.137 |
| Algeria | 360 | 34.7 | 0.18 | 0.525 |

Table 2: descriptive statistics at the country level

distributions of HISEI appears to be in tune with the countries' per capita GDP. Exceptions to these are Lebanon, Jordan and Russia whose relatively good standing in terms of average and/or first order dominance of their HISEI distributions (they stand above western European countries such as Germany, Switzerland, France or the Netherlands) seems at odd with the modest ranking of these countries in terms of per capita GDP.

Figure 5 shows the first and second order dominance rankings of countries based on the marginal distributions of mathematics PISA scores (ignoring therefore family background). The second order criterion considered is the "elitist" one that can be expressed in terms of the generalized dual Lorenz criterion. According to this criterion, country $A$ ranks above country $B$ if, for every $k$, the sum of proficiency levels achieved by the $k$ most proficient children is higher in $A$ than in $B$. First order dominance implies second order dominance but the converse does not hold. On Figure 5, the countries are vertically ordered by their average PISA score in mathematics. We can observe that first order dominance applied to the marginal distributions of math scores - irrespective of the children family backgrounds - leads to a rather precise rankings of the countries. Indeed, out of the 903 possible pairs of distinct countries, 783 are ranked conclusively by the first-order criterion (a rate of comparability of almost 87%). Moving from first to second order dominance obviously increases to 825 the number of pairs of countries that can be conclusively compared (a comparability rate sightly above 91%). Hence, the elitist value judgement according to which increasing the spread of the distribution of PISA scores in math - given the mean of that score and the background of the children - does not increase immensely the discriminatory power of dominance when applied to the marginal distributions of math scores. Some countries see their standing significantly increased as a result of valuing favorably increasing dispersion in cognitive skill. The biggest winner in this respect is Korea, which dominates 37 countries (out of 42) on the basis of first-order dominance, but which second order-dominates all countries but Japan and Singapore.

The rankings provided in Figures 4 and 5 are useful preliminaries for understanding those produced by our criteria. Indeed the criteria defined in statements (b) of Theorems 1 and 2 require - as necessary conditions - that there be (inverse) first-order Dominance for the marginal distribution of family backgrounds. Moreover, the first order stochastic dominance of the marginal distribution of skills is also necessary for the criterion of statement (b) of Theorem 1 to hold. Hence, the fact of observing country $A$ dominating country $B$ in Figure 5 while observing the inverse dominance relationship in Figure 4 is a useful starting ground for observing the dominance of $B$ by $A$ by our main dominance criterion. However, the ranking of countries provided by the criterion of Statement (b) of Theorem 1 is *not* the mere intersection of the first order dominance of Figure 5 and the (inverse) first order dominance of Figure 4. In addition to these two first order dominance conditions on the marginals, dominance requires also reductions in the correlation between skill and background.

Figure 6 provides the Hasse dominance diagram of the 43 countries provided by statements (b) of Theorems 1 (criterion C1 of the diagram) and 2 (criterion C2). As can be seen, there are many pairs of countries - more than 82.5% - that can not be compared at all by the very demanding agreement among all transitive rankings that record favorably improvements in a child's cognitive skill, deteriorations in a child's family background and reductions in the correlation between skill and background. Moreover, restricting the unanimity to those elitist principles that also record favorably an increase in the dispersion of cognitive skill only increases the fraction of conclusively compared pairs of countries from 17.5% to 18.7%. As was the case for marginal distributions of skill levels, the gain of comparability obtained by adding elitism to the list of the three core principles is rather modest.

Nonetheless, it happens that every country can be compared with at least one other. However,
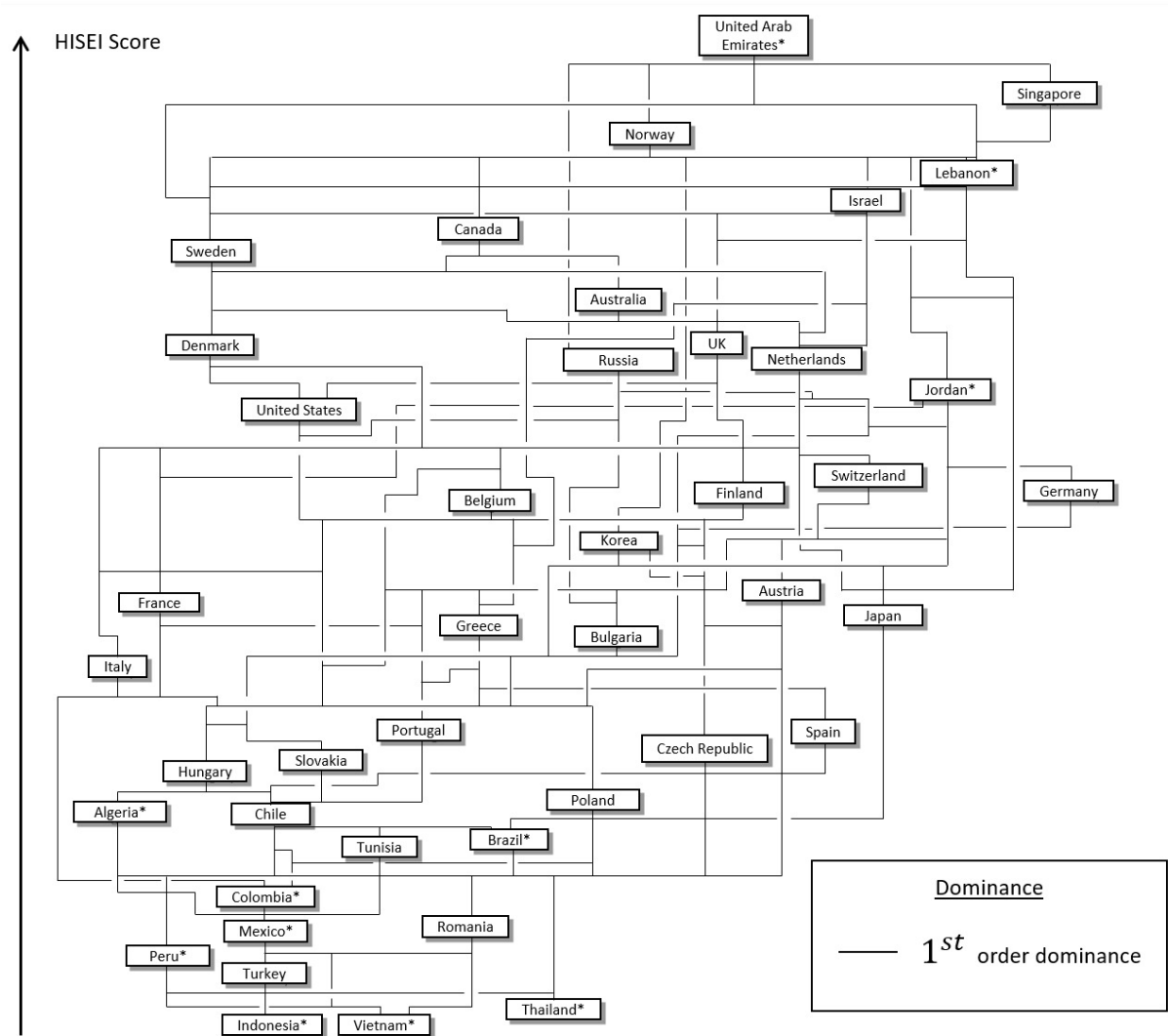
Figure 4: 1st order dominance ranking of countries based on their marginal distributions of family backgrounds (discretized HISEI).
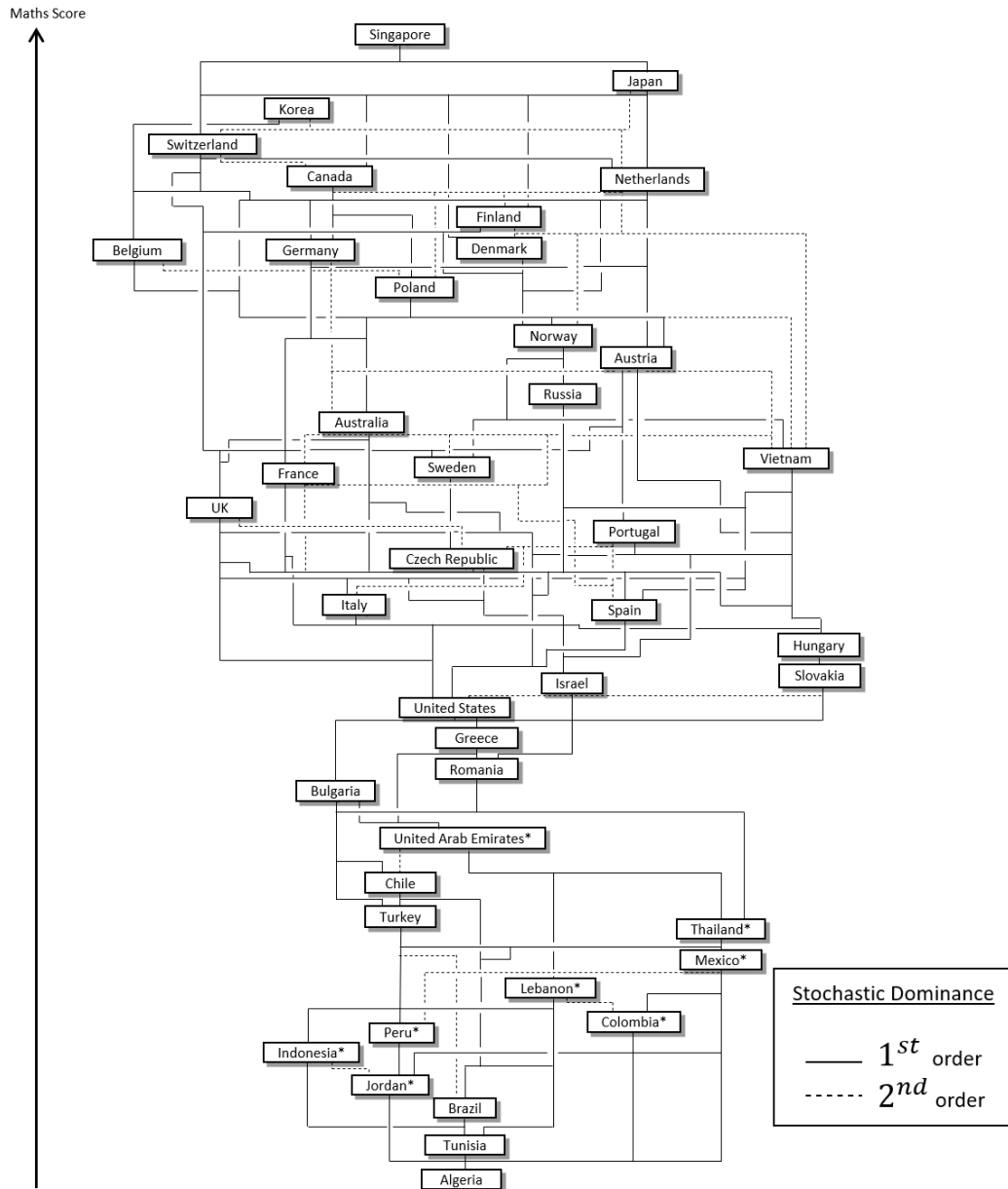
Figure 5: Ranking of countries based on their marginal distributions of children skills (1st order and 2nd order (elitist) dominance).

a country like Singapore can only be compared with the United Arab Emirates (the only country in the sample whose marginal distribution of HISEI first-order dominates that of Singapore). It is also interesting to notice that the ranking of countries provided on Figure 6 is less complete than the intersection of the rankings based on (inverse) first-order dominance applied to HISEI alone (as per Figure 4) and first and second order dominance applied to math scores only (Figure 5). For example the marginal distribution of math scores in Finland first-order dominates that of UK, while the marginal distribution of HISEI in the UK dominates the corresponding distribution in Finland. However, Finland does not dominate the UK as per the criteria of Theorems 1 and 2. The reason for this absence of dominance comes from the unfavorable correlation between skill and family background in Finland (as compared to UK). Put differently, the change in the British joint distribution of skills and background that would be needed to make it identical to that of Finland includes improvements in children' skills as per Definition 1 and deteriorations in family background as per Definition 2. If it did not, then one could not have dominance relations connecting the marginals of skills and backgrounds of the two countries. But the changes in the British distribution of backgrounds and skills that would be required to make it identical to that of Finland also involve additional features that do not reduce correlation between skill and background as per Definition 3. These unfavorable correlation features that prevent Finland from dominating UK can not be seen from a look at the standard correlation coefficients between skills and backgrounds provided by Table 2. Pearson Correlation between skills and background appears indeed to be marginally lower in the UK than in Finland but the ranking of the two countries is reversed if one uses the regression coefficient as a measure of correlation. Similar remarks can be made, using order 2 of dominance and the criterion C2, on Turkey and Brazil, Hungary and the US, France and Russia and Vietnam and Russia. The definite ranking of all other pairs of countries observed on Figure 6 is, however, *exactly* the intersection of the rankings by (inverse) first-order dominance of the marginal distributions of family background (Figure 4) and either first or second (elitist) dominance on the marginal of skills (Figure 5).

The resistance of Russia to domination by any other country illustrates well the role played by the aversion to correlation between skills and backgrounds in our criteria. Indeed, among the 43 countries depicted in Figure 6, Russia is the country where the correlation between skills and backgrounds - measured by either the regression coefficient or Pearson - is the lowest. There are many other countries beside Russia which, by dominating some others, and being dominated by none, appear at the "top" of the "dominance tree" of Figure 6. These countries are (in alphabetic order) Belgium, the Czech Republic, Finland, Germany, Indonesia, Japan, Korea, Poland, Portugal, Singapore, Switzerland, Thailand,Turkey and Vietnam. This latter country, who dominates 17 others, is clearly a major player in this top group. This prominence is explained by the fact that Vietnam pupils are by large coming from less favored families than those of the other countries. Since Vietnam pupils are doing rather well at PISA test, the Vietnam education system appears to outperform many others. This impressive performance of Vietnam is something that does not emerge spontaneously from a casual look at PISA rankings, in which the average Math Score of Vietnam (494) makes this country only a bit above the OECD average. However, considering that the average social status of Vietnamese parents is by far the lowest of all countries, the Math attainments of their children appears remarkable. To highlight just how impressive the attainments of Vietnamese children are at Math PISA test, one can compare them with those of children from countries with similar average parental social-status. Peru and Indonesia for example have HISEI averages roughly comparable to that of Vietnam. Yet their rather average standing at average of 368 and 375 at the Math PISA tests, making them the 2 lowest ranked countries of the official PISA ranking. Another good perform in this "top club" is Poland who dominates 12 other countries, and Japan (who dominates nine). At the bottom of the tree, one finds countries dominated by several others, but who dominate none. Members
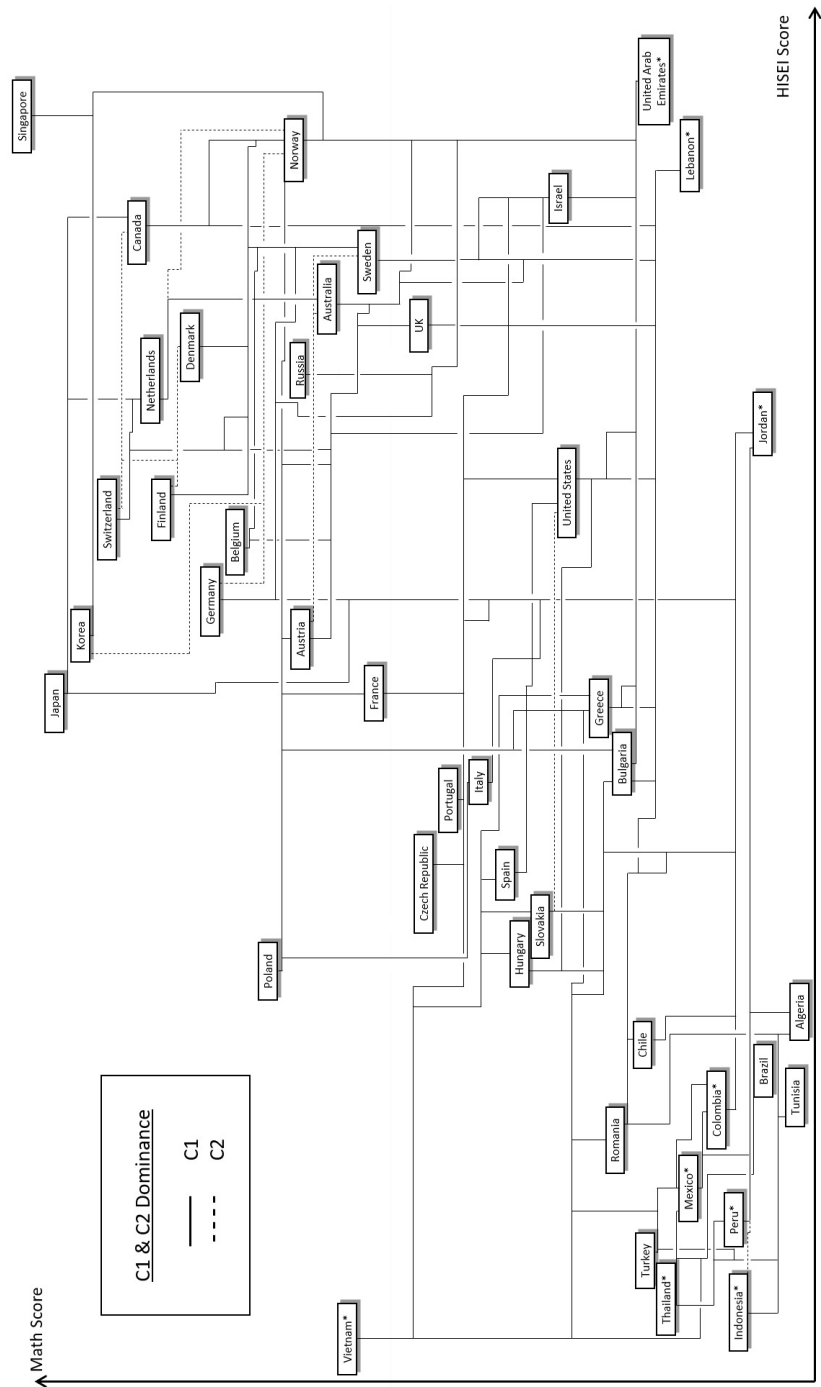
21

Figure 6: Rankings of countries as per the criteria of Theorems 1 and 2.

of this "bottom" pool are Algeria, Brazil, Jordan, Lebanon and UAE. Particularly bad per-

formance are that of the UAE (dominated by 29 countries), Lebanon (dominated by 25) and Jordan (dominated by 18). The main reason for this bad performance is the polar opposite of what explains the good performance of Vietnam .Indeed, the children trained in the education systems of Jordan, Lebanon and UAE countries are coming from family with relatively good HISEI categories, even though they perform somewhat poorly in PISA Math tests.

# 4    Conclusion

This paper has provided and implemented a methodology for appraising the performance of education systems, described by the (joint) distribution of cognitive skills and family backgrounds of the children that they educate. The methodology stands on a wide consensus about principles that could underlie such an appraisal. Three principles were given a particular emphasis:

1) Favorable recording of any improvement in the skill of a child with a given background.

2) A given skill acquired by a child is appraised all the more favorably as the child is coming from a relatively unfavorable background and,

3) For any two children differing in both their skill and background, it is preferable that the child with the low background be given the high skill than the other way around.

We identified a simple test that coincides with the unanimity of all anonymous transitive rankings of education systems that agree with these principles. The test amounts to verifying, for any levels of background and skill, if there is a higher fraction of children from a lower background and a better skill in one system than in another. Because the unanimity of rankings that satisfy these three principles is large, the test that corresponds exactly to it is not extremely discriminatory. We have therefore also considered the possibility of complementing these three basic principles with the more controversial elitist idea that a given mean preserving spread in the skill of two children with the same background should also been considered favorably. We have also identified an empirically testable criterion that coincides with the unanimity of all transitive and anonymous rankings that are compatible with elitism and the three other principles. The test amounts to verifying two sets of conditions:

1) If the expected success of children of a given background over any target assigned to their background is larger in the dominating than in the dominated distribution for any assignment of skill target that is increasing with the background and

2) if the fraction of children coming from a lower background than any given level is larger in the dominating then in the dominated distribution.

We have also illustrated the usefulness of our criteria by comparing the distributions of PISA scores in mathematics and family backgrounds based on the 2015 wave of PISA data. In this empirical application, we have taken the highest of the two parents' index of social status as the measure of background. The empirical analysis reveals that the criteria are easy to use, and do generate interesting conclusions. The most salient of them is the excellent performance of Vietnam . While the 19% of conclusive rankings obtained with the more demanding elitist criterion can be considered disappointing, the robustness of the conclusions are worth stressing. In effect, any conceivable assessment of national education systems that records favorably the three main principles above as well as dispersions in the children skill will agree with those conclusions. The fraction of conclusive rankings falls at 17.5% when one focuses only on the three core principles.

Two limitations of our analysis are, however, worth stressing.

The first one is our inability to find an operational criterion that would have replaced elitism by inequality aversion. This inability, that seems to be based on somewhat difficult features of the two-dimensional dominance theory developed in this paper, is clearly disappointing. Problems

of similar nature where encountered by Muller and Trannoy (2012) in a somewhat different context. The difficulty clearly calls for future research in theoretical dominance analysis. Why is it that some combinations of elementary principles convert better than others into easy-to-use dominance criteria ?

A second limitation concerns the relevance of the country as a natural example of an education system. Consider for example the extremely favorable appraisal of Vietnam that results from our empirical analysis. Is the high performance of Vietnam really the result of exceptionally well-designed education institutions developed in this country ? Or is it not, at least to some extent, the result also of a cultural attitude toward skill acquisition that has little to do with the actual working of Vietnamese education institutions? While the country level may not be the more appropriate to appraise the performance of education systems, it is our hope that researchers will find the methodology developed herein useful for making the required more-fine grained assessment of education systems performance.

# 5 Appendix

## 5.1 Proof of Theorem 1

The proof proceeds in several steps. We first establish, in the following Lemma, that if one wants to compare two education systems $\mathbf{d}^{\mathbf{e}^*}$ and $\mathbf{d}^{\mathbf{e}^\circ}$ by means of an additively separable - across children - evaluation function in a way that is sensitive to improvements in a child's cognitive skill, deteriorations in a child's background and reductions in correlation between skills and background, then it is necessary and sufficient that the evaluation function satisfies specific monotonicity and submodular properties.

**Lemma 1** *Let $\mathbf{d}^{\mathbf{e}^*}$ and $\mathbf{d}^{\mathbf{e}^\circ}$ be two education systems in $\mathcal{E}$ such that $\mathbf{d}^{\mathbf{e}^*}$ has been obtained from $\mathbf{d}^{\mathbf{e}^\circ}$ by either an improvement in a child's cognitive skill, a deterioration in family background, or a reduction in correlation between skill and background per Definitions 1-3. Then one has:*

$$\sum_{\beta=1}^{\overline{b}} \sum_{\sigma=1}^{\overline{z}} \mathbf{d}_{bs}^{\mathbf{e}^*} \Phi(\beta, \sigma) \geq \sum_{\beta=1}^{\overline{b}} \sum_{\sigma=1}^{\overline{z}} \mathbf{d}_{bs}^{\mathbf{e}^\circ} \Phi(\beta, \sigma) \tag{8}$$

*for a function $\Phi : \mathcal{B} \times \mathcal{S} \longrightarrow \mathbb{R}$ if and only if $\Phi$ satisfies:*
*(i) $\Phi(b, s+1) \geq \Phi(b, s)$ for all $b \in \mathcal{B}$ and $s \in \{1, ..., \overline{s} - 1\}$,*
*(ii) $\Phi(b-1, s) \geq \Phi(b, s)$ for all $b \in \{2, ..., \overline{b}\}$. and $s \in \mathcal{S}$ and,*
*(iii) $\Phi(s', b) - \Phi(s, b) \geq \Phi(s', b') - \Phi(s, b')$ for all $(b, s)$ and $(b', s')$ in $\mathcal{B} \times \mathcal{S}$ such that $b' \geq b$ and $s' \geq s$.*

**Proof.** *Assume first that $\mathbf{d}^{\mathbf{e}^*}$ has been obtained from $\mathbf{d}^{\mathbf{e}^\circ}$ by an improvement in a child's cognitive skill as per Definition 1. Then, there exists some $b \in \mathcal{B}$ and $s \in \{1, ..., \overline{s} - 1\}$ such that:*

$$\begin{aligned}
\mathbf{d}_{\beta\sigma}^{\mathbf{e}^*} &= \mathbf{d}_{\beta\sigma}^{\mathbf{e}^\circ} \forall (\beta, \sigma) \in (\mathcal{B} \times \mathcal{S}) \backslash \{b, s\}, (b, s+1) \\
\mathbf{d}_{bs}^{\mathbf{e}^*} &= \mathbf{d}_{bs}^{\mathbf{e}^\circ} - 1 \text{ and,} \\
\mathbf{d}_{bs+1}^{\mathbf{e}^*} &= \mathbf{d}_{bs}^{\mathbf{e}^\circ} + 1.
\end{aligned}$$

*Hence:*

$$\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e}^*}\Phi(\beta,\sigma) \quad \geq \quad \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e}^\circ}\Phi(\beta,\sigma)$$

$$\Longleftrightarrow$$

$$\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}(\mathbf{d}_{bs}^{\mathbf{e}^*}-\mathbf{d}_{bs}^{\mathbf{e}^\circ})\Phi(\beta,\sigma) \quad \geq \quad 0$$

$$\Longleftrightarrow$$

$$\Phi(s+1,b)-\Phi(s,b) \quad \geq \quad 0$$

*which, when applied to arbitrary $b \in \mathcal{B}$ and $s \in \{1,...,\overline{s}-1\}$ , is equivalent to (i). Assume now that $d^{\mathbf{e}^*}$ has been obtained from $d^{\mathbf{e}^\circ}$ by means of a deterioration in a child's family background as per Definition 2. Then, there exists some $b \in \{2,...,\overline{b}\}$ and $s \in \mathcal{S}$ such that:*

$$\begin{aligned}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^*} &= \mathbf{d}_{\beta\sigma}^{\mathbf{e}^\circ} \; \forall \; (\beta,\sigma) \in (\mathcal{B}\times\mathcal{S})\backslash\{(b-1,s),(b,s)\} \\ \mathbf{d}_{bs}^{\mathbf{e}^*} &= \mathbf{d}_{bs}^{\mathbf{e}^\circ}-1 \; and, \\ \mathbf{d}_{(b-1)s}^{\mathbf{e}^*} &= \mathbf{d}_{(b-1)s}^{\mathbf{e}^\circ}+1.\end{aligned}$$

*Hence:*

$$\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e}^*}\Phi(\beta,\sigma) \quad \geq \quad \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e}^\circ}\Phi(\beta,\sigma)$$

$$\Longleftrightarrow$$

$$\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}(\mathbf{d}_{bs}^{\mathbf{e}^*}-\mathbf{d}_{bs}^{\mathbf{e}^\circ})\Phi(\beta,\sigma) \quad \geq \quad 0$$

$$\Longleftrightarrow$$

$$\Phi(b-1,s)-\Phi(b,s) \quad \geq \quad 0$$

*which, when applied to arbitrary $b \in \{2,...,\overline{b}\}$ and $s \in \mathcal{S}$, is equivalent (ii). Finally, assume that $\mathbf{d}^{\mathbf{e}^*}$ has been obtained from $\mathbf{d}^{\mathbf{e}^\circ}$ by means of a reduction in correlation between skill and family background as per Definition 3. This means that there are $b$ and $b'$ in $\mathcal{B}$ and $s$ and $s'$ in $\mathcal{S}$ satisfying $b < b'$ and $s < s'$ such that:*

$$\begin{aligned}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^*} &= \mathbf{d}_{\beta\sigma}^{\mathbf{e}^\circ} for \; all \; (\beta,\sigma) \in (\mathcal{B}\times\mathcal{S})\backslash\{(b,s),(b',s),(b,s'),(b',s')\} \\ \mathbf{d}_{bs}^{\mathbf{e}^*} &= \mathbf{d}_{bs}^{\mathbf{e}^\circ}-1 \; and \; \mathbf{d}_{b's'}^{\mathbf{e}^*} = \mathbf{d}_{b's'}^{\mathbf{e}^\circ}-1 \; , \\ \mathbf{d}_{bs'}^{\mathbf{e}^*} &= \mathbf{d}_{bs'}^{\mathbf{e}^\circ}+1 \; and \; \mathbf{d}_{b's}^{\mathbf{e}^*} = \mathbf{d}_{b's}^{\mathbf{e}^\circ}+1\end{aligned}$$

*Hence:*

$$\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e}^*}\Phi(\beta,\sigma) \quad \geq \quad \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e}^\circ}\Phi(\beta,\sigma)$$

$$\Longleftrightarrow$$

$$\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}(\mathbf{d}_{bs}^{\mathbf{e}^*}-\mathbf{d}_{bs}^{\mathbf{e}^\circ})\Phi(\beta,\sigma) \quad \geq \quad 0$$

$$\Longleftrightarrow$$

$$\Phi(b,s')-\Phi(b,s)-[\Phi(b',s')-\Phi(b',s)] \quad \geq \quad 0$$

*as required by (iii).* ∎

In the next proposition, we establish,an equivalence between Statement (b) of Theorem (1) and the fact of observing, for the two education systems $\mathbf{d^{e^*}}$ and $\mathbf{d^{e^\circ}}$ mentioned in this statement, Inequality (8) for all functions $\Phi$ satisfying properties (i)-(iii) of Lemma 1. We specifically establish the following.

**Proposition 1** *Let* $\mathbf{d^{e^*}}$ *and* $\mathbf{d^{e^\circ}}$ *be two education systems in* $\mathcal{E}$*. Then, the two following statements are equivalent:*

*(a)* $\displaystyle\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{\beta\sigma}^{\mathbf{e^*}}\Phi(\beta,\sigma) \geq \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{\beta\sigma}^{\mathbf{e^\circ}}\Phi(\beta,\sigma)$ *for all functions* $\Phi : \mathcal{S}\times\mathcal{B} \longrightarrow \mathbb{R}$ *satisfying Properties (i)-(iii) of Lemma 1 and,*

*(b)* $\displaystyle\sum_{\beta=1}^{b}\sum_{\sigma=s}^{\overline{s}}\mathbf{d}_{\beta\sigma}^{\mathbf{e^*}} \geq \sum_{\beta=1}^{b}\sum_{\sigma=s}^{\overline{s}}\mathbf{d}_{\beta\sigma}^{\mathbf{e^\circ}}$ *for all* $(b,s) \in \mathcal{B}\times\mathcal{S}.$.

**Proof.** *(a)* $\implies$ *(b).*
*Assume that the inequality:*

$$\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e^*}}\Phi(\beta,\sigma) \geq \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e^\circ}}\Phi(\beta,\sigma) \tag{9}$$

*holds for all functions* $\Phi$ *satisfying Properties (i)-(iii) of Lemma 1. In particular therefore, inequality (9) holds for the function* $\Phi^{sb}$ *defined, for any* $(s,b) \in S\times B$*, by:*

$$\begin{aligned}\Phi^{bs}(\beta,\sigma) &= 1 \ \textit{if } \beta \leq b \textit{ and } \sigma \geq s\\ &= 0 \ \textit{otherwise}\end{aligned} \tag{10}$$

*which clearly satisfies Properties (i)-(iii) of Lemma 1. Since Inequality (9) holds for any such function* $\Phi^{sb}$*, one has, for every* $(b,s) \in \mathcal{B}\times\mathcal{S}$*:*

$$\begin{aligned}\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e^*}}\Phi^{bs}(\beta,\sigma) \quad &\geq \quad \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{bs}^{\mathbf{e^\circ}}\Phi^{bs}(\beta,\sigma)\\ &\Longleftrightarrow\\ \sum_{\beta=1}^{b}\sum_{\sigma=s}^{\overline{s}}\mathbf{d}_{\beta\sigma}^{\mathbf{e^*}} \quad &\geq \quad \sum_{\beta=1}^{b}\sum_{\sigma=s}^{\overline{s}}\mathbf{d}_{\beta\sigma}^{\mathbf{e^\circ}}\end{aligned}$$

*for all such* $(b,s)$ *as required by (b).*
*(b)* $\implies$ *(a).*
*Assume that (b) holds. In order to demonstrate that Inequality (9) must hold for all functions* $\Phi$ *satisfying Properties (i)-(iii) of Lemma 1, we start by writing Inequality (9) as:*

$$\sum_{b=1}^{\overline{b}}\sum_{s=1}^{\overline{s}}\Delta\mathbf{d}_{bs}\Phi(b,s) \geq 0 \tag{11}$$

*where, for every* $(b,s) \in S\times B$*:*

$$\Delta\mathbf{d}_{bs} = \mathbf{d}_{bs}^{\mathbf{e^*}} - \mathbf{d}_{bs}^{\mathbf{e^\circ}}$$

*Inverting the order of the summation of the skill variable, we can write alternatively Inequality (11) as:*

$$\sum_{b=1}^{\overline{b}} \sum_{s=\overline{s}}^{1} \Delta \mathbf{d}_{bs} \Phi(b,s) \geq 0 \tag{12}$$

*Doing a discrete (or Abelian, see e.g. Fishburn and Vickson (1978) eq. 2.49), sum by part of Expression (12) yields (after exploiting the fact that $\sum_{b=1}^{\overline{b}} \sum_{s=\overline{s}}^{1} \Delta d_{bs} = 0$):*

$$\sum_{b=1}^{\overline{b}} \sum_{s=\overline{s}}^{2} \Delta \mathbf{d}_{bs} \Phi_2(\overline{b}, s-1)$$
$$-\sum_{b=1}^{\overline{b}-1} \sum_{s=\overline{s}}^{1} \Delta \mathbf{d}_{bs} \Delta \Phi_1(b,1)$$
$$-\sum_{b=1}^{\overline{b}-1} \sum_{\beta=1}^{b} \sum_{s=\overline{s}}^{2} \sum_{\sigma=\overline{s}}^{s} \Delta \mathbf{d}_{\beta\sigma} \Phi_{12}(b, s-1)$$
$$\geq \quad 0 \tag{13}$$

*where, for every $(b,s) \in \{1, ..., \overline{b}-1\} \times \{1, ..., \overline{s}-1\}$:*

$$\Phi_1(b,s) = \Phi(b+1,s) - \Phi(b,s)$$
$$\Phi_2(s,b) = \Phi(b,s+1) - \Phi(b,b)$$

*denote the partial discrete (to the right) difference of $\Phi$ with respect to family background and skill respectively and where:*

$$\Phi_{12}(b,s) = \Phi(b+1,s+1) - \Phi(b+1,s) - \Phi(b,s+1) + \Phi(b,s) \tag{14}$$

*denote the discrete (to the right) difference of difference of $\Phi$ first with respect to background and second with respect to skill. Hence, a sufficient condition for (13) to hold for all functions $\Phi$ satisfying $\Phi_1 \geq 0 \geq \Phi_2$ and $\Phi_{12} \leq 0$ as required by Properties (i)-(iii) of Lemma 1 is to have:*

$$\sum_{\beta=1}^{b} \sum_{\sigma=\overline{s}}^{s} \Delta \mathbf{d}_{\beta\sigma} \geq 0$$

*for all $s \in \{1, ..., \overline{s}\}$ and $b \in \{1, ..., \overline{b}\}$, as required by Statement 2 of Theorem 1.* ∎

We propose to prove Theorem 1 as a corollary of the just proved Proposition 1 and the next Proposition, which establishes an equivalence between the fact of going from an education system to another by a finite sequence of improvements in cognitive skills, deteriorations in family backgrounds and reductions in correlation between cognitive skill and family background on the one hand, and the fact of observing Inequality 9 for all functions $\Phi$ satisfying Properties (i)-(iii) of Lemma 1 on the other. In order to state and prove this Proposition, we introduce some additional notation and auxiliary results.

We start by defining the binary relation $\succsim_1$ on $\mathcal{E}$ as follows: $\mathbf{d^{e^*}} \succsim_1 \mathbf{d^{e^\circ}}$ if and only if $\mathbf{d^{e^*}}$ has been obtained from $\mathbf{d^{e^\circ}}$ by either (non-inclusively) an improvement in a child's cognitive skill, a deterioration in family background, or a reduction in correlation between skill and background per Definitions 1-3.

This binary relation is anti-symmetric and reflexive but is not transitive. The transitive closure of $\succsim_1$, denoted $\widehat{\succsim}_1$, and defined by:

$$
\mathbf{d^{e^*}} \ \widehat{\succsim}_1 \ \mathbf{d^{e^\circ}} \quad \Longleftrightarrow \quad \exists \text{ a sequence } \{\mathbf{d^{et}}\}_{t=0}^{\bar{t}} \text{ of education systems such that:}
$$
$$
\mathbf{d^{e0}} \ = \ \mathbf{d^{e^*}},
$$
$$
\mathbf{d^{e\bar{t}}} \ = \ \mathbf{d^{e^\circ}} \text{ and}
$$
$$
\mathbf{d^{et}} \ \succsim \ {}_1\mathbf{d^{et+1}} \text{ for } t = 0, ..., \bar{t} - 1
$$

We now observe that the fact for education system $\mathbf{d^{e^*}}$ to result from system $\mathbf{d^{e^\circ}}$ by means of either an improvement in a child's cognitive skill, a deterioration in family background, or a reduction in correlation between skill and background per Definitions 1-3 - that is $\mathbf{d^{e^*}} \succsim_1 \mathbf{d^{e^\circ}}$ - can equivalently be described by the fact that the difference $\boldsymbol{\Delta d} = \mathbf{d^{e^*}} - \mathbf{d^{e^\circ}}$ belongs to some set $\mathcal{T}^1 \subset \mathbb{Z}^{\overline{bs}}$. Define specifically the sets $\mathcal{T}_{skill}$, $\mathcal{T}_{fam}$ and $\mathcal{T}_{correl}$ by:

$$
\mathcal{T}_{skill} \ = \ \{\Delta\mathbf{d} \in \mathbb{Z}^{\overline{bs}} : \Delta\mathbf{d}_{\beta\sigma} = 0 \ \forall \ (\beta, \sigma) \in (\mathcal{B} \times \mathcal{S}) \backslash \{(b,s), (b, s+1)\},
$$
$$
\Delta\mathbf{d}_{bs} \ = \ -1 \ \text{ and } \Delta\mathbf{d}_{bs+1} = 1 \text{ for some } (b,s) \in \mathcal{B} \times \mathcal{S} \backslash \{\overline{s}\}\}
$$

$$
\mathcal{T}_{fam} \ = \ \{\Delta\mathbf{d} \in \mathbb{Z}^{\overline{bs}} : \Delta\mathbf{d}_{\beta\sigma} = 0 \ \forall \ (\beta, \sigma) \in (\mathcal{B} \times \mathcal{S}) \backslash \{(b-1,s), (b, s)\},
$$
$$
\Delta\mathbf{d}_{bs} \ = \ -1 \ \text{ and } \Delta\mathbf{d}_{bs+1} = 1 \text{ for some } (b,s) \in \mathcal{B} \backslash \{1\} \times \mathcal{S}\}
$$

$$
\mathcal{T}_{correl} \ = \ \{\Delta\mathbf{d} \in \mathbb{Z}^{\overline{bs}} : \Delta\mathbf{d}_{\beta\sigma} = 0 \ \forall \ (\beta, \sigma) \in (\mathcal{B} \times \mathcal{S}) \backslash \{(b,s), (b',s), (b,s'), (b',s')\},
$$
$$
\Delta\mathbf{d}_{bs} \ = \ -1 = \Delta\mathbf{d}_{b's'} \ \text{ and}
$$
$$
\Delta\mathbf{d}_{bs'} \ = \ 1 = \Delta\mathbf{d}_{b's} \text{ for some } (b,s) \text{ and } (b',s') \in \mathcal{B} \times \mathcal{S} \text{ such that } (b,s) \ << \ (b',s')\}
$$

Each of these set is finite, as is the set $\mathcal{B}(\mathcal{T}^1) = \mathcal{T}_{skill} \cup \mathcal{T}_{fam} \cup \mathcal{T}_{correl}$. Hence $\mathbf{d^{e^*}} - \mathbf{d^{e^\circ}} \in \mathcal{B}(\mathcal{T}^1)$ if and only if $\mathbf{d^{e^*}} \succsim_1 \mathbf{d^{e^\circ}}$. Define now the set $\mathcal{T}^1$ as follows:

$$
\mathcal{T}^1 = \{m \in \mathbb{Z}^{\overline{bs}} : m = \sum_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} \lambda_t m_t \text{ for } \lambda_t \in \mathbb{N} \text{ and } m_t \in \mathcal{B}(\mathcal{T}^1)\} \tag{15}
$$

Hence $\mathcal{T}^1$ is the set of all differences in education systems $\mathbf{d^{e^*}} - \mathbf{d^{e^\circ}}$ (whatever the involved education systems $\mathbf{d^{e^*}}$ and $\mathbf{d^{e^\circ}}$ are) generated by the fact of obtaining $\mathbf{d^{e^*}}$ from $\mathbf{d^{e^\circ}}$ by a finite sequence of improvements in a child's cognitive skill, deteriorations in family background and reductions in correlation between skill and background as per Definitions 1-3. Put differently, $\mathbf{d^{e^*}} - \mathbf{d^{e^\circ}} \in \mathcal{T}^1$ if and only if $\mathbf{d^{e^*}}$ $\widehat{\succsim}_1 \mathbf{d^{e^\circ}}$. It must be remarked that $0 \in \mathcal{T}^1$. It can also be noticed that $\mathcal{B}(\mathcal{T}^1)$ is a finite basis of the set $\mathcal{T}^1$ in the (obvious sense) that any $m \in \mathcal{T}^1$ can be written as $\sum_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} \lambda_t m_t$ for $\lambda_t \in \mathbb{R}_+$, $m_t \in \mathcal{B}(\mathcal{T}^1)$ and $t = 1, ..., \#\mathcal{B}(\mathcal{T}^1)$. As can be noticed also, one can never have $m = m' + m"$ for $m, m'$ and $m" \in \mathcal{B}(\mathcal{T}^1)$. Hence, the elements of $\mathcal{B}(\mathcal{T}^1)$ are "irreducible" and $\mathcal{B}(\mathcal{T}^1)$ is therefore a Hilbert basis of the set $\mathcal{T}^1$ (see e.g. Cook, Fonlupt, and Schrijver (1986)).

We now show that the set $\mathcal{T}^1$ is a discrete and pointed convex cone. That is $\mathcal{T}^1$ is such that for any $m$ and $m'$ such that $m \in \mathcal{T}^1$ and $m' \in \mathcal{T}^1$, it is the case that $\lambda_1 m + \lambda_2 m' \in \mathcal{T}^1$ for all $\lambda_1$ and $\lambda_2 \in \mathbb{N}$ (discrete cone) and if $m \in \mathcal{T}^1$ and $-m \in \mathcal{T}^1$, one has $m = 0$ (pointed cone).

**Lemma 2** *The set $\mathcal{T}^1$ is a discrete and pointed cone. That is, it is such that for any $m$ and $m'$ such that $m \in \mathcal{T}^1$ and $m' \in \mathcal{T}^1$, it is the case that $\lambda_1 m + \lambda_2 m' \in \mathcal{T}^1$ for all $\lambda_1$ and $\lambda_2 \in \mathbb{N}$ and $m = 0$ if $m \in \mathcal{T}^1$ and $-m \in \mathcal{T}^1$.*

**Proof.** *Suppose that $m$ and $m'$ belong to $\mathcal{T}^1$. Using Expression (15), one has $m = \sum\limits_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} \lambda_t m_t$*

*and $m' = \sum\limits_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} \lambda'_t m_t$ for $\{\lambda'_t\} \cup \{\lambda_t\} \subset \mathbb{N}$ and $m_t \in \mathcal{B}(\mathcal{T}^1)$. For any $\lambda_j \in \mathbb{N}$ $(j = 1, 2)$.*

*one has $\lambda_1 m + \lambda_2 m' = \sum\limits_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} (\lambda_1 \lambda_t + \lambda_2 \lambda'_t) m_t$ for $m_t \in \mathcal{B}(\mathcal{T}^1)$ and $(\lambda_1 \lambda_t + \lambda_2 \lambda'_t) \in \mathbb{N}$ and*

*$t = 1, ..., \#\mathcal{B}(\mathcal{T}^1)$, which implies (using Expression (15) again) that $\lambda_1 m + \lambda_2 m' \in \mathcal{T}^1$. Hence $\mathcal{T}^1$ is a discrete cone. To show that it is pointed, consider any $m$ such that $m \in \mathcal{T}^1$ and $m \neq 0$. This means that $m = \sum\limits_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} \lambda_t m_t$ for $\lambda_t \in \mathbb{N}$ and $m_t \in \mathcal{B}(\mathcal{T}^1)$ and $-m = \sum\limits_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} \lambda'_t m_t$ for $\lambda'_t \in \mathbb{N}$ and $m_t \in \mathcal{B}(\mathcal{T}^1)$*

*for $t = 1, ...., \#\mathcal{B}(\mathcal{T}^1)$. One therefore has:*

$$\sum_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} \lambda_t m_t \quad = \quad - \sum_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} \lambda'_t m_t$$

$$\Longleftrightarrow$$

$$\sum_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} (\lambda_t + \lambda'_t) m_t \quad = \quad 0-$$

*For this equality to hold for some $(\lambda_t + \lambda'_t) > 0$, there must exists some $m_t$ and $m' \in \mathcal{B}(\mathcal{T}^1)$ such that $m_t = -\gamma m_{t'}$ for some strictly positive $\gamma$. But this is impossible given the definition of the elementary operations associated to the set $\mathcal{B}(\mathcal{T}^1)$.* ∎

With these preliminaries, we are ready to prove the following proposition, which, when combined with the equivalence established in Proposition 1, proves Theorem 1.

**Proposition 2** *.Let $\mathbf{d}^{\mathbf{e}^*}$ and $\mathbf{d}^{\mathbf{e}^\circ}$ be two education systems in $\mathcal{E}$. Then, the two following statements are equivalent:*

*(a) $\sum\limits_{\beta=1}^{\overline{b}} \sum\limits_{\sigma=1}^{\overline{z}} \mathbf{d}^{\mathbf{e}^*}_{\beta\sigma} \Phi(\beta, \sigma) \geq \sum\limits_{\beta=1}^{\overline{b}} \sum\limits_{\sigma=1}^{\overline{z}} \mathbf{d}^{\mathbf{e}^\circ}_{\beta\sigma} \Phi(\beta, \sigma)$ for all functions $\Phi : \mathcal{S} \times \mathcal{B} \longrightarrow \mathbb{R}$ satisfying Properties (i)-(iii) of Lemma 1 and*
*(b) $\mathbf{d}^{\mathbf{e}^*} - \mathbf{d}^{\mathbf{e}^\circ} \in T^1$.*

**Proof.** *The fact that (b) implies (a) is an immediate consequence of Lemma 1 and the fact that $\mathbf{d}^{\mathbf{e}^*} - \mathbf{d}^{\mathbf{e}^\circ} \in T^1$ if and only if $\mathbf{d}^{\mathbf{e}^*} \widehat{\succsim}_1 \mathbf{d}^{\mathbf{e}^\circ}$. We therefore only prove that (a) implies (b). For this sake, we write (a) as:*

$$\sum_{\beta=1}^{\overline{b}} \sum_{\sigma=1}^{\overline{z}} \Delta \mathbf{d}_{\beta\sigma} \Phi(\beta, \sigma) \geq 0$$

*for all $\Phi : \mathcal{S} \times \mathcal{B} \longrightarrow \mathbb{R}$ satisfying Properties (i)-(iii) of Lemma 1. Let $\mathcal{F}$ denote the set of all such functions. It is easy to check that $\mathcal{F}$ is a linear space, as is obviously $\mathbb{R}^{\overline{bs}}$ in which the discrete and*

*pointed cone $\mathcal{T}^1$ lies. It is also reasonably straightforward to check that the two linear spaces $\mathcal{F}$ and $\mathbb{R}^{\overline{bs}}$ are in strict duality under the bilinear mapping $g : \mathbb{R}^{\overline{bs}} \times \mathcal{F} \to \mathbb{R}$ defined by:*

$$g(m, \Phi) = \sum_{\beta=1}^{\overline{b}} \sum_{\sigma=1}^{\overline{z}} m_{\beta\sigma} \Phi(\beta, \sigma)$$

*That is, for all $\Phi : \mathcal{S} \times \mathcal{B} \longrightarrow \mathbb{R}$ satisfying Properties (i)-(iii) of Lemma 1 such that $\Phi(\beta, \sigma) \neq \Phi(\beta', \sigma')$ for at least two distinct pairs $(\beta, \sigma)$ and $(\beta', \sigma')$ in $\mathcal{S} \times \mathcal{B}$, there exists some $m \in T^1$ such that $g(m, \Phi) \neq 0$ and, conversely,for any $m \in T^1$ such that $m \neq 0$, one can find some function $\Phi \in \mathcal{F}$ such that $g(m, \Phi) \neq 0$. We now observe that, by virtue of Lemma 1, $\mathcal{F}$ is the dual cone of the set $\mathcal{T}^1$ under the function $g(m, \Phi)$. That is, $\mathcal{F} = \{\Phi : \mathcal{S} \times \mathcal{B} \longrightarrow \mathbb{R} : g(m, \Phi) \geq 0 \text{ for all } m \in T^1\}$. By assumption, $\sum_{\beta=1}^{\overline{b}} \sum_{\sigma=1}^{\overline{z}} \Delta\mathbf{d}_{\beta\sigma} \Phi(\beta, \sigma) \geq 0$ for all $\Phi \in \mathcal{F}$. Hence $\mathbf{d}^{\mathbf{e}^*} - \mathbf{d}^{\mathbf{e}^\circ}$ belongs to the dual cone of $\mathcal{F}$ or, equivalently, to the dual cone of the dual cone of $\mathcal{T}^1$. Call this set $\mathcal{T}^{1**}$. By virtue of the Bipolar theorem (see Wikipedia or,for example,Muller and Scarsini (2012) (Theorem 4.3)), $\mathcal{T}^{1**}$ is the smallest closed set containing the convex cone generated by $\mathcal{T}^1$. The convex cone generated by $\mathcal{T}^1$, denoted $Co(\mathcal{T}^1)$, is defined by*

$$Co(\mathcal{T}^1) = Co\{\lambda m : \lambda \in \mathbb{R}_+, \, m \in \mathcal{T}^1\}$$

*where $Co(A)$ denote the convex hull of $A$. Now $\mathbf{d}^{\mathbf{e}^*} - \mathbf{d}^{\mathbf{e}^\circ} \in \mathcal{T}^{1**}$ and $\mathbf{d}^{\mathbf{e}^*} - \mathbf{d}^{\mathbf{e}^\circ} \in Z^{\overline{bs}}$. Since $\mathcal{T}^1 \subset Z^{\overline{bs}}$, the set $Co(\mathcal{T}^1)$ is obviously a rational cone in the sense that it is generated by the rational pointed cone $\mathcal{T}^1$. By Hilbert theorem, the rational pointed cone $\mathcal{T}^1$ has a unique Hilbert basis $\mathcal{B}(\mathcal{T}^1)$ that spans $Co(\mathcal{T}^1)$ (and $\mathcal{T}^1$). Since $\mathbf{d}^{\mathbf{e}^*} - \mathbf{d}^{\mathbf{e}^\circ} \in \mathcal{T}^{1**} \cap Z^{\overline{bs}}$, $\mathbf{d}^{\mathbf{e}^*} - \mathbf{d}^{\mathbf{e}^\circ} \in \mathcal{T}^1$ and this completes the proof.* ∎

## 5.2   Proof of Theorem 2

The proof strategy follows that used in the proof of Theorem 1. We start by establishing, in the following extension of Lemma 1, that the comparison of two education systems $\mathbf{d}^{\mathbf{e}^*}$ and $\mathbf{d}^{\mathbf{e}^\circ}$ by means of an additively separable - across children - evaluation function in a way that is sensitive to improvements in a child's cognitive skill, deteriorations in a child's background, reductions in correlation between skills and background and dispersions in children cognitive skill requires the evaluation function to satisfy monotonicity, submodularity and restricted convexity properties. We specifically prove the following.

**Lemma 3** *Let $\mathbf{d}^{\mathbf{e}^*}$ and $\mathbf{d}^{\mathbf{e}^\circ}$ be two education systems in $\mathcal{E}$ such that $\mathbf{d}^{\mathbf{e}^*}$ has been obtained from $\mathbf{d}^{\mathbf{e}^\circ}$ by either an improvement in a child's cognitive skill, a deterioration in family background, a reduction in correlation between skill and background or a dispersion in children cognitive skills as per Definitions 1-4. Then one has:*

$$\sum_{\beta=1}^{\overline{b}} \sum_{\sigma=1}^{\overline{z}} \mathbf{d}_{bs}^{\mathbf{e}^*} \Phi(\beta, \sigma) \geq \sum_{\beta=1}^{\overline{b}} \sum_{\sigma=1}^{\overline{z}} \mathbf{d}_{bs}^{\mathbf{e}^\circ} \Phi(\beta, \sigma) \tag{16}$$

*for a function $\Phi : \mathcal{B} \times \mathcal{S} \longrightarrow \mathbb{R}$ if and only if $\Phi$ satisfies:*
*(i) $\Phi(s+1, b) \geq \Phi(s, b)$ for all $b \in \mathcal{B}$ and $s \in \{1, ..., \overline{s} - 1\}$,*
*(ii) $\Phi(b-1, s) \geq \Phi(b, s)$ for all $b \in \{2, ..., \overline{b}\}$.and $s \in \mathcal{S}$,*
*(iii) $\Phi(s', b) - \Phi(s, b) \geq \Phi(s', b') - \Phi(s, b')$ for all $(b, s)$ and $(b', s')$ in $\mathcal{B}$ such that $b' \geq b$ and $s' \geq s$ and,*
*(iv) $\Phi(s', b) - \Phi(s' - 1, b) \geq \Phi(s+1, b) - \Phi(s, b)$ for all $b$ in $\mathcal{B}$ and $s$ and $s'$ in $\mathcal{S}$ such that $s' > s+1$.*

**Proof.** *The proof of the necessity and sufficiency of conditions (i)-(iii) for having Inequality (16) with any two education systems $\mathbf{d^{e^*}}$ and $\mathbf{d^{e^\circ}}$ such that $\mathbf{d^{e^*}}$ has been obtained from $\mathbf{d^{e^\circ}}$ by either an improvement in a child's cognitive skill, a deterioration in family background, a reduction in correlation between skill and background has already been established in Lemma 1. We therefore only establish here the necessity and sufficiency of (iv) for the satisfaction of Inequality (16) for education systems $\mathbf{d^{e^*}}$ and $\mathbf{d^{e^\circ}}$ such that $\mathbf{d^{e^*}}$ has been obtained from $\mathbf{d^{e^\circ}}$ by a dispersion in cognitive skill as per Definition 4. Let indeed $\mathbf{d^{e^*}}$ and $\mathbf{d^{e^\circ}}$ be two such systems for which, as per Definition 4, there is a background $b \in \mathcal{B}$ and skill levels $s$ and $s'$ in $\mathcal{S}$ satisfying $s' > s+1$ such that :4*

$$\mathbf{d}^{e^*}_{\beta\sigma} = \mathbf{d}^{e^\circ}_{\beta\sigma} \text{ for all } (\beta,\sigma) \in (\mathcal{B} \times \mathcal{S})\backslash(\{(b,s'),(b,s'+1)\} \cup \{(b,s-1),(b,s)\})$$
$$\mathbf{d}^{e^*}_{bs'} = \mathbf{d}^{e^\circ}_{bs'}+1, \; \mathbf{d}^{e^*}_{bs'-1} = \mathbf{d}^{e^\circ}_{bs'-1}-1 \;,\; \mathbf{d}^{e^*}_{bs} = \mathbf{d}^{e^\circ}_{bs}+1 \text{ and } \mathbf{d}^{e^*}_{bs+1} = \mathbf{d}^{e^\circ}_{bs+1}-1$$

*if $s'-1 > s+1$ and:*

$$\mathbf{d}^{e^*}_{bs'} = \mathbf{d}^{e^\circ}_{bs'}+1, \; \mathbf{d}^{e^*}_{bs'-1} = \mathbf{d}^{e^*}_{bs+1} = \mathbf{d}^{e^\circ}_{bs'-1}-2 \;,\; \mathbf{d}^{e^*}_{bs} = \mathbf{d}^{e^\circ}_{bs}+1$$

*if $s'-1 = s+1$. Hence inequality (16) holds with such $\mathbf{d^{e^*}}$ and $\mathbf{d^{e^\circ}}$ if and only if one has*

$$\Phi(s',b) - \Phi(s'-1,b) - \Phi(s+1,b) + \Phi(s,b) \geq 0$$

*for any $b \in \mathcal{B}$ and skill levels $s$ and $s'$ in $\mathcal{S}$ satisfying $s' > s+1$, which is nothing else than what is required by (iv).* ∎

We now establish the analogue of Proposition 3 of the last subsection. Specifically, we prove the equivalence between Statement (b) of Theorem 2 and the fact of observing, for the two education systems $\mathbf{d^{e^*}}$ and $\mathbf{d^{e^\circ}}$ mentioned in this statement, Inequality (16) for all functions $\Phi$ satisfying properties (i)-(iv) of Lemma 3. We specifically establish the following.

**Proposition 3** *Let $\mathbf{d^{e^*}}$ and $\mathbf{d^{e^\circ}}$ be two education systems in $\mathcal{E}$. Then, the two following statements are equivalent:*

*(a)* $\displaystyle\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}^{e^*}_{\beta\sigma}\Phi(\beta,\sigma) \geq \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}^{e^\circ}_{\beta\sigma}\Phi(\beta,\sigma)$ *for all functions $\Phi : \mathcal{S} \times \mathcal{B} \longrightarrow \mathbb{R}$ satisfying Properties (i)-(iv) of Lemma 3 and,*

*(b)* $\displaystyle\sum_{b\in\mathcal{B}} S^{\mathbf{e}}(b,s(b)) \geq \sum_{b\in\mathcal{B}} S^{\mathbf{e}^\circ}(b,s(b))$ *for all increasing functions $s : \mathcal{B} \to \mathcal{S}$ and $\displaystyle\sum_{\beta=1}^{b}\sum_{\sigma=1}^{\overline{s}}\mathbf{d}^{e^*}_{\beta\sigma} \geq \sum_{\beta=1}^{b}\sum_{\sigma=1}^{\overline{s}}\mathbf{d}^{e^\circ}_{\beta\sigma}$ for every background $b$.*

**Proof.** *(a) $\Longrightarrow$ (b).*
Assume that $\mathbf{d^{e^*}}$ and $\mathbf{d^{e^\circ}}$ are two education systems in $\mathcal{E}$ for which Inequality (16) holds for all functions $\Phi : \mathcal{S} \times \mathcal{B} \longrightarrow \mathbb{R}$ satisfying Properties (i)-(iv) of Lemma 3. In particular therefore, inequality (16) holds for the function $\Phi^{s(\cdot)}$ defined, for any increasing function $s : \mathcal{B} \to \mathcal{S}$, by:

$$\Phi^{s(\cdot)}(b,s) = \max[s - s(b), 0]$$

Indeed, $\Phi^{s(.)}$ satisfies Properties (i)-(iv) of Lemma 3. Since Inequality (16) holds for any such function $\Phi^{s(.)}$, one has:

$$\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^*}\max[\sigma-s(\beta),0] \quad\geq\quad \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^{\circ}}\max[\sigma-s(\beta),0]$$
$$\Longleftrightarrow$$
$$\sum_{b\in\mathcal{B}}S^{\mathbf{e}}(s(b),b) \quad\geq\quad \sum_{b\in\mathcal{B}}S^{\mathbf{e}^{\circ}}(s(b),b)$$

as required by the first criterion of Statement (b). For the second criterion of this statement, simply observe that, for every $b \in \mathcal{B}$, the function $\Phi^b : \mathcal{S} \times \mathcal{B} \longrightarrow \mathbb{R}$ defined, for every $\sigma \in \mathcal{S}$, by:

$$\Phi^b(\beta,s) \quad=\quad 1 \;\; if \; \beta \leq b$$
$$=\quad 0 \;\; otherwise$$

also satisfies (more than often trivially) Properties (i)-(iv) of Lemma 3. Hence Inequality (16) holds for $\Phi^b$ so that one has, for every $b$:

$$\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^*}\Phi^b(\beta,\sigma) \quad\geq\quad \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^{\circ}}\Phi^b(\beta,\sigma)$$
$$\Longleftrightarrow$$
$$\sum_{\beta=1}^{b}\sum_{\sigma=1}^{\overline{s}}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^*} \quad\geq\quad \sum_{\beta=1}^{b}\sum_{\sigma=1}^{\overline{s}}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^{\circ}}$$

for every $b$, as required by the second part of Statement (b).

(b) $\Longrightarrow$ (a)

We proceed just as in the corresponding step of the proof of Proposition 1 (with the same notation) by writing Inequality (16) as:

$$\sum_{b=1}^{\overline{b}}\sum_{s=\overline{s}}^{1}\Delta\mathbf{d}_{bs}\Phi(b,s) \geq 0 \tag{17}$$

or, after performing an Abel decomposition of the inner term of Inequality (17) only:

$$\sum_{b=1}^{\overline{b}}\sum_{s=\overline{s}}^{1}\Delta\mathbf{d}_{bs}\Phi(b,1)$$
$$+\sum_{b=1}^{\overline{b}}\sum_{s=\overline{s}}^{2}\sum_{\sigma=\overline{s}}^{s}\Delta\mathbf{d}_{b\sigma}\Phi_2(b,s) \quad\geq\quad 0 \tag{18}$$

Paralleling the ingenious approach proposed by Bourguignon (1989) for his ordered poverty gap criterion, one can observe that the second term of the left hand side of Inequality (18) can be written as:

$$\sum_{b=1}^{\overline{b}}\sum_{s=\overline{s}}^{2}\sum_{\sigma=\overline{s}}^{s}\Delta\mathbf{d}_{b\sigma}\Phi_2(b,s) \quad=\quad \sum_{b=1}^{\overline{b}}\sum_{s=\overline{s}}^{2}\sum_{\sigma=\overline{s}}^{s}[\Delta\mathbf{d}_{b\sigma}+g_{b-1}^s-g_b^s]\Phi_2(b,s)$$
$$-\sum_{b=2}^{\overline{b}}\sum_{s=\overline{s}}^{2}\Phi_{12}(b-1,s-1)g_{b-1}^s \tag{19}$$

*for any list of $\overline{b}\overline{s}$ non-negative numbers $g_b^s$ (for $b = 1, ..., \overline{b}$ and $s = 1, ..., \overline{s}$) satisfying $g_0^s = g_{\overline{b}}^s = 0$ for every $s \in \{1, ..., l\}$. Doing an additional Abel decomposition of the first term of the right hand side of Expression (19) enables one to write this expression as:*

$$\sum_{b=1}^{\overline{b}}\sum_{s=\overline{s}}^{2}\sum_{\sigma=\overline{s}}^{s}\Delta\mathbf{d}_{b\sigma}\Phi_2(b,s) =$$

$$\sum_{b=1}^{\overline{b}}[\sum_{s=\overline{s}}^{2}\sum_{\sigma=\overline{s}}^{s}\Delta\mathbf{d}_{b\sigma} + \sum_{s=\overline{s}}^{2}g_{b-1}^s - \sum_{s=\overline{s}}^{2}g_b^s]\Phi_2(b,1)$$

$$-\sum_{b=1}^{\overline{b}}\sum_{s=\overline{s}}^{2}[\sum_{\sigma=\overline{s}}^{s}\sum_{j=\overline{s}}^{\sigma}\Delta\mathbf{d}_{bj} + \sum_{\sigma=\overline{s}}^{s}g_{b-1}^\sigma - \sum_{\sigma=\overline{s}}^{s}g_b^\sigma][\Phi_2(b,s-1) - \Phi_2(b,s)]$$

$$-\sum_{b=2}^{\overline{b}}\sum_{s=\overline{s}}^{2}\Phi_{12}(b-1,s-1)g_{b-1}^s \tag{20}$$

*Substituting (20) back into inequality (18) enables one to write this latter inequality as follows (after performing an Abel decomposition of the first term of the left hand side of this inequality this time with respect to the outer term of Inequality (17):*

$$-\sum_{b=1}^{\overline{b}}\sum_{\beta=1}^{b}\sum_{s=\overline{s}}^{1}\Delta\mathbf{d}_{\beta s}\Phi_1(b,1)$$

$$+\sum_{b=1}^{\overline{b}}[\sum_{s=\overline{s}}^{2}\sum_{\sigma=\overline{s}}^{s}\Delta\mathbf{d}_{b\sigma} + \sum_{s=\overline{s}}^{2}g_{b-1}^s - \sum_{s=\overline{s}}^{2}g_b^s]\Phi_2(b,1)$$

$$+\sum_{b=1}^{\overline{b}}\sum_{s=\overline{s}}^{2}[\sum_{\sigma=\overline{s}}^{s}\sum_{j=\overline{s}}^{\sigma}\Delta\mathbf{d}_{bj} + \sum_{\sigma=\overline{s}}^{s}g_{b-1}^\sigma - \sum_{\sigma=\overline{s}}^{s}g_b^\sigma]\Phi_{22}(b,s-1)$$

$$-\sum_{b=2}^{\overline{b}}\sum_{s=\overline{s}}^{2}\Phi_{12}(b-1,s-1)g_{b-1}^s$$

$$\geq \quad 0 \tag{21}$$

*where $\Phi_{22}(b,s-1) = \Phi_2(b,s) - \Phi_2(b,s-1)$). We now observe that, for every $b \in \mathcal{B}$ and $s \in \mathcal{S}$, one has:*

$$\sum_{\sigma=\overline{s}}^{s}(s-\sigma)\mathbf{d}_{b\sigma} = s\sum_{\sigma=\overline{s}}^{s}\mathbf{d}_{b\sigma} - \sum_{\sigma=\overline{s}}^{s}\sigma\mathbf{d}_{b\sigma}$$

$$= s\sum_{\sigma=\overline{s}}^{s}\mathbf{d}_{b\sigma} - s\sum_{\sigma=\overline{s}}^{s}\mathbf{d}_{b\sigma} - \sum_{\sigma=\overline{s}}^{s+1}\sum_{j=\overline{s}}^{\sigma}\Delta\Delta\mathbf{d}_{bj}$$

$$= -\sum_{\sigma=\overline{s}}^{s+1}\sum_{j=\overline{s}}^{\sigma}\Delta\mathbf{d}_{bj} \tag{22}$$

*Using this expression, one can write Inequality (21) as:*

$$-\sum_{b=1}^{\overline{b}}\sum_{\beta=1}^{b}\sum_{s=\overline{s}}^{1}\Delta\mathbf{d}_{\beta s}\Phi_1(b,1)$$

$$+\sum_{b=1}^{\overline{b}}[\sum_{s=\overline{s}}^{1}(s-1)\Delta\mathbf{d}_{bs}+\sum_{s=\overline{s}}^{2}g_{b-1}^s-\sum_{s=\overline{s}}^{2}g_b^s]\Phi_2(b,1)$$

$$+\sum_{b=1}^{\overline{b}}\sum_{s=\overline{s}}^{2}[\sum_{\sigma=\overline{s}}^{s-1}(\sigma-s+1)\Delta\mathbf{d}_{b\sigma}+\sum_{\sigma=\overline{s}}^{s}g_{b-1}^\sigma-\sum_{\sigma=\overline{s}}^{s}g_b^\sigma]\Phi_{22}(b,s-1)$$

$$-\sum_{b=2}^{\overline{b}}\sum_{s=\overline{s}}^{2}\Phi_{12}(b-1,s-1)g_{b-1}^s$$

$$\geq\quad 0 \tag{23}$$

*As is clear, a sufficient condition for Inequality (23) to hold for all functions $\Phi$ satisfying properties (i)-(iv) of Lemma 3 (that are all such functions such that $\Phi_1 \leq 0$, $\Phi_{12} \leq 0$, $\Phi_2 \geq 0$ and $\Phi_{22} \geq 0$ is to have:*

$$\sum_{\beta=1}^{b}\sum_{s=\overline{s}}^{1}\Delta\mathbf{d}_{\beta s}\quad\geq\quad 0 \ and, \tag{24}$$

$$\sum_{\sigma=\overline{s}}^{s-1}(\sigma-s+1)\Delta\mathbf{d}_{b\sigma}+\sum_{\sigma=\overline{s}}^{s}g_{b-1}^\sigma-\sum_{\sigma=\overline{s}}^{s}g_b^\sigma\quad\geq\quad 0 \tag{25}$$

*for all $b \in \mathcal{B}$ and $s \in \mathcal{S}$ for some list of $(\overline{b}+1)\overline{s}$ non-negative numbers $g_{b-1}^s$ (for $b \in \mathcal{B}$ and $s \in \mathcal{S}$) satisfying $g_0^s = g_{\overline{b}}^s = 0$ for every $s \in \mathcal{S}$. Condition (24) for every $b$ is nothing else than the second requirement Statement (b) of Theorem 2. We now establish that a sufficient condition for the existence of a list of $(\overline{b}+1)\overline{s}$ non-negative numbers $g_{b-1}^s$ (for $b \in \mathcal{B}$ and $s \in \mathcal{S}$) satisfying $g_0^s = g_{\overline{b}}^s = 0$ for every $s \in \mathcal{S}$ for which (25) holds for every $(b,s) \in \mathcal{B} \times \mathcal{S} \in \{1,...,m\}$ is to have:*

$$\sum_{b\in\mathcal{B}}\sum_{\sigma\in\{s^b,...,\overline{s}\}}\max(\sigma-s^b)\Delta\mathbf{d}_{b\sigma}\geq 0 \tag{26}$$

*for every $(s^1,...,s^{\overline{b}}) \in \mathcal{S}^{\overline{b}}$ such that $s^1 \leq ... \leq s^{\overline{b}}$. To see this, suppose that Inequality (26) holds for all such $(s^1,...,s^{\overline{b}})$. For every $(b,s) \in \mathcal{B} \times \mathcal{S}$, define the numbers $\sum_{\sigma=\overline{s}}^{s}g_{b-1}^\sigma$ recursively as follows:*

$$\sum_{\sigma=\overline{s}}^{s}g_{b-1}^\sigma=\min_{s\geq s^b}\sum_{\sigma=\overline{s}}^{s}(\sigma-s)\Delta\mathbf{d}_{b\sigma}+\sum_{\sigma=\overline{s}}^{s}g_b^\sigma$$

*starting from:*

$$\sum_{\sigma=\overline{s}}^{s}g_{\overline{b}-1}^\sigma=\min_{s\geq s^b}\sum_{\sigma=\overline{s}}^{s}(\sigma-s)\Delta\mathbf{d}_{\overline{b}-1}$$

*and setting $\sum_{\sigma=l}^{s}g_0^\sigma = 0$ for every $s$. Hence $\sum_{\sigma=\overline{s}}^{s}g_{b-1}^\sigma$ is the minimal difference in success between the two education systems $\mathbf{e}$ and $\mathbf{e}^\circ$ for all backgrounds above $b$ and all success target above $s$. These numbers are clearly non-negative (since inequality (26) holds) and decreasing, as required. Moreover, the min condition in the definition of these numbers guarantees that condition (25) holds and this completes the proof. ∎*

34

Just like for Theorem 1, we prove Theorem 2 as a corollary of the just proved Proposition 3 and Proposition 4 below. The latter proposition states the equivalence between the fact of going from an education system to another by a finite sequence of improvements in cognitive skills, deteriorations in family backgrounds, reductions in correlation between cognitive skill and family background and increasing dispersions in skills on the one hand, and the fact of observing Inequality (16) for all functions $\Phi$ satisfying Properties (i)-(iv) of Lemma 3 on the other.

For this sake, we define the binary relation $\succsim_2$ on $\mathcal{E}$ as follows: $\mathbf{d^{e^*}} \succsim_2 \mathbf{d^{e^\circ}}$ if and only if $\mathbf{d^{e^*}}$ has been obtained from $\mathbf{d^{e^\circ}}$ by either (non-inclusively) an improvement in a child's cognitive skill, a deterioration in family background a reduction in correlation between skill and background or an increase in children skill dispersion as per Definitions 1-4. Just like its subrelation $\succsim_1$ defined in the proof of Theorem 1, $\succsim_2$ is anti-symmetric and reflexive but not transitive. Its transitive closure, denoted $\widehat{\succsim}_2$, and is defined :

$$\mathbf{d^{e^*}} \ \widehat{\succsim}_1 \ \mathbf{d^{e^\circ}} \iff \exists \text{ a sequence } \{\mathbf{d^{et}}\}_{t=0}^{\bar{t}} \text{ such that:}$$
$$\mathbf{d^{e0}} = \mathbf{d^{e^*}},$$
$$\mathbf{d^{e\bar{t}}} = \mathbf{d^{e^\circ}} \text{ and}$$
$$\mathbf{d^{et}} \succsim_2 \mathbf{d^{et+1}} \text{ for } t = 0, ..., \bar{t} - 1$$

As in the proof of Theorem 1, we can equivalently write $\mathbf{d^{e^*}} \succsim_2 \mathbf{d^{e^\circ}}$ by the fact that the difference $\mathbf{\Delta d} = \mathbf{d^{e^*}} - \mathbf{d^{e^\circ}}$ belongs to the set $\mathcal{B}(\mathcal{T}^2) = \mathcal{B}(\mathcal{T}^1) \cup \mathcal{T}_{spread}$ where the set $\mathcal{T}_{spread}$ is defined by:

$$\mathcal{T}_{spread} =$$

$$= \{\Delta\mathbf{d} \in \mathbb{Z}^{\overline{bs}} : \exists b \in \mathcal{B}, \ s, \ s' \in \mathcal{S} \text{ s.t. } s + 1 < s', \ \mathbf{d}_{\beta\sigma} = 0 \text{ for which:}$$
$$\Delta\mathbf{d}_{\beta\sigma} = 0 \ \forall \ (\beta, \sigma) \in (\mathcal{B} \times \mathcal{S}) \backslash (\{(b, s), (b, s+1)\} \cup \{(b, s'-1), (b, s')\}),$$
$$\Delta\mathbf{d}_{bs'} = 1, \Delta\mathbf{d}_{bs'-1} = -1, \ \Delta\mathbf{d}_{bs} = 1 \text{ and } \Delta\mathbf{d}_{bs+1} = -1 \text{ if } s' > s + 2 \text{ and,}$$
$$\Delta\mathbf{d}_{bs'} = 1, \Delta\mathbf{d}_{bs'-1} = \Delta\mathbf{d}_{bs+1} = -2, \ \Delta\mathbf{d}_{bs} = 1 \text{ if } s' = s + 2\}$$

We finally define the set $\mathcal{T}^2$ of all differences in education systems $\mathbf{d^{e^*}} - \mathbf{d^{e^\circ}}$ resulting from the fact of obtaining $\mathbf{d^{e^*}}$ from $\mathbf{d^{e^\circ}}$ by a finite sequence of improvements in a child's cognitive skill, deteriorations in family background and reductions in correlation between skill and background as per Definitions 1-4.as follows:

$$\mathcal{T}^2 = \{m \in \mathbb{Z}^{\overline{bs}} : m = \sum_{t=1}^{\#\mathcal{B}(\mathcal{T}^1)} \lambda_t m_t \text{ for } \lambda_t \in \mathbb{N} \text{ and } m_t \in \mathcal{B}(\mathcal{T}^2)\} \tag{27}$$

Just like for in the proof of Theorem 1, $\mathbf{d^{e^*}} - \mathbf{d^{e^\circ}} \in \mathcal{T}^2$ if and only if $\mathbf{d^{e^*}} \ \widehat{\succsim}_1 \ \mathbf{d^{e^\circ}}$. It is clear that $0 \in \mathcal{T}^2$, and that $\mathcal{B}(\mathcal{T}^2)$ is a finite basis of the set $\mathcal{T}^2$ in the (obvious sense) that any $m \in \mathcal{T}^2$ can be written as $\sum_{t=1}^{\#\mathcal{B}(\mathcal{T}^2)} \lambda_t m_t$ for $\lambda_t \in \mathbb{R}_+$, $m_t \in \mathcal{B}(\mathcal{T}^2)$ and $t = 1, ..., \#\mathcal{B}(\mathcal{T}^2)$. It can also be observed that one can never have $m = m' + m''$ for $m, m'$ and $m'' \in \mathcal{B}(\mathcal{T}^2)$ so that $\mathcal{B}(\mathcal{T}^2)$ is a Hilbert basis for the set $\mathcal{T}^2$ (see e.g. Cook, Fonlupt, and Schrijver (1986)). Using similar arguments as in Lemma 2 above, one can verify that, just like $\mathcal{T}^1$, $\mathcal{T}^2$ is a discrete and pointed convex cone.

These remarks enable us to prove the following proposition which, when combined with the equivalence established in Proposition 3, proves Theorem 2. The proof of this proposition follows exactly the same structure as that of Proposition 2 and is therefore omitted.

**Proposition 4** .*Let* $\mathbf{d^{e^*}}$ *and* $\mathbf{d^{e^\circ}}$ *be two education systems in* $\mathcal{E}$. *Then, the two following statements are equivalent:*

*(a)* $\displaystyle\sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^*}\Phi(\beta,\sigma) \geq \sum_{\beta=1}^{\overline{b}}\sum_{\sigma=1}^{\overline{z}}\mathbf{d}_{\beta\sigma}^{\mathbf{e}^\circ}\Phi(\beta,\sigma)$ *for all functions* $\Phi: \mathcal{S}\times\mathcal{B} \longrightarrow \mathbb{R}$ *satisfying Properties (i)-(iii) of Lemma 1 and*

*(b)* $\mathbf{d^{e^*}} - \mathbf{d^{e^\circ}} \in \mathcal{T}^2$.

## 5.3 Statistical Inference

Comparing distributions on the basis of the criteria identified in statements (b) of Theorems 1 and 2 amounts to verifying that a certain number of inequalities are connecting two distributions. Specifically comparing education systems $e^*$ and $e^o$ by the criterion of Part (b) of Theorem 1 consists in verifying that the inequality

$$\sum_{\beta=1}^{b}\sum_{\sigma=s}^{\overline{s}}(\mathbf{d}_{\beta\sigma}^{\mathbf{e}^*} - \mathbf{d}_{\beta\sigma}^{\mathbf{e}^\circ}) \geq 0$$

$$\overset{\Longleftrightarrow}{\gamma_{bs}^{1e^*} - \gamma_{bs}^{1\mathbf{e}^\circ} \quad \geq \quad 0}$$

holds for all pairs $(b,s)$ observed on the discrete grid $\{1,...\overline{b}\}\times\{1,...\overline{s}\}$ where, for $j = \mathbf{e}^*, \mathbf{e}^\circ$:

$$\gamma_{bs}^{1j} = \sum_{\beta=1}^{b}\sum_{\sigma=s}^{\overline{s}}\mathbf{d}_{\beta\sigma}^{j}$$

Analogously, comparing the very same education systems by the criterion of Part (b) of Theorem 2 amounts to verifying that the inequalities:

$$\sum_{\beta=1}^{b}\sum_{\sigma=1}^{\overline{s}}(\mathbf{d}_{\beta\sigma}^{\mathbf{e}^*} - \mathbf{d}_{\beta\sigma}^{\mathbf{e}^\circ}) \geq 0 \tag{28}$$

$$\overset{\Longleftrightarrow}{\gamma_{b}^{2e^*} - \gamma_{b}^{2\mathbf{e}^\circ} \quad \geq \quad 0}$$

and

$$\sum_{b\in\{1,...,\overline{b}\}}[S^{\mathbf{e}^*}(b,s_b) - S^{\mathbf{e}^\circ}(b,s_b)] \geq 0 \tag{29}$$

$$\overset{\Longleftrightarrow}{\gamma_{s}^{2e^*} - \gamma_{s}^{2\mathbf{e}^\circ} \quad \geq \quad 0}$$

hold for all $b$ in the finite grid $\{1,...,\overline{b}\}$ and all $(s_1,...,s_{\overline{b}}) \in \{1,...\overline{s}\}^{\overline{b}}$ such that $s_1 \leq s_2 \leq ... \leq s_{\overline{b}}$ where, for $j = \mathbf{e}^*, \mathbf{e}^\circ$:

$$\gamma_{b}^{2j} = \sum_{\beta=1}^{b}\sum_{\sigma=s}^{\overline{s}}\mathbf{d}_{\beta\sigma}^{j}$$

36

and:

$$\gamma_s^{2j} = \sum_{b=1}^{\overline{b}} S^j(b, s_b)$$

We base the test for the statistical significance of these collections of inequalities on the statistics $T_{bs}^1$, $T_{1b}^2$ and $T_{2bs(b)}^2$ defined by:

$$T_{bs}^1 = \frac{\widehat{\gamma}_{bs}^{1e^*} - \widehat{\gamma}_{bs}^{1e^\circ}}{\left(\frac{\widehat{\omega}_{bs}^{1e^*}}{N^{e^*}} + \frac{\widehat{\omega}_{bs}^{1e^\circ}}{N^{e^\circ}}\right)^{\frac{1}{2}}},$$

$$T_b^2 = \frac{\widehat{\gamma}_{1b}^{2e^*} - \widehat{\gamma}_{1b}^{2e^\circ}}{\left(\frac{\widehat{\omega}_b^{2e^*}}{N^{e^*}} + \frac{\widehat{\omega}_b^{2e^\circ}}{N^{e^\circ}}\right)^{\frac{1}{2}}}$$

and,

$$T_s^2 = \frac{\widehat{\gamma}_s^{2e^*} - \widehat{\gamma}_s^{2e^\circ}}{\left(\frac{\widehat{\omega}_s^{2e^*}}{N^{e^*}} + \frac{\widehat{\omega}_s^{2e^\circ}}{N^{e^\circ}}\right)^{\frac{1}{2}}}$$

where, for $j = e^*, e^\circ$:

(i) $N^j$ is the size of the sample drawn from population $j$,

(ii) $\widehat{\gamma}_{bs}^{1j}$, $\widehat{\gamma}_b^{2j}$ and $\widehat{\gamma}_s^{2j}$ are the sample estimates of $\gamma_{bs}^{1j}$, $\gamma_{bs}^{2j}$ and $\gamma_s^{2j}$ respectively and,

(iii) $\widehat{\omega}_{bs}^{1j}$, $\widehat{\omega}_b^{2j}$ and $\widehat{\omega}_s^{2j}$ are the estimates of the variance of $\widehat{\gamma}_{bs}^{1j}$, $\widehat{\gamma}_b^{2j}$ and $\widehat{\gamma}_s^{2j}$ respectively.

Since the PISA samples of the different countries can be considered independent, one can use Davidson and Duclos (2000) (p. 1445) argument based on the law of large numbers and the Central limit theorem to provide the following estimates of those variances:

$$\widehat{\omega}_{bs}^{1j} = \frac{1}{N^j} \sum_{\beta=1}^{b} \sum_{\sigma=s}^{\overline{s}} \mathbf{d}_{\beta\sigma}^j - (\widehat{\gamma}_{bs}^{1j})^2,$$

$$\widehat{\omega}_b^{2j} = \frac{1}{N^j} \sum_{\beta=1}^{b} \sum_{\sigma=1}^{\overline{s}} \mathbf{d}_{\beta\sigma}^j - (\widehat{\gamma}_b^{2j})^2 \text{ and,}$$

$$\widehat{\omega}_s^{2j} = \frac{1}{N^j} \sum_{b=1}^{b} \sum_{\sigma=s(b)}^{\overline{s}} \mathbf{d}_{b\sigma}^{\mathbf{e}}(\sigma - s(b))^2 - (\widehat{\gamma}_s^{2j})^2$$

We test dominance based on the liberal Union-Intersection criterion (UI) of Bishop, Formby, and Thistle (1989). This amounts to conclude that education system $\mathbf{e}^*$ dominates system $\mathbf{e}^\circ$ for the dominance criterion of Theorem 1 if and only if:

$$\min_{(b,s)\in\{1,...,\overline{b}\}\times\{1,...,\overline{s}\}} T_{bs}^1 > -C_\alpha^1 \text{ and } \max_{(b,s)\in\{1,...,\overline{b}\}\times\{1,...,\overline{s}\}} T_{bs}^1 > C_\alpha^1$$

where $C_\alpha^1$ is the critical value for a significance level of $\alpha$ ($\alpha$ being the probability of rejecting $H_0$ when $H_0$ is true) derived from the Studentized Maximum Modulus (SMM) distribution provided by Stoline and Ury (1979) with the appropriate degrees of freedom. Similarly, one concludes that system $\mathbf{e}^*$ dominates system $\mathbf{e}^\circ$ for the dominance criterion of Theorem 2 if and only if:

$$\min_{b\in\{1,...,\overline{b}\}} T_b^2 > -C_\alpha^{2b} \text{ and } \max_{b\in\{1,...,\overline{b}\}} T_{1b}^2 > C_\alpha^{2b}$$

and:

$$\min_{(s_1,...,s_{\overline{b}})\in\{1,...,\overline{s}\}^{\overline{b}}:s_1\leq s_2\leq...\leq s_{\overline{b}}} T_s^2 > -C_\alpha^{2s} \text{ and } \max_{(s_1,...,s_{\overline{b}})\in\{1,...,\overline{s}\}^{\overline{b}}:s_1\leq s_2\leq...\leq s_{\overline{b}}} T_s^2 > C_\alpha^{2s}$$

where again $C_\alpha^{2b}$ and $C_\alpha^{2s}$ are the critical values for a significance level of $\alpha$ taken from the SMM distribution with the number of degree of freedom appropriate for the number of inequalities associated to (28) and (29) respectively. All dominance test performed in this paper have been done at the 95% confidence interval.

# References

ALLISON, R. A., AND J. E. FOSTER (2004): "Measuring health Inequality Using Qualitative Data," *Journal of Health Economics*, 23, 505–524.

ATKINSON, A. B. (1981a): "The Measurement of Economic Mobility," in *Inkomensverdeling en Openbare Financien*, ed. by P. Eggelshoven, and L. van Gemerden. Het Spectrum.

———— (1981b): "On Intergenerationnal Income Mobility in Britain," *Journal of Post-Keynesian Economics*, 3, 194–218.

ATKINSON, A. B., AND F. BOURGUIGNON (1982): "The Comparison of Multi-dimensioned Distribution of Economic Status," *Review of Economic Studies*, 49, 183–201.

———— (1987): "Income Distributions and Differences in Needs," in *Arrow and the Foundation of the Theory of Economic Policy*, ed. by G. R. Feiwel. Macmillan, London.

BARRETT, G. J. (2012): "The Returns to Cognitive Skills in the Australian Labour Market," *The Economic Record*, 88, 1–17.

BAZEN, S., AND P. MOYES (2012): "Elitism and Stochastic Dominance," *Social Choice and Welfare*, 39, 207–251.

BÉNABOU, R., F. KRAMARZ, AND C. PROST (2009): "The French Zones D'Education Prioritaire: Much Ado About Nothing ?," *Economics of Education Review*, 28, 345–356.

BISHOP, C. M., AND J. P. FORMBY (1999): "Test of Significance for Lorenz Partial Orders," in *Handbook of Inequality Measurement*, ed. by J. Silber. Kluwer Academic Press.

BISHOP, C. M., J. P. FORMBY, AND P. D. THISTLE (1989): "Statistical Inference, Income Distributions and Social Welfare," in *Research on Economic Inequality*, ed. by D. J. Slotje. JAI Press.

BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2005): "Why the Apple doesn't fall far: understanding intergenerational transmission of human capital," *American Economic Review*, 95, 437–439.

BOURGUIGNON, F. (1989): "Family Size and Social Utility: Income Distribution Dominance Criteria," *Journal of Econometrics*, 42, 67–80.

CARVALHO, M. D., L. F. GAMBOA, AND F. D. WALTENBERG (2012): "Equality of Educational Opportunity Employing Pisa Data: Taking Both Achievement and Access into Account," ECINEQ working paper no. 2012-277.

COOK, W., J. FONLUPT, AND A. SCHRIJVER (1986): "An Integer Analogue of Caratheodory's Theorem," *Journal of Combinatorial Theory, Serie B*, 40, 63–70.

DAHL, G. B., AND L. LOCHNER (2012): "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax," *American Economic Review*, 102, 1927–1956.

DAVIDSON, R., AND J. Y. DUCLOS (2000): "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, 58, p.1435–1465.

DUCLOS, P. Y., D. SAHN, AND S. D. YOUNGER (2006): "Robust Multidimensional Poverty Comparisons," *The Economic Journal*, 116, 943–968.

FERREIRA, F. H. G., AND J. GIGNOUX (2014): "The Measurement of Educational Inequality: Achievement and Opportunity," *The World Bank Economic Review*, 28, 210–246.

FIELDS, J. S., AND E. A. OKE (1999): "The measurement of income mobility: an introduction to the literature," in *Handbook on Income Inequality Measurement*, ed. by J. Silber, pp. 557–596. Kluwer, Norwell, MA.

FISHBURN, P. C., AND R. G. VICKSON (1978): "Theoretical Foundations of Stochastic Dominance," in *Stochastic Dominance*, ed. by G. A. Withmore, and M. C. Findlay. Lexington Books.

GANZEBOOM, H. B. G., P. D. GRAAF, AND D. J. TREIMAN (1992): "A standard international socio-economic index of occupational status," *Social Science Research*, 21, 1–56.

GOUSSÉ, M., AND N. LEDONNÉ (2015): "Why do inequality in 15-year-old cognitive skill increases so much in France between 2000 and 2009 ?," Mimeo, Université Laval.

GRAVEL, N., B. MAGDALOU, AND P. MOYES (2015): "Ranking Distributions of an Ordinal Attribute," AMSE working paper, no 2015-50.

GRAVEL, N., AND P. MOYES (2012): "Ethically Robust Comparisons of bi-dimensionnal Distributions with an Ordinal Attribute," *Journal of Economic Theory*, 147, 1384–1426.

GRAVEL, N., P. MOYES, AND B. TARROUX (2009): "Robust International Comparisons of Distributions of Disposable Income and Access to Regional Public Goods," *Economica*, 76, 432–461.

GRAVEL, N., AND A. MUKHOPADHYAY (2010): "Is India better off now than fifteen years ago ? A robust multidimensional answer," *Journal of Economic Inequality*, 8, 173–195.

GREEN, D. A., AND W. C. RIDDELL (2003): "Literacy and Earnings: An Investigation of the Interaction of Cognitive and Unobserved Skills in Earnings Generation," *Labour Economics*, 10, 165–184.

HANUSHEK, E. A., G. SCHWERDT, AND L. WOESSMANN (2015): "Returns to Skills around the World: Evidence from PIAAC," *European Economic Review*, 73, 103–130.

HANUSHEK, E. A., AND L. WOESSMANN (2008): "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 46, 607–668.

——— (2011): "The Economics of International Differences in Educational Achievement," in *Handbook of the Economics of Education, Vol. 3*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann. North Holland, Amsterdam.

HECKMAN, J. J., J. J. STIXRUD, AND S. UZRUA (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482.

HUSSAIN, M. A., M. M. JORGENSEN, AND L. P. OSTERDAL (2016): "Refining Population Health Comparisons: A Multidimensional First Order Dominance Approach," *Social Indicators Research*, forthcoming.

JACOB, B., AND J. ROTHSTEIN (2016): "The Measurement of Student Ability in Modern Assessment Systems," *Journal of Economic Perspectives*, 30, 85–108.

JENKINS, S. P., AND P. J. LAMBERT (1993): "Ranking Income Distributions when Needs Differ," *Review of Income and Wealth*, 39, 337–356.

LEHMANN, E. L. (1955): "Ordered Family of Distributions," *The Annals of Mathematics and Statistics*, 26, 399–419.

MAGDALOU, B. (2018): "An Model of Social Welfare Improving Transfers," Center for Environmental Economics – Montpellier, Working paper no. 2018-13.

MAYER, S. A. (1997): *What Money Can't Buy: Family Income and Children's Life Chances.* Harvard University Press, Cambridge, MA.

MOYES, P. (2012): "Comparaisons of Heterogeneous Distributions and Dominance Criteria," *Journal of Economic Theory*, 147, 1351–1383.

MULLER, A., AND M. SCARSINI (2012): "Fear of Loss, Inframodularity and Transfers," *Journal of Economic Theory*, 147, 1490–1500.

MULLER, C., AND A. TRANNOY (2012): "Multidimensionnal Inequality Comparisons: A Compensation Perspective," *Journal of Economic Theory*, 147, 1427–1449.

NICKELL, S. (2004): "Poverty and Worklessness in Britain," *Economic Journal*, 114, C1–C25.

OREOPOULOS, P., AND K. G. SALVANES (2011): "Priceless: the Non-Pecuniary Benefit of Education," *Journal of Economic Perspectives*, 25, 159–184.

PISA (2017): "PISA 2015 Technical Report," Paris, OECD, 463 pages.

SCHUTZ, G., H. W. URSPRUNG, AND L. WOESSMANN (2008): "Education Policy and Equality of Opportunity," *Kiklos*, 61, 279–308.

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

STOLINE, M. R., AND H. K. URY (1979): "Tables of the Studentized Maximum Modulus Distributions and an Application to Multiple Comparisons among Means," *Technometrics*, 21, 87–93.