Evaluations of compensatory educational programs: which lessons have to be learned?

Working Paper

Fanny ALIVON[1]
Laboratoire d'Économie de Dijon (LEDi)
UMR CNRS 6307
Université de Bourgogne

Rachel GUILLAIN
Laboratoire d'Économie de Dijon (LEDi)
UMR CNRS 6307
Université de Bourgogne

Abstract: This paper aims to realize a critical overview of empirical studies on evaluation of compensatory educational programs from urban segregation point of view. This programs are constructed for struggle with effects of segregation, namely fill the real gaps between students from disadvantaged background and those more advantaged. The efficiency of those programs is regularly criticized, particularly because of resources deployed. We wonder about the lessons given by the evaluations studies of those programs. Conclusions are not evident: consequences of those programs are not linked at a specific application, but those policies show differentiated, conditional effects to school characteristics (such as the size or years of experiences of the establishment in the program) and to student characteristics (such as their age or initial abilities). Through this overview, we establish a set of methodological suggestions and recommendations for improving the efficiency of those programs.

---

[1] Corresponding author: fanny.alivon@u-bourgogne.fr ; Address : Laboratoire d'Économie de Dijon (UMR 6307 CNRS-Université de Bourgogne) – Pôle d'Économie et de Gestion – 2 boulevard Gabriel – BP26611 – 21066 DIJON Cedex, FRANCE ; Tel : +33 672 145 069

1. Introduction

Inequalities between populations tend to increase, between countries, within countries, but also inside each area of cities. This translates into the setting up or the amplification of ghettos in some disconnected districts, with the concentration of poor people who accumulate socioeconomics difficulties. Thus, we can see the emergence of phenomena such as urban riots like in 2007 in French big cities.

Cities are not homogenous territories: inequalities are identified within different neighborhoods of a same town. Some gather wealth and are free from socioeconomic difficulties. Others gather difficulties like unemployment, insecurity, exclusion and crime in a cumulative way. Those factors reflect the terms of urban segregation: it is the spatial mark of socioeconomic disparities. Urban segregation is defined by Baumont & Guillain (2013) like "a joint process of geographical clustering of communities whose share a set of socioeconomic characteristics and whose are relatively exclusive from other groups". This is a multidimensional and interdependent phenomenon, studying through three points of view (employment, education and crime) by simplification.

When we talk about employment we principally refer to the hypothesis of "spatial mismatch" express by Kain (1968) which is that the fact of reside in places far and misconnected from job centers could have really essential consequences in terms of level of wage and unemployment. This link between residential location and probability of unemployment can be explained by several factors: (i) people who live in disconnected areas have less information about jobs opportunities, because of the distance but also because of the weakness of their network, (ii) incentives for intensive jobs research for people living in a disconnect area could be weak because of house prices which are more expensive and the acceptation of a job could be synonym of moving, (iii) people could refuse a distant job because it implies long and expensive costs of commuting compared with the proposed wage, (iv) employers can practice strategies of redlining, based on prejudice about people living in those type of living area, (v) employers can considerate that people living far from jobs are less productive considering long commuting and are less ready to accept flexible schedules.

Segregation can also be linked to crime. Economic models consider that the decision to commit a crime is rational, so depend on the level of repression, but is also influence by social consideration such as local environment and peer pressure. In a spatial point of view, the analysis of crime show that individual incitation to commit crime are affected by the probability to be arrested, but also by the place of resident of criminals.

If we concentrate on the link between segregation and education is based on the concept of human capital externalities. We observe spillovers or peer effects at school. A pupil has more chance to acquire a good level of education if peers from the same school have themselves a good education. The student achievement strongly depends on socioeconomic characteristics of their classmates because each student is source of a spillover effect on the others. Social interactions among individuals can also have important effects on educational achievement. The group norm, the social pressure has an influence on educational decisions.

In each case, location choices underlie the phenomenon. Segregation is initially the resultant of mechanism of housing market, and moreover, locations tend to fuel the most flagrant inequalities instead of correct them. So, segregation is not socially desirable.

To fight against it, and thus to make cities more integrative space, to introduce social diversity in each neighborhood, public policies are implemented. But it is not so easy, it requires major operations because it implies to overcoming the spontaneous market forces. Furthermore, like we said before segregation is a multidimensional phenomenon which concerns several facets of urban society. Thus, public policies are not global; they seek to act on a particular mechanism. And so it is the same for their evaluations.

The evaluation of public policies began in the 1950s in the USA, developed in the 1980s around the world and became inescapable nowadays, at the point of even be mandatory in some countries. The

principle is to assess the effectiveness of a policy by comparing its results to the objective and resources implemented.

This paper falls within the same tendency of assessments of evaluations than Mayneris & Py (2013) or also Briant & Lafourcade (2014) on the French Enterprise Zone. We not focus like them on one particular policy but on a large set of programs on different countries.

We talk about public policies in plural because it is not a uniform policy. All had the same ambition but a plurality of application, so we wonder if one is more efficient than the others.

There is a general feeling that those policies, which are implemented since more than thirty years, would be moderately efficient compared to the resources deployed. Besides, we haven't the possibility to see clearly the effects. These latters would be, according to the collective belief, modest compared to the efforts. In this context, what say the economic studies? Is it real or is it a bad relay media?

The question is: is there a way to conduct these programs which works better? It underlies the questioning on the appropriate methods for evaluate these policies.

This overview is necessary because the responses are more nuanced than a categorical one. Other factors are involved in the effectiveness of the policies. There is a duration effect, i.e. it's take time before policies act and before we can assess them. It is necessary to have a step back to perceive the transcription of the effects.

2. Why studying policies in education for analyzing process of segregation?

The link between education and segregation is dual. First, the role of public policies is to fight against effects of segregation. Otherwise, its goal is to limit future segregation. These two points are closely linked.

Policies who struggle against segregation in the field of education are based on the existence of peer effects. The implicit mechanism is the follow: living in a particular neighborhood influence positively or negatively life course of residents. Public policies aim to counter the negative neighborhood effects and to give same chances of success to everybody, the effect of the territory have to be then neutralized.

We observe that children, who are in situation of learning difficulties, of school dropout, or the simple fact to not have a full schooling that would allow them to obtain a degree or qualifications, will probably have issues of insertion on the labor market, hence a strengthening of the current segregation. Therefore, to intervene on pupils and students is equivalent to try to counter the segregation process at its base, i.e. try to improve the student's path for ensuring them a better occupational integration and then reduce segregation.

3. Context of the survey

Commonly, when we talk about educational program, we think at policies which aim to standardize and provide the same knowledge to all. However, observing a widening gap and an accentuation of segregation, the authorities implement also public policies of positive discrimination to give more to those who have less.

So, we interest on compensatory educational programs from the prism of econometric studies on evaluation of educational programs.

There is a general feeling that many studies have been done on the subject. But not from an economic point of view. Evaluations of a whole educational program are not so common. Most of the time, evaluations are about particular measures. As an example, we can find a lot of studies on the effects of class size on pupil's performances. But the reduction of size of a class is just one of possible ways to reach the ambition of a specific program.

We only interest on programs not on each measure which composed it. This choice is mad for the reason that measures can differ. They are not necessarily same everywhere or not even applied every time. Generally measures are more recommendation than real obligation of application. So we prefer to analyze programs as a whole. Moreover, educational programs have approximately the same

ambition, namely to fill the performance gap between the most disadvantaged/in difficulties students and their more advantaged peers.

The question behind this review is: we hear a lot of critics against educational programs, from a common point of view they would not be effective, but, what does the economics studies said? Have these programs borne fruits? If this is the case, are the impacts the same everywhere for everybody?

4. Educational programs
4.1. The individual or the territory?

The type of educational programs where are studied is based on the idea of positive discrimination, i.e. "give more to those who have the least".
All began with the Title I of the law component the educational section of the "poverty war" started by president Johnson in 1965. These policies have rapidly spread on the Anglo-Saxon world.
Educational policies are all started with the same observation: there are differences between students with more difficulties and/or most disadvantaged and their pairs more advantaged. So compensatory policies aim to fill these gaps, but also to help those which are more in difficulties, from the perspective of, at term, improving life course and reducing segregation.
Behind this shared goal, we can see a plurality of implementation. We find, like for other policies, the opposition highlighted by Glaeser & Gottlieb (2008) between place-based and agent-based policies. Individualized policy focuses on the person regardless of its location in the territory. On the contrary, territorialized policy provides financial resources to areas which concentrate socioeconomic difficulties.
Some policies fall under individualized logic while others fall under territorialized logic.
But, as we will see later, this distinction can be exceeded. Indeed, unlike other public policies, educational ones produce, indifferently of their implementation, heterogeneous effects.

4.2. Description of programs

Program Name	Location	Date of creation	Program Description
Title I	United-States	1965: Elementary and Secondary Education Act	Funding is distributed to schools and school districts with a high percentage of students from low-income families.
Individuals with Disabilities Education Act	United-States	1975: IDEA	This act prescribed a series of diagnostics, counseling activities, and services for students with disabilities (both physical and mental).
Special Educational Needs	United Kingdom – England	1981: Education Act (England)	Schools identify children with learning difficulties. Gradual approach, individual treatment.
Éducation Prioritaire	France	1982: Zones d'Éducation Prioritaire	Government provides supplementary resources to establishments located in disadvantaged areas.
Chicago's Social Promotion Policy	United-States (Chicago)	1996: CSP	It imposed grade retention for students who don't hit a certain score on standardized tests. Retention occurs after failure at second chance exams taken after summer schools.
Excellence in Cities	England	1999: White Paper Excellence in Schools	Urban schools, specifically in disadvantaged areas, received resources to raise standards.
Bagrut 2001	Israel	2001: Bagrut	Additional instructions are provided to underperforming high school students for preparing them to the matriculation exams.

Table 1: Description of programs

Title I, the historical program in USA
Since 1965 and the Elementary and Secondary Education Act, Title I's main goal has been to "help close the educational achievement gap between economically disadvantaged students and their more advantaged peers by providing funding for supplementary educational services in reading and mathematics to low-achieving students in low-income elementary and secondary schools". This program falls within the "War on Poverty" and was based on the argument that students from low-income families who live in areas with a high concentration of other poor families were twice disadvantaged.

Title I provide financial assistance to local educational agencies (LEAs) and schools with high numbers or high percentages of children from low-income families to help ensure that all children meet challenging state academic standards. Federal funds are currently allocated through four statutory formulas that are based primarily on census poverty estimates and the cost of education in each state : (i) Basic Grants provide funds to LEAs in which the number of children counted in the formula is at least 10 and exceeds 2 percent of an LEA's school-age population; (ii) Concentration Grants flow to LEAs where the number of formula children exceeds 6,500 or 15 percent of the total school-age population; (iii) Targeted Grants are based on the same data used for Basic and Concentration Grants except that the data are weighted so that LEAs with higher numbers or higher percentages of children from low-income families receive more funds; (iv) Targeted Grants flow to LEAs where the number of schoolchildren counted in the formula (without application of the formula weights) is at least 10 and at least 5 percent of the LEA's school-age population.

Education Finance Incentive Grants (EFIG) distribute funds to states based on factors that measure: (i) a state's effort to provide financial support for education compared to its relative wealth as measured by its per capita income; and (ii) the degree to which education expenditures among LEAs within the state are equalized.

Special education in United-States
The Individuals with Disabilities Education Act (IDEA) translated concerns about the education of children with both physical and mental disabilities into federal law. It prescribed a series of diagnostics, counseling activities, and services for disabled students.

Special Educational Needs Programs in England
In 2013, in England, about 20% of children have special needs or disabilities, like behavioral/social difficulties, troubles of reading and/or writing, of understanding or concentrating that affect their ability. To address this problem, since 1981 Education Act, the government is implementing the *Special Educational Needs* program (hereafter SEN). It is a highly decentralized policy, at the individual level, whose purpose is to help each child with specific difficulties in learning. Generally, it consists of asking schools to identify pupils with learning difficulties and then adapt teaching strategy according with a national Code of Practice. Interventions are decided at school level and include for example one-to-one tuition, teaching assistance etc. The SEN Code of Practice recommends a graduated approach. First stages are at the discretion of the school (both in identification and type of provision), later stages involve the Local Authority. Three types of actions can be distinguished:

- "School Action", when the school identifies a child as having SEN and sets about providing an intervention that is additional to or different from that which is provide to the rest of the pupils; Schools that identify a child as requiring school action will themselves determine the level of additional support necessary for the pupil and provide it from within existing resources.
- "School Action Plus", if School Action is deemed inadequate, and the pupil needs help from outside the school to fully address its needs. A specialist may act in an advisory capacity or provide additional specialist assessment or be involved in teaching the child directly.
- "Statement", the school may request a statutory assessment if all schools action fails.

The French policy of priority education
The French government had observed that children are facing inequalities at school owing to great diversity of social and cultural backgrounds. In consequence, since 1982 and the creation of the Priority Education Zones (*Zones d'Education Prioritaire* en French, and herafter ZEP), it has

developed a priority education policy for helping pupils leaving in most disadvantaged neighborhoods. This policy concern both primary (1st to 5th grades) and junior-high schools (6th to 9th grades), plus a small number of high schools (10th to 12th grades).

Years	Name program	Primary schools	Junior High Schools
1982-1990	*Zones Prioritaires*	3 730	503
1990-1999	*Zones d'Education Prioritaire* (ZEP)	5 503	796
1999-2008	*Zones d'Education Prioritaire* (ZEP) and *Réseaux d'Education Prioritaire* (REP)	7329	1 053
2008-2012	*Réseaux Ambition Réussite* (RAR) and *Réseaux de Réussite Scolaire* (RRS)	6969	1105
2012-today	*Ecoles collèges lycées pour l'ambition, l'innovation et la réussite* (ECLAIR) and *Réseaux de Réussite Scolaire* (RRR)	6 770	1 099

Table 2: Priority Education Policy – Wages of expansion
Source: French Ministry of National Education

The aim of the actual legislation (loi de Refondation de l'Ecole de la République du 8 juillet 2013) is to bring at less than 10% the achievement gap between students of the priority education and the rest of the students in France.

In operating terms, the government gives more resources to establishments located in disadvantaged neighborhood. This inequality of means aims for offsetting the effects of socioeconomics difficulties, for obtain an equality of outcomes in particularly disadvantaged areas. These resources may take the form of supplementary financial support but mostly supplementary teaching hours and compensatory credits, and they are attributed depending on essentially social criteria.

The Chicago's Social Promotion Policy
In 1996-1997, Chicago launched a policy to end social promotion, i.e. the practice of passing students to the next grade regardless of their academic skills or school performance. Then, students in third, sixth and eighth grades are required to perform at predefined levels in both reading and mathematics in order to be promoted to the next grade. Students who do not meet the standard in June are required to attend a six-week summer school program, after which they can retake the exams. Those who pass the August exams move on to the next grade. Students who fail againl are required to repeat the grade.
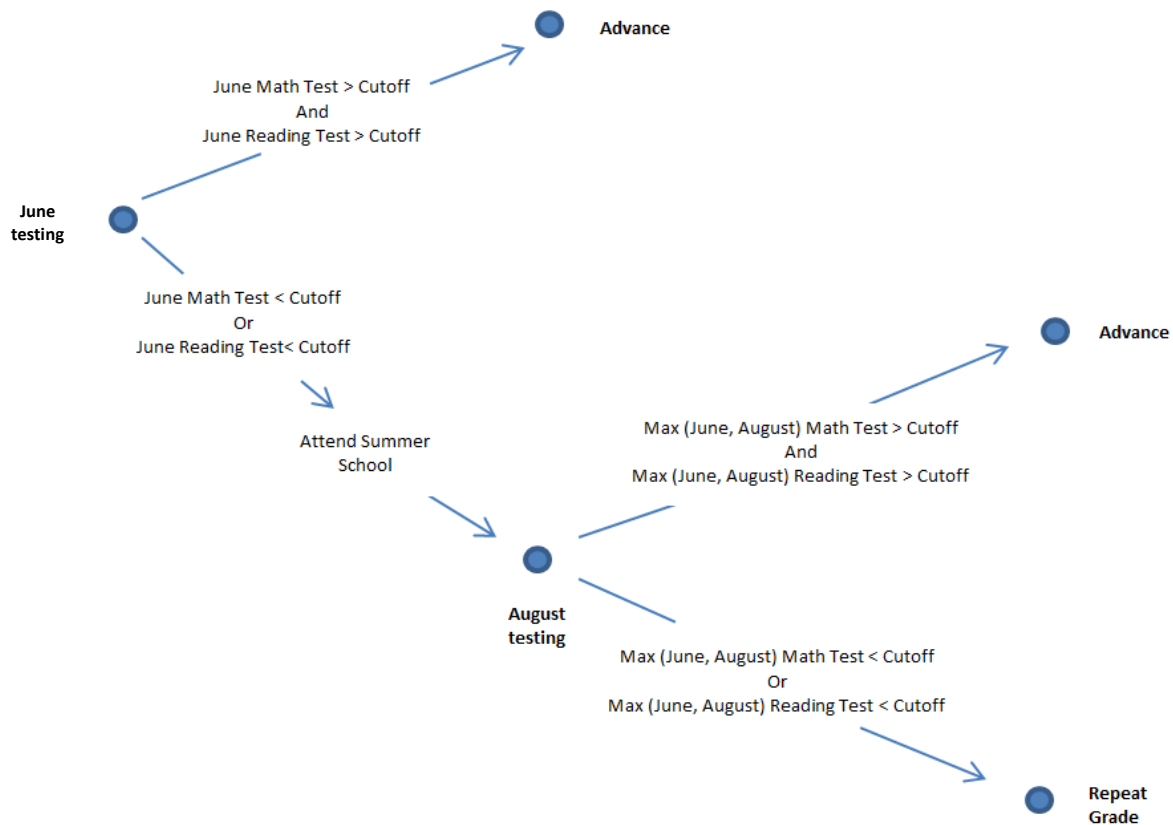
Figure 1: Student Progress under the Chicago Accountability Policy
Source: Jacob and Lefgren (2004)

Excellence in Cities: a British program

Excellence in Cities (EiC) is a major government policy designed to raise standards in urban schools. The launch of EiC, in 1999, was one of the outcomes of the 1997 White Paper Excellence in Schools which indicated an intention to create "inclusive schooling which provide a broad, flexible and motivating education that recognizes the different talents of all children and delivers excellence for everyone". One of the challenges facing such an inclusive system has been ongoing difficulties presented by problems of socioeconomic disadvantaged in major urban areas of England.

The overall vision of EiC was "to drive up standards in our schools in the major cities higher and faster; to match the standards of excellence found in our best schools. The output must be that city parents and city children expect and gain as much from their schools as their counterparts anywhere else in the country. A vision of what city education can become is what EiC is all about. Excellence must be the norm" (DfEE, 1999).

It aims to offer diversity of provision so that the needs of all pupils are met within a framework of cooperation and partnership between schools. EiC is organized through partnerships, and each partnership includes a local education authority (LEA) and all its secondary schools. Funding is allocated to each partnership, which is responsible for deciding how the resources should be used.

There are three core strands of EiC that affect all schools in treatment areas : (i) Learning Mentors, to help students to overcome educational or behavioral problems, (ii) Learning Support Units, to provide short-term teaching and support programs for difficult pupils; and a Gifted and Talented program, to provide extra support for 5-10% of pupils in each school. Some schools are designated as more Specialist (i.e., in particular subjects) or Beacon (to disseminate good practice), and received substantial sums of money. Other components of EiC are City Learning Centers (to provide ICT facilities) and Education Action Zones.

An Israeli program: 2001 Bagrut Program
In 2001, the Israel Ministry of Education has singled out as its top priority the need to raise the matriculation rate of the Bagrut especially among disadvantaged students and students in peripheral communities. The Bagrut is a diploma, completing by passing a series of national exams in core and elective subjects beginning in tenth grade, with some tests taken in eleventh and most taken in twelfth grade.
The Bagrut 2001 program targets low-achieving high schools. The intervention included individualized instruction in small study groups of up to five students for tenth, eleventh and twelfth graders. The aims of the intervention where (i) to design individualized instruction based on students' needs; (ii) to increase the matriculation rate; (iii) to enhance the scholastic and cognitive abilities, self-image, and leadership aptitudes of underperforming students. Participants were chosen by their teachers based on the likelihood of their passing the matriculation exams.

5. Which results from the empirical evaluations of these programs?
5.1. The interest of an evaluative approach

A new tool for the decision support, the evaluation of public policies, was created in the sixties in USA before to develop in United-Kingdom, in Scandinavian countries and, twenty years later, in all the western countries. This tool is strongly related to the rationalization of public action, and aims to confront a program to its results and to its initial objectives.
In France, the decree n°90-82 of the 22th January 1990 specify that "the evaluation of a policy is to examine whether the legal, administrative or financial means implemented had produce the expected effects of the policy and the objectives set for it".
Since the nineties the European Commission has mad of evaluation a regulatory systematic requirement in context of funding allocated to the Member-States.
The aim of evaluation is to determine the extent to which policy has achieved the objectives assigned and has produced the expected impacts on the public concerned.

A good evaluation makes possible to sort the accumulation of policies through time. Accumulation which surely conceal obsolete policies (objectives have already been achieved), inefficient policies (objectives are badly affected and/or the program is too expensive) or distorted policies (used for, *de facto*, other purposes than those displayed).

The evaluation is a technically complex exercise. This is to determine whether any improvements observed can be attributed to the implementation of the program.

5.2. The econometric evaluations of the compensatory educational programs
5.2.1. Global effects

As we said in section 3, we focus our interest on econometric evaluations of compensatory educational programs realized in the last fifteen years. Contrary to what we might think, it does not exist so much. We have identified twelve studies on several programs implemented in different countries.

Four recent evaluations study the impacts of the French compensatory education at different time of implementation. Benabou, Kramartz and Prost (2009) had studying the impact of the French *zones d'éducation prioritaire* (ZEP), the first phase of the French program, in a first time, on resources, theirs utilization and key establishment characteristics, and in a second time, on four measures of individual student achievement. They had found that ZEP program, which was in place until 2008 have had effects very slight on school characteristics, with an extremely slow decrease in class sizes (-0.2 students per year on average) and increase in teaching hours per student (+1.2% per year). These theoretically beneficial effects do not translate into an improvement of student achievement, as Benabou *et al.* find no effect of the ZEP program on students' attainment. The ZEP status was even a negative signaling effects for teachers and probably even for students (they found a decrease in the number of students entering in 6th grade).

Next generations of French compensatory education policy were analyzed by Beffy and Davezies (2013), Caille, Daviezes and Garrouste (2013) and Davezies and Garrouste (2014). They had all analyzed the impact of the RAR[2] program and had found at best no effect and at worst a negative impact. In this way, Caille *et al*. have studying several dimension of educational achievement and have found no effects on school performance at short and long term, nor on school trajectories. While, Beffy and Davezies had observed a decrease in pupil achievement, and had concluded that "the program increases the sorting on pupil's ability between schools and/or have a negative effect on the pupil's achievement in the schools". Davezies and Garrouste (2014), do not study the impact on achievement, but provide evidence of the negative signal produced by the implantation of the RAR program. They found that "living near a RAR junior high-school decreases the probability to attend the closest school and increases that probability to go to a private school locally". They highlight then the existence of effects of avoidance, parents prefer to pay and send their child to a private school rather they attend a RAR junior high-school.

The assessment provide by these four evaluations is far from being positive, and justify the criticisms directed to this policy. Those studies even show that over time, and so usually with improvements and adjustments that are supposed to be made in such case, the effects are bad and tend to exacerbate existing problems and segregation.

In the same vein, the evaluation of the Title I program done by Van Der Klaauw (2008) in New York City highlights that this "program was unsuccessful in improving student outcomes in high-poverty school in New York City during the 1993, 1997 and 2001 schools years, and may in fact have had adverse effects during earlier years. Less evidence of a negative effect is found for the 2001 school year". The resources are not optimally spent and thus call into question the efficiency of this program.

Another program is reported as not having the desired effects: the Special Educational Needs program which was assessed by Crawford and Vignoles (2010) on the Avon region (England) and by Keslair, Maurin and McNally (2012) on the entire British territory.

Keslair *et al*. found results suggest that the program is ineffective for children with moderate difficulties, and more generally that there is no overall effect on account of the combined direct and indirect effects.

Crawford and Vignoles go further and show that "pupils with SEN label score about 0.3 standard deviations lower at Key Stage 2 than otherwise identical pupils without a SEN label".

Hanushek, Kain and Rivkin (2002) have evaluated special education programs in Texas. They found that "special-education programs on average have significantly beneficial effect on performance" and this effects are larger for certain part of the population suffering from lighter disability.

Another program who is reported as efficient and as having positive effects is the Excellence in Cities, who was assessed by Machin, McNally and Meghir (2004, 2010). In both studies, they found a positive impact on school attendance and pupil attainment within EiC schools relative to comparable but non-EiC schools.

Jacob and Lefgren (2004) are interested in the Chicago's Social Promotion policy (CSP), and had concluded that summer school and grade retention in this case have a modest but positive net impact on student achievement scores for third-grade students.

Lavy and Schlosser (2005) had shown that the Bagrut 2001 Program had positive effect on matriculation rates. But with a cost-benefice analysis of this program and others implemented in the same, the authors show that this program had relatively less positives important effects in the light of these costs than the other policies.

---

[2] Réseaux Ambition Réussite in French that we can translate to Ambition Success Network

5.2.2. Differentiated effects

As we can see from previous section, the effects of the compensatory educational programs are various and varied. Some are reported to be effective, others are ineffective and some are even counterproductive. Results of compensatory educational programs, as they are globally presented, are mitigated. But by analyzing all specifics results of the studies, we can go further than general conclusions put forward by the authors.

Our goal is to search specifics effects which can allow us to highlight heterogeneous effects, differentiated effects that could qualify the conclusions sometimes too general. Indeed, the fact to say that overall impact of a program is null can hide the compensation between to effects: one positive and one negative. In the same way, saying that a policy is globally negative doesn't mean that every measure which composed it, every effect produced, is bad. It can simply be the predominance of a negative effect among positive or neutral ones.

The objective of this advanced analysis of papers is to identify factors of differentiation who offer ways to improve the effectiveness of compensatory educational programs.

Table 1 lists all the effects identified by the authors. From the analysis of every evaluation, we have identified two main aspects: the role of student characteristics and the role played by school characteristics.

The differential effects depending on the school characteristics

The school context seems to play a key role in the effects that a program could have. The question is: in which direction? Is an advantaged background more profitable to student's achievement? Or are programs more efficient for pupils in disadvantaged schools? No clear answer is provided on the analyzed studies. We know that the effects of the program could be different in function of the socio-economic context of each school where the program is implemented. But the evaluations are not unanimous. Some like Crawford and Vignoles (2010) reports that even a gap subsist between treated and non-treated students; the SEN program induces quicker progress for treated students in the more advantaged schools. This may suggest that a favorable context of studying is beneficial for treated students, who can hope to catch up "normal" students. That can also suggest that advantaged schools are making more effective use of their SEN budget.

Conversely, Machin *et al.* (2010) tells that the EiC program have had much larger effects in more disadvantaged schools.

Others characteristics of schools can play a role in the effectiveness of an educational program. The size of the school and the years of experience in the program influence its results.

The evaluation of the Bagrut 2001 Program, made by Lavy and Schlosser (2005), shows that effects of this policy are much more important in smaller schools. While on its side, Machin *et al.* (2010) have shown that the length of exposure of the schools to the program leads to a greater impact of the program. Indeed, "EiC increases the probability of attaining level 5 or above by 2.9 percentage points in EiC phase 1 schools, by 1.5 percentage points in phase 2 schools, by 1.7 percentage points in cluster 1 schools, and by hardly anything for schools in phase 3 or in cluster 2 (i.e., more recent entrants to the policy)" (Machin *et al.* 2010, p.379)

The differential effects depending on the student characteristics

In almost every evaluation, the impact by age or class level is tested. On the whole, when results are significant, programs have greater impact on the younger. For example, Jacob and Lefgren (2004) who assess the Chicago's Social Promotion policy found no effect of grade retention on the whole sample of students, but they find a positive effect for the younger (3rd graders comparatively to 6th grades). Indeed, for 3rd graders, summer school and grade retention increase student achievement roughly 20% of a year's worth of learning, while the effect is null on the 6th graders.

In addition to the students' age, their abilities are also factors of differentiation. As for the school context, the effects of those factors are ambiguous. Almost every evaluation studies the impact of the program according to initials abilities of pupils. Some, like Machin *et al.* (2004, 2010) for EiC, show that students who derive the most benefits from the program are high-ability students. While Lavy and

Schlosser (2005) find that the effects of program decline monotonically with student ability. They show that the effect on the lowest ability quartile is twice that on the third ability quartile.

Article	Year	Policy analyzed	Data	Sample	Research Design	Results
Beffy M. and Davezies L.	2013	RAR	Different administrative files of the Ministry of Education: . an exhaustive pupil-level cross-sectional data (*Scolarité*) for every student in junior or secondary high school . An exhaustive teacher-level panel data (*Relais*) for every teacher in junior or secondary high school . An exhaustive pupil-level data of their scores to the final national exam (*Brevet des Collèges*) . An exhaustive panel of French junior and secondary high schools	5,000 public juniors high schools in metropolitan France, including 206 affected to the RAR program	RD	. For treated schools, the results to the final junior high-school exam are worse than in similar non-treated schools . The effects on class size and number of hours per class are far from those expected . Slight ↘ in independent and intermediate profession children recruitment, and ↗ of workers' children (adverse effect) . ↗ teachers with non-standard qualifications, ↘ of the proportion of teachers having the "agrégation" (i.e. the top teaching diploma) . When it is significant, the treatment has always a negative effect on achievement
Benabou R., Kramarz F. and Prost C.	2009	ZEP	. FSE (*Fichers Standards Enrichis*) administrative files of the Ministry of Education : school-level data . The 1980 panel: student-level data . The 1989 panel: student-level data . Political dataset	. FSE files: 4,743 establishments per year between 1987 and 1992 . The 1980 panel: 20,691 students who entered 6th grade in 1980 . The 1989 panel: 24,455 students who entered 6th grade in 1989 . Political dataset:	DD, IV	. <u>General impact of ZEP</u>: extremely slow ↘ but continuous of class size (-0.2 students per year in average); ↗ teaching hours per student (+1.2% per year); slight ↘ in teacher experience despite the bonuses offered . <u>Impacts of ZEP status on school characteristics</u>: loss of 7 students in average by year for schools that became ZEP in 1989, 4 students per year between 1987 and 1999 for schools that became ZEP in 1982, 6 students per year between 1995 and 1999 for schools that became ZEP in 1990 (due to reduced entry rather than to increased exit); no major changes in the social composition of schools . <u>Impacts of ZEP status on individual schooling achievement</u>: whatever they use DD or IV, the impact is never significantly different from 0, regardless of the measure of achievement used
Caille J.P, Davezies L. and Garrouste M.	2013	RAR	. 2007 student panel	. 30,924 students	RD	. No effects of RAR on achievement and educational trajectories of students
Crawford C. and Vignoles A.	2010	SEN	. Avon Longitudinal Study of Parents and Children (ALSPAC)	. 7,742 pupils attending 278 different schools	PSM	. Pupils with SEN score, on average, 0.408 SD lower at KS2 than pupils with the same prior attainment who do not have SEN . Pupils with SEN seem to make relatively greater progress in more advantaged schools
Davezies L. and Garrouste M.	2014	RAR	. Exhaustive individual data of French pupils entering 6th grade in	. 1,098,636 individuals,	RD, IV	. Living near a RAR junior high school tends to ↘ the individual probability to attend the default option junior high school, for pupils just above the thresholds

			2006 and 2007 . Exhaustive data at the school level	with 531,729 entering 6th grade in 2006 and 566,907 in 2007		. Living near a RAR junior high school tends to ↘ the probability to go to another public school, for pupils just above the thresholds . Living near a RAR junior high school tends to ↗ the individual probability to attend a private school, for pupils just above the thresholds . Those effects are even more important that students come from privileged backgrounds
Hanushek E., Kain J. and Rivkin S.	2002	IDEA	. The Texas Schools Microdata Panel constructed by the UTD Texas Schools Project	. 3 successive cohorts of Texas public elementary school students beginning in 1993 200,000 students by cohort in over 3,000 public schools		. <u>Effects on special education students</u>: the average achievement of students classified as speech-impaired, in all grades, is at least 0.7 SD higher than the average for those classified as learning-disabled and at least 0.5 SD higher than for those classified as emotionally disturbed; the average effect of special education for all disabilities is positive once student heterogeneity is allowed for with fixed effects; estimated program effects are much larger when derived from the students entering special education than from those who exit; program impacts declines by ∼25% in the 2nd year for the average participant and ∼40% for the average learning-disabled student. . <u>Effects on regular-education students</u>: no evidence that special education harms achievement in regular classrooms; a 10-percentage-point ↗ in the percentage of students classified as disabled ↗ achievement ∼0.016 SD
Jacob B. and Lefgren L.	2004	CSP	. Administrative data from the CPS system, individual level	. Cohort of students who were in 3rd and 6th grades from 1993-1994 school year to the 1998-1999 school year (total of 402,924 obs.)	RD, IV	. For 3rd graders, summer school and grade retention ↗ student achievement ∼20% of a year's worth of learning. By the second year after the program, the effects had faded by ∼25% to 40% . The net effect for 6th graders was essentially 0 in reading, and close to 0 in mathematics, particularly by year 2
Keslair F., Maurin E. and McNally S.	2012	SEN	. National Pupil Database (NPD) . Consistent Financial Reporting (CFR) data	. All pupils observed in the NPD at age 11 between 2002 and 2008	DD, IV	. Highly able pupils are almost never assigned to a SEN program, regardless of the school context . Substantial gap between the probability of being assigned to a SEN program in a high context school compared to a low context school . Very significant variation in SEN resources across pupils with different abilities and across school contexts . No net effect of being assigned to SEN program on the educational performance of pupils with moderate difficulties compared to other pupils in the same year group (but no spillover effects)
Lavy V. and Schlosser A.	2005	Bagrut	. Unspecified	. 163 schools . 4,100 students, 1/5 of all students in treated schools	DD, IV, SPM	. Positive impact of the program on treated schools: ∼3.3 percentage point ↗ in the mean matriculation rate, implying an improvement of 6% . The program boosted the matriculation rate of treated schools by 3 or 4 percentage points . The higher treatment intensity, the greater the improvement in the school mean matriculation rate . The program affected the achievements of treated students only, ↗ their probability of earning a matriculation certificate by 13 percentage points on average . The effects of the program ↘ monotonically with student ability

						. Program less efficient than other programs implemented in the same period: Bagrut 2001 program produced a gain similar to that of the teacher bonus program but at almost twice the per student cost
Machin S., Meghir C. and McNally S.	2004	EiC	. Administrative records of pupil-level attainment . Administrative school-level data	. 241,789 students and 699 schools	DD, matching	. A bigger improvement in the average Maths and English performance of EiC pupils relative to the comparison group: average level of attainment ↗ by 0.18 for EiC pupils, but by 0.16 in all secondary schools not in EiC . The average percentile in KS3 English performance ↗ by 0.62 of a percentile, as compared to a fall in the non-EiC schools .The group of EiC schools is the only one where absences ↘ . Improvement in maths of ~ 0.03 of a level for pupils in EiC schools . Results also positive, but weaker in statistical terms for English. . The average policy impact is of the order of 0.5 to 0.8 of a percentile
Machin S., Meghir C. and McNally S.	2010	EiC	. Pupil-level: National Pupil Database . School-level: LEASIS, school performance tables, school change file	. 3,157 schools (including 1,009 EiC schools)	DD, matching	. ↗ the probability of attaining level 5 by 1.8 percentage points if one considers outcomes in the most recent year of the policy (2003) with the prepolicy year . Larger effects are shown for schools that have been in the EiC policy for longer: EiC increased the probability of attaining level 5 or above by 2.9 percentage points in EiC phase 1 schools, by 1.5 percentage points in phase 2 schools, by 1.7 percentage points in cluster 1 schools, and by hardly anything for schools in phase 3 or in cluster 2 (i.e. more recent entrants to the policy) . EiC policy has had an impact on raising achievement in maths in schools exposed to the policy. However for achievement in English, no effects are found . The EiC policy led to a ↘ in the percentage of half days missed by about 1 percentage point (0.59 with controls) . The effect on absences varies by EiC phase: for the three main phases of EiC (phase 1-3) the effects are -.762, -.634, and -.349, respectively . Larger effects of the policy are found for pupils of high or medium ability than for those of lower ability and are generally found only within more disadvantaged schools
Van der Klaauw W.	2008	Title I	. School-based data collected by the New-York City Board of Education's Office of Research, Evaluation and Assessments, and provided by the NYU's Institute for Education and Social Policy	. All public elementary and middle schools in 1993, 1997 and 2001	RD, sensitivity analysis	. Title I has been ineffective at raising student performance, and appears to have had adverse effects during the 1993 and 1997 school years. Less evidence of adverse effects is found for 2001 . Title I status has led to slightly lower attendance rates, slightly higher grade repetition rates and higher rates at which students entered and left the school during the school year . Small positive effect on teacher absence rates . Most of the effect estimates imply a negative effect on performance, which in several cases is statistically significant

DD= difference in differences; IV= instrumental variables; RD= regression discontinuity; PSM= propensity score matching

Table 3: Evaluation of educational programs and theirs results

6. Which program is the one?

The previous analysis made of the evaluations of compensatory educational programs brings us to take stock of the lessons learned from the evaluations, and from there to provide recommendations.
The fact that we can bring out global effects but also differentiated ones leads us to think at potential ways of improving programs, or at least, to think at some points which can underpin a more important efficiency of educational policies.
We make a comparison of the different evaluations in order to identified aspects of policies that work well. This could help to implement improvements to policies whose results are much nuanced and often criticized.
Thus, we believe it would be wise to rethink policy targets, but also to develop/strengthen accompanying measures of programs.

Rethink policy targets

As we said previously, the age of the beneficiaries play a key role in the success of an educational policy.
Based on our analysis, but also on the study of Currie (2001) on early childhood education in United-States, we affirm that more the child is treated young, more the program would likely have an positive impact on him. And so, more the program would gain in efficiency.
However, the majority of educational policy concerned students from primary school to the end of senior high-school. We believe that this represents a dilution of allocated resources to too many people. Then, we think that it would be necessary to refocus policies on younger as guarantee of efficiency.
This intuition seems to be confirmed in the case of France, where the policy of priority education is inefficient. An evaluation report conducted by the Ministry of Education (2013) explain that: (i) expenditures per students of secondary school are 15% higher than the OECD average, and that, on the contrary, (ii) expenditures per students of primary school are 17% below the OECD average.
OECD tells that the French education system is among those with the worst results. This may be due to a misallocation of resources between age groups.
Unlike many OECD countries, France is investing heavily in junior high-school level. But according to the results of the latest PISA survey (OECD, 2014), the French system does not produce good results and is increasingly unequal. So one might think that, maybe, in France, the government does not intervene on the right people to produce the desired results.

Student's abilities are also a criterion for differentiation in policy effectiveness.
Several studies highlight that positive effects of programs are more important for students who having medium to high abilities. One might think that to improve efficiency, it is necessary to focus on these special populations. But doing that would be totally contrary to the principle, the aim of these policies. Remember, the basis for these policies is based on the principle of positive discrimination. According to this principle, the objective is to help those who need it most. In attempt to improve efficiency while maintaining consistent population treated, one can see the treatment of students with high abilities as a dilution of resources, and therefore as a lower effort headed to those who really needed it.

Develop/strengthen accompanying measures

We also asked whether the lack of relative effectiveness of policies was not due more to the fact that they are insufficient to counter alone the problems of underperforming at school and segregation. Maybe their current working is not bad, and can be successful if it could be supported by additional measures.

In individualized policy frameworks, these programs could be accompanied by measure to raise the self-confidence and the self-esteem of treated students. Measure of reintegration into the classroom and within classmates can also be a good support action. Indeed, individualized treatment requires

regular exits of class, avoiding the student from taking the same course that these classmates. Then, treated pupils can feel excluded, and the treatment is a source of stigmatization. These feelings can bring to an introversion which could have harmful consequences on educational achievement despite the treatment.

In the place-based policy framework, territories receive extra funding because of their disadvantaged socio-economic background. These neighborhoods suffer from a poor image, from stigmatization. Strategies of avoidance of these neighborhoods are common. Those who can afford it, will choose another place of residence. For those unable to afford housing in another part of town, they may seek to avoid any activity in this neighborhood, especially the schooling of their children. These territories suffer from a growing segregation with the pauperization of such neighborhood. The idea of an accompanying measure is to try to enhance the image of these neighborhood, with a view to re-introduce more diversity. This diversity, by peer effects, will pull educational attainment up. This type of measure can be towards families. Another type of intervention can be considered. Chiapa *et al.* (2012) have shown that the fact that poor families rub educated people (in this case medical personnel) increases the ambitions of parents for their children's school careers. Interventions in families can show all the prospects for their children, and so boost motivation of parents to direct their children to school so that hey then have all the cards to get by in life.

A much more extreme solution could also be considered. If compensatory educational programs are necessary, it is because the current educational system is not efficient. The solution could be to try to replicate the best performing education systems, namely the Asian systems. They place great emphasis on the selection and training of teachers. They also set them clear objectives while allowing them ample freedom to achieve these aims.
Rather than hanging the problems generated by a system, we should take the problem at its base and reform this system.

We have seen that evaluations of compensatory educational programs explain that these programs have very mitigated effects, which is consistent with criticisms that are usually expressed: a lot of money is spent on these programs, for only few results.
But maybe this is not programs that are inefficient, because otherwise they would have been stopped or replace long time ago. Maybe it is the evaluations, by their technical nature, which are not able to translate real effects of the programs. We will see in the next part the difficulties of econometrical evaluations.

7. Which method is the one?
7.1. Theoretical framework[3]

A lot of econometric methods have been developed over the last twenty years for realizing evaluation. The principle of the economic evaluation falls within the framework of Rubin (1974). The vocabulary used borrows heavily from the field of medical experimentation which it is derived.
Empirically, we are interested in the evaluation of a measure T. Basically, we can distinguish people who benefit from this measure ($T = 1$) of those who do not benefit ($T = 0$). We are interested then in the effect of this measure on a variable of interest, often called *outcome*. Everybody had virtually potential outcomes: the outcome Y_0 associated to $T = 0$ (non-treated) and Y_1 associated to $T = 1$ (treated).
From these two potential outcomes, we can determine the causal effect Δ_i of the measure:
$$\Delta_i = Y_{i1} - Y_{i0}$$
This causal effect has to central characteristics:
  1- It is individual. Nothing constrains the effect to being the same in the whole population or homogeneous in some subpopulations. There is a distribution of the program's effect in the population.

---

[3] Elements from Crepon and Jacquemet (2010), Givord (2010)

2- It is unobservable. Indeed, we observe either Y_1, if $T = 1$, or Y_0, if $T = 0$, but not both simultaneously. Information available are on the form :

$$\begin{cases} T \in \{0,1\} \\ Y = TY_1 + (1 - T)Y_0 \end{cases}$$

An important hypothesis of the Rubin's model is that there is no externality. Treatment of the individual i has an effect only on his own outcome, not on the others ones. This hypothesis is called SUTVA, for Stable Unit Treatment Value Assumption.

Evaluation is thus a problem of missing data. Identifiability and identification of the causal effect of treatment are the key issue of the evaluation. This identification is not acquired *a priori*.
All methods of evaluation involve offering different ways for recreate this alternative and unobserved situation for individuals, or in other words, reconstruct the missing information.

7.2. Relevance of the control group: selection bias and endogeneity

The ideal evaluation consists in comparing the situation resulting from the public policy to the hypothetical situation that would have occurred in the absence of implementation of this policy. But this counterfactual situation is not observable. The main difficulty of the evaluation is to reconstitute what would happen in the absence of the policy.

The naïve estimator will compare the average individual situations as they received treatment or not:

$$\hat{\Delta}^0 = \bar{Y}^{T=1} - \bar{Y}^{T=0}$$
$$\Delta^0 = E(Y_1|T = 1) - E(Y_0|T = 0)$$
$$\Delta^0 = \underbrace{E(Y_1|T = 1) - E(Y_0|T = 1)}_{\text{ATT: Average Treatment on the Treated}} + \underbrace{E(Y_0|T = 1) - E(Y_0|T = 0)}_{\text{Population effect}}$$

The first term measures the causal effect which is interesting us. The second term, the population effect correspond to the gap in the absence of treatment for individuals who have benefited and those who are excluded. The naïve estimator compares the situation of beneficiaries to the average situation of a control group. The population effect correspond to the fact that the control group is not good: it is not representative of what would be the situation of beneficiaries in the absence of the treatment. The naïve estimator confuses the both components "program effect" and "population effect" and make them indistinguishable. In this sense, the naïve estimator is biased. It is a selection bias. Nevertheless, this estimator is a convergent estimator of :
- Δ^{ATT}, the average treatment on the treated, if the affectation to the treatment is independent of the potential outcome Y_0, i.e. $Y_0 \perp T$
- Δ^{ATE}, the average treatment effects, is the affectation is independent of the two potential outcomes Y_0 et Y_1.
But generally this is not the case; the naïve estimator is probably consistently biased.

7.3. Different methods of estimation

Different empirical methods are developed, and each one brings on its way an answer to problems of autoselection and heterogeneity. The main issue is the ability to find data, situations that neutralize the effect of selection.
Experimental methods based on randomization as a source of identification are used to circumvent the selection bias. These methods involve assigning randomly individuals between the two groups. There is no longer reason for individuals in the treatment group to be different on average from those in the control group.
There are two main cases where the principle of experimental methods is used:

17

- Fields experiments. This consists to distribute randomly, usually by lottery, individuals from a population between (i) the control group (individuals will continue their activities without any change in their situation) and (ii) the treatment group (individuals will operate in an environment that differs owing to the implementation of a measure). This ensures the exogeneity of treatment participation.
- Lab experiments. This consists to build a controlled environment in which participants interact. Participants were recruits through advertising campaigns that highlight the opportunity to earn money or gifts. Behaviors detected in the experiments are used to observe the reaction of individual behavior to a given environment.

Such methods are rarely used, especially in Europe and France, and even less in terms of public policy.

Pubic policies implemented generally target a specific audience; individuals are treated according to their characteristics.

Most of public policies are non-random experiments, there have eligibility criteria based on specific characteristics of individuals. So individuals who are treated, by nature, differ from non-treated individuals.

To solve this, we can us methods of matching, which are combined with estimators of propensity score to highlight the effect of the program. Other methods attempt to use exogenous changes in the economics environment, i.e. implementation of the program to create quasi-experimental situations or natural experiments. This is the case of difference-in-differences, the method of instrumental variables, or regressions discontinuity.

7.3.1. Difference in differences

Principle

The difference in differences (DD) estimator aim at measuring difference of outcome between treated and non-treated populations before and after the policy and at performing the difference of these two evolutions.

As we seen previously, compare beneficiaries to non-beneficiaries is likely to be biased by the presence of selection effects. A comparison before/after is no more satisfying. A lot of factors are likely to explain variations of outcomes in time, independently of program's effects.

The idea of the method of DD is to mix these two approaches. In the simplest case, we have two groups observed before and after the implementation of the program, with only one of them affected by the program. The estimator consists in comparing the evolution of the beneficiaries' outcome before and after the treatment to the evolution of the control group over the same period. We make two differences:
- One should give the opportunity to eliminate systematic differences between the treated group and the other one;
- The other should give the opportunity to elimination the temporal evolution, supposed to be identical for both groups in the absence of program.
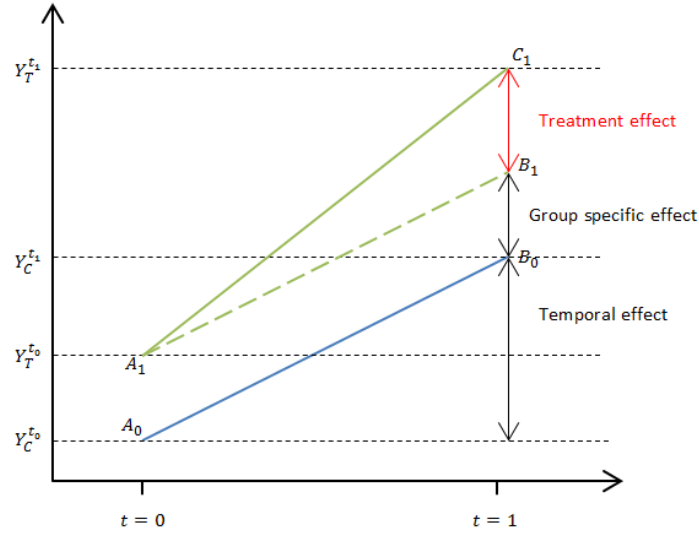
Identifying assumption

Figure 2: Difference in difference method

Evolutions of the outcome without treatment of both groups are identical. But the implementation of the program has modified the outcome for treated individuals. The naïve estimator which compare treated and non-treated (distance B_0C_1) is biased by selection effect, because beneficiaries are per se different from non-treated (*ex ante* difference B_0B_1). The real effect of the treatment is given by B_1C_1.

Determination of the control group

Most of studies who mobilize DD method define the control group of an *ad hoc* manner. In general, comparing the situation with a group considered being similar, but there is always a degree of arbitrariness in this choice.
Some use matching to achieve the most relevant control group.

Estimation

The estimator of DD is as follow:
$$\Delta = E(Y^{t_1} - Y^{t_0}|T = 1) - E(Y^{t_1} - Y^{t_0}|T = 0)$$

In which Y^{t_1} is the outcome observed after the implementation of the program and Y^{t_0} the one before the implementation. In practice, this corresponds to the OLS estimator in the regression:
$$y_{it} = \beta_0 + \beta_1 1_{t_1} + \beta_2 1_{T_i=1} + \delta 1_{t_1} 1_{T_i=1} + u_{it}$$

With t_1 is the date after the application of the program.
In the case where there are several groups and time periods, the corresponding estimation is:
$$y_i = e_t + e_g + \delta T_{gt} + u_{it}$$

With T_{gt} is an indicator that the group g has been treated at time t, e_t show temporal fixed effects and e_g are group fixed effects.

Limits

The main limit of the method of difference-in-differences is that the identifying hypothesis is not testable, very fragile and sometime even not credible. Indeed, that would mean it would be possible to observe the counterfactual evolution of the treatment group in the absence of treatment.

The hypothesis of the identical evolution between groups is *a priori* more credible that the time period is short (Bertrand *et al.*, 2004). But a public policy has rarely short-term effects, it is much more

interesting to analyze these effects in the medium to long term. *De facto*, in practice, studies are based on estimated over pretty long time periods.

The assumption that there is no cross effects group/time periods is not plausible. And not taking into account can biased the inference.

7.3.2. Regression discontinuity

Principle

The method of regression discontinuity relies on the existence of a selection variable (usually denoted S) which has a discontinuous effect on the probability of being treated. More precisely, it is to exploit the fact that many policies have threshold attribution rules. The intuition behind this method is that around this threshold individuals are almost identical even through some are beneficiaries of the treatment and some other not.

In general case, we focus on the effect of the treatment T_i in an outcome variable y_i:

$$y_i = \alpha + \Delta_i T_i + u_i$$

In which $\Delta_i = Y_{i1} - Y_{i0}$ is the treatment effect for individual i, $\alpha = E(Y_0)$ the average potential outcome without treatment, and residuals $u_i = Y_{i0} - E(Y_0)$.

In practice, we must distinguish two cases.

Sharp design

In this case, treatment T depends deterministically of the selection variable S:

$$T_i = T(S_i) = 1(S_i > \underline{S})$$

Identification hypothesis are pretty general. The observed outcome must actually have a discontinuity at the selection point. It pass "suddenly" from outcome without treatment Y_0 to outcome with treatment Y_1.

The identification hypothesis is then: potential outcomes Y_1 and Y_0 are continuous around the discontinuity point of T. In other words, the inobserved component of the outcome $E(u_i/S)$ is continuous in $\underline{S}$, just as the average treatment effect $E(\Delta_i/S)$

Then, the treatment effect at discontinuity point is:

$$E(\Delta_i | S = \underline{S}) = \lim_{S \searrow \underline{S}^+} E(Y|S) - \lim_{S \nearrow \underline{S}^-} E(Y|S)$$
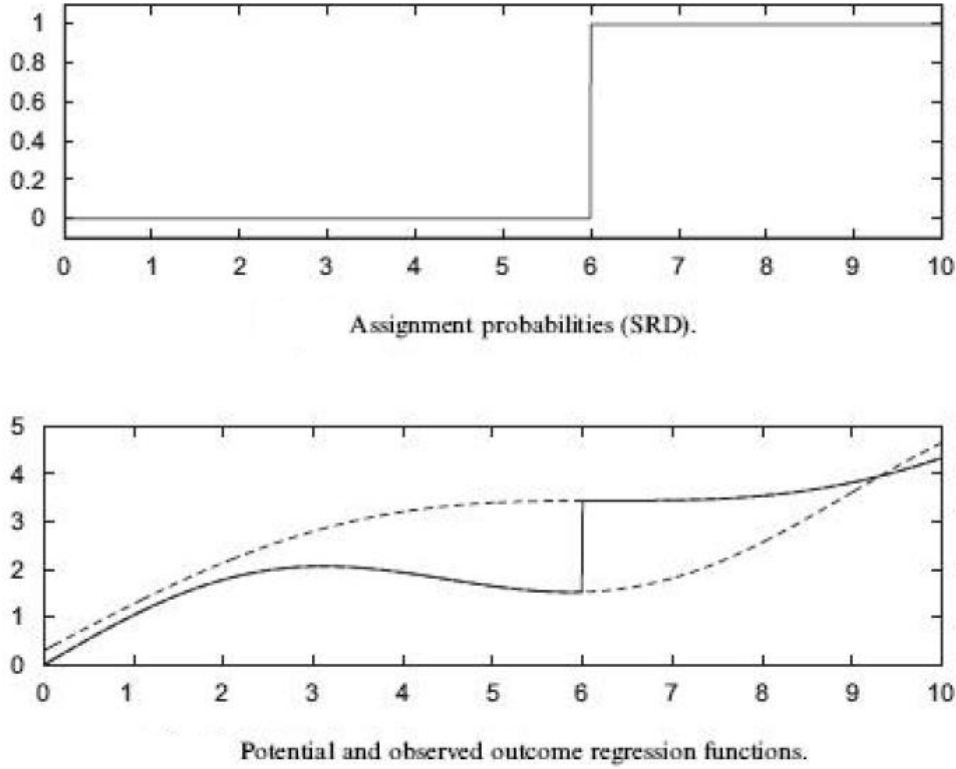
Assignment probabilities (SRD).

Potential and observed outcome regression functions.

Figure 3: Potential outcomes with and without treatment, case of sharp design
Source: Imbens and Lemieux (2008)

Fuzzy design

In this case, treatment depends of the selection variable S, but the assignment is not perfect. It is only $P(T = 1|S)$ which having a discontinuity on $\underline{S}$.

If the treatment effect is steady is the area of discontinuity point, it is given by:

$$\Delta = \frac{\lim\limits_{S \searrow \underline{S}^+} E(Y|S) - \lim\limits_{S \nearrow \underline{S}^-} E(Y|S)}{\lim\limits_{S \searrow \underline{S}^+} E(T|S) - \lim\limits_{S \nearrow \underline{S}^-} E(T|S)}$$

If the treatment effect varies with individuals, an additional hypothesis is necessary.

The "local" independence" of the treatment can be supposed: $T_i \perp \Delta_i | S_i$.

Or, we can use a less restrictive hypothesis: the one of monotoniciy: $T_i(S)$ is nondecreasing in S, for all i, around the discontinuity point.

Estimations
Semi-parametric estimation

We can use a kernel estimator to estimate $\lim_{S \searrow \underline{S}^+} E(Y/S)$:

$$\frac{\sum_i 1_{(S_i > \underline{S})} k_h(S_i - \underline{S}) Y_i}{\sum_i 1_{(S_i > \underline{S})} k_h(S_i - \underline{S})}$$

Sharp design:

To estimate outcome value ar the right of the threshold $\lim_{S \searrow \underline{S}^+} E(Y|S)$, we look for parameters $(a_-, b_1, b_2, \ldots, b_p)$ such as:

$$\min_{(a_-, b_1, \ldots, b_p)} \sum_i 1_{(S_i - \underline{S})} k_h(S_i - \underline{S}) \left(Y_i - a_- - b_1(S_i - \underline{S}) - \cdots - b_p(S_i - \underline{S})^p \right)^2$$

The estimator of $\lim_{S \searrow \underline{S}^+} E(Y|S)$ is then:

21

$$\hat{\mu}(\underline{S}) = \hat{a}_- + \hat{b}_1(\underline{S} - \underline{S}) + \cdots = \hat{a}_-$$

It is the same for the estimator of the left boundary.

Causal effect of the treatment in a sharp design is estimated by:

$$\hat{\delta} = \hat{a}_+ - \hat{a}_-$$

Fuzzy design:

The principle is the same, but in addition it is necessary to estimate $\lim_S E(T|S)$ at right and left of the threshold.

The causal effect of the treatment is:

$$\frac{\hat{a}_{y+} - \hat{a}_{y-}}{\hat{a}_{t+} - \hat{a}_{t-}}$$

Parametric estimation

it may be that we do not have enough detailed information in order to have enough observations around the point of discontinuity.

We can overcome these deficits using larger samples, even using more remote observations from the point of discontinuity. We control the dependent variable of interest in the selection variable in a parametric specification generally polynomial.

Sharp design:

Linear classic model:

$$y_i = m(S_i) + T_i\Delta + e_i$$

With $E(e_i|T_i, S_i) = 0$, $\Delta = E(\Delta_i|S_i = \underline{S})$ and $T_i = 1_{(S_i > \underline{S})}$.

We have $e_i = y_i - E(y_i|T_i, S_i)$ wich guarantee by construction the idenfitication hypothesis $e_i \perp T_i$.

$$m(S_i) = \alpha + E(u_i|S_i) + T_i(E(\Delta_i|S_i) - E(\Delta_i|S_i = \underline{S}))$$

According to the assumption of local continuity, $m(S)$ is continuous in S.

If $m(S)$ is well specified, we can estimated the average treatment effect by a simple linear regression. In practice, $m(S)$ is often approximated by polinomial functions.

Fuzzy design:

$$y_i = m(S_i) + E(T_i|S_i)\Delta + e_i$$

With $e_i = y_i - E(y_i|S_i)$, $m(S_i) = E(u_i|S_i) + E(T_i\Delta_i|S_i) - E(T_i|S_i)E(\Delta_i|\underline{S})$.

We can estimate the effect by a two stages procedure:

- $E(T_i|S_i) = f(S_i) + \gamma 1_{(S_i \geq \underline{S})} + v_i$
- We use the estimation of $E(T_i|S_i)$

If $f(S)$ and $m(S)$ are identical, this a 2OLS procedure avec an endogeneous variable T and an instrument $1_{(S_i \geq \underline{S})}$.

Limits

The limitation of this method is that we estimate only a local effect of treatment. Conclusions are only valid around threshold. If the treatment is not steady on the population, then the estimates interpretation would be limited.

So that the hypothesis is valid, it must be sure that people cannot "handle" the threshold value, in order to benefit from the treatment or not.

7.3.3.Instrumental variables

Principle

Instrumental variables can be used in the evaluation of public policies, but have a much broader scope. These are used for many years (dating back to Wright in 1928) to address endogeneity issues.

We are in the standard linear framework where regressor T_i is suspected to be endogenous, i.e. correlated with residuals u_i:

$$y_i = \alpha + \beta T_i + u_i$$

This is typically the case when there are selection issues. Their use is therefore suitable for evaluation of public policy.

An instrument is then simply a variable Z which satisfies two conditions:

1. It is correlated to the endogenous variable T: $cov(T, Z) \neq 0$
2. It is not correlated to residuals u: $cov(u, Z) = 0$

If such variable exists, it is possible to obtain a consistent estimated of the parameter of interest δ. Classic estimator is 2OLS. It is obtain in regressing the outcome Y not on the endogenous regressor T, but on its prediction $\hat{T}$ from the first stage of the regression of T on Z:

$$\hat{\delta}^{IV} = \left(\hat{T}'\hat{T}\right)^{-1} \hat{T}'Y$$

When the instrument is binary, a common estimator in the evaluative literacy is the Wald estimator:

$$\hat{\delta}_{IV} = \frac{E(Y|Z = 1) - E(Y|Z = 0)}{E(T|Z = 1) - E(T|Z = 0)}$$

The choice of the instrument can be very complicated. A valid instrument is a variable that is related to the fact of getting in a program for a reason that can explain, but otherwise has no direct impact on other determinants of outcome.

Limits

Most of the time, the discussion focuses on the quality and validity of instruments. Validity is based on the fact that the instrument is not correlated with unobserved determinants of outcome. This property is crucial; it ensures that the estimator obtained by instrumental variables is asymptotically unbiased. However, no test exists to allow rigorously validate this property.

Moreover, the fact to research instruments for which the exogeneity cannot be blamed makes them sometimes weakly correlated with treatment. Then, we are talking about low instrument. When we are interested on the effect of the treatment T on the outcome Y. T is endogenous, but we have a vector Z of K instruments such as $E(Z'u) = 0$ related to the treatment by the relation:

$$\begin{cases} Y = T\delta + u \\ T = Z\Pi + v \end{cases}$$

In this case, common estimator is 2OLS obtained by regressing the outcome not on T but on its projection $\hat{T} = Z\hat{\Pi}$. Π is the measure of the intensity of correlation between instrument and endogenous variable.

In practice, for measuring the strength of an instrument, we can use the "concentration" which is defined like a ratio between the part of the endogenous variables explained by the instrument and the part which is not:

$$\mu^2 = \Pi'Z'Z\Pi/\sigma_v^2$$

In this case, estimators properties obtained are not satisfactory; estimates can be imprecise, or even more biased than naïve estimators. What is even more serious when the number of instruments used is high.

7.4. Implication of the existence of various methods

We have seen previously that various methods exist, with different estimation methods. This variety of methods, with their own limits, associated with the recurring problem of formation of the control group, can generate a large number of biases which can distort the analysis of studied policies. It is this diversity, and the high probability of existence of bias (despite efforts made by researchers to avoid it) that led us to consider the possibility of misinterpretation of policy outcomes.

Actually, evaluations between them question their results. For example, Beffy and Davezies (2013), by challenging the method used by Benabou *et al.* (2009), also question the results found by these latters. Indeed, Benabou *et al.* use the difference in difference method to estimate the effect of ZEP.

As we seen in previous section, this method relies on the existence of a common trend between the potential outcome of the treated group without treatment and the outcome of the control group. However, according Beffy and Davezies, this common trend is not obvious. So should we take seriously the results provided by Benabou *et al.*? To what extent the bias induced by poor specification prevente from reflecting reality? But, are results of Beffy really better? Although no specification problem is found, the choice of the regression discontinuity implies that their results cannot be generalized, and say nothing about the effect of the policies on the whole treated (and non-treated) population.

7.5. Necessity to take into account the individual or group comportment

Even if any problem was found from a technical point of view, evaluations cannot entirely reflect effects of the programs. Indeed, we are in presence of economic behaviors which are modified by the existence itself of the policy. And these behaviors are not taking into account on the estimations of the programs effects.

As we mentioned earlier, the stigmatizing effects of these policies lead to avoidance strategies. These behaviors, which result from the implementation of the policy, play heavily on the effects of the program, and are not yet included in the estimate. Their estimation and their inclusion in the model are extremely complicated, but the fact to do doing it leads to poor estimation of the effects.

Mimetism and conformism are two behaviors that imply a modification of the real effects of educational programs. We may think that a student, enrolled in a treated school, a school suffering from a bad image, whose students are considered as difficult, unruly and of low abilities, does not seek to make an effort, as in all cases he will be stigmatized. He will therefore conform to the image that we have of students frequenting these establishments.

Behavioral problem can also come from modification of teachers expectations. In the presence of students in difficulties, teachers tend to have lower expectations and requirements than for "normal" students. They do not seek to take students up. They tend to be resigned about the difficulty of the situation.

Another type of comportment can biased effects announced by evaluations: substitution effect. This means that additional funds provided can be used instead of the usual funds, annihilating any compensatory effect. Thus, treated schools do not use additional resources compared to non-treated schools. So it is normal that these programs appear to be ineffective. In this case, the inefficiency does not mean that the program is bad, just that fund management is not done carefully. Van der Klaauw (2008) explains that this is what is happening in New York City with the Tittle I. So, his evaluation reports that Title I has been ineffective at raising performance, and appears to have had adverse effects during the 1993 and 1997 schools years. Maybe if all available funds had been used, the program would raise positive effects.

The problem with all these behaviors is that they are induced by the existence itself of the policy. Taking them into account in the estimate is almost impossible. Indeed, these behaviors are the results of the implementation of the policy but impact also on its other effects. But do not take them into account leads to a bias.
But, some researchers like Zenou (e.g. Calvo-Armengol, Patacchini and Zenou, 2009; Patacchini and Zenou, 2012) open new opportunities to better understand the peer effects through the formalization of networks and their integration into the estimates. A reflection could be conducted to consider the use of such methods in the evaluation of public policies.

8. Conclusion

The present study provides a snapshot of the economic literature on compensatory educational programs.

Usually, people have a fairly critical look at these programs. They are considered to be money pits who having very little effect compared to the means employed. We are therefore questioned the veracity of these remarks.

From our analysis, we cannot be categorical on the legitimacy of these programs. Some programs stand out as having positive effects (but they are a minority), while others appear to be ineffective or even counterproductive. But behind these very general conclusions are hiding differentiated, heterogeneous effects.

Some programs are efficient over a part of the population, but not on another. Anyway, these heterogeneous effects have allowed us to look what might be the good program. We believe that to improve the efficiency of compensatory educational programs, it should avoid excessive dilution of resources by refocusing actions on youngest and most in difficulties publics located in disadvantaged neighborhoods. It might also be useful to support these programs of additional measures to help the program to take its full extent.

We also asked about the following point: if the effects reported by the evaluations did not reflect the reality? And if these evaluations, owing to their technical implementation, were biased? Several methods of evaluations are currently used. Each method has its own limitations, which can lead to biased results (e.g. if the hypothesis of common trend of difference in difference method does not hold) or incomplete results (e.g. in the case of a regression discontinuity, where the conclusions are apply around thresholds of assignation to the treatment). Add to this, the current problem in policy evaluation that is the formation of a relevant control group. One must also know that individual economic behaviors caused by the implementation of the policy, such as conformism or avoidance strategies, modify the effects of these policies. Their estimation is complicated and generally not included in the models, so the results from the evaluations are biased by this lack of information.

Thus, our analysis, although it does not really statue on the effectiveness of compensatory educational programs, sheds light on the improvements which could be made for these policies. Our analysis also learns to be careful with the interpretation and use of results from evaluations of public policies.

However, we believe that, for lack of a better alternative, these evaluations are a good indication of the health of the compensatory educational programs, and they help to provide tracks and arguments when debates are engaged. They have the advantage, when they are finely analyzed, to be a tool for decision support for governments wishing to establish or improve their existing policies.

9. References

BAUMONT C., GUILLAIN R. (2013) Introduction. *Revue Économique*, 64(5), 765-774.

BEFFY M. and DAVEZIES L. (2013) Has the "ambition success" educational program achieved its ambition? *Annals of Economics and Statistics*, 111/112, 271-294

BENABOU R., KRAMARZ Y. and PROST C. (2009) The French zone d'éducation prioritaire: Much ado about nothing? *Economics of Education Review*, 28, 245-356

BERTRAND M., DUFLO E. and MULLAINATHAN S. (2004) How much should we trust difference-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249-275

BRIANT A., LAFOURCADE M., SCHMUTZ B. (2014). Can Tax Breaks Beat Geography? Lessons from the French Enterprise Zone Experience. *American Economic Journal: Economic Policy*, forthcoming.

CAILLE J.P., DAVEZIES L. and GARROUSTE M. (2013) Les réseaux ambition réussite. Une analyse en régression par discontinuité, *forthcoming*

CALVO-ARMENGOL A., PATACCHINI E. and ZENOU Y. (2009) Peer effects and social networks in education, *Review of Economic Studies*, 76, 1239-1267

CHIAPA C., GARRIDO J.L. and PRINA S. (2012) The effect of social programs and exposure to professionals on the educational aspirations of the poor, *Economics of Education Review*, 31, 778-798

CRAWFORD C. and VIGNOLES A. (2010) An analysis of the educational progress of children with special educational needs, DoQSS Working Paper No. 10-19

CREPON B. and JACQUEMET N. (2010) Économétrie : méthodes et applications, Ouvertures Économiques, De Boeck

CURRIE J. (2001) Early childhood education programs, *Journal of Economic Perspectives*, 15(2), 213-238

DAVEZIES L. and GARROUSTE M. (2014) More harm than good? Sorting effects in a compensatory education program, Presentation at PUCA seminar, 20th of march

DfEE (1999) *Excellence in Cities* (Issue 1) [online]. Available: http://www.standards.dfes.gov.uk/local/excellence/nlett1/news1.html [24 June,2005].

GIVORD P. (2010) Méthodes économétriques pour l'évaluation de politiques publiques, WPD3E n°G2010-08

GLAESER E.L., GOTTLIEB J.D. (2008) The economics of place-making policies (No. w14373). National Bureau of Economic Research.

HANUSHEK E., KAIN J. and RIVKIN S. (2002) Inferring program effects for special populations: does special education raise achievement for students with disabilities? *The Review of Economics and Statistics*, 84(4), 584-599

IMBENS G. and LEMIEUX T. (2008) Regression discontinuity designs: A guide to practice, *Journal of Econometrics*, 142(2), 615-635

JACOB B. and LEFGREN L. (2004) Remedial education and student achievement. A regression-discontinuity analysis, *The Review of Economics and Statistics*, 86(1), 226-244

KAIN J. (1968) Housing segregation, negro employment, and metropolitan decentralization, *The Quarterly Journal of Economics*, 82(2), 175-197

KESLAIR F. and MC NALLY S. (2009) Special Educational Needs in England. Final report for the National Equality Panel, Centre for Economic Performance, London School of Economics

KESLAIR F., MAURIN E. and McNALLY S. (2012) Every child matters? An evaluation of "Special Educational Needs" programmes in England, *Economics of Education Review*, 31, 932-948

LAVY V. and SCHLOSSER A. (2005) Targeted remedial education for underperforming teenagers: Costs and benefits, *Journal of Labor Economics*, 23(4), 839-874

MACHIN S., McNALLY S. and MEGHIR C. (2004) Improving pupil performance in English secondary school: Excellence in Cities, *Journal of the European Economic Association*, 2(2-3), 396-405

MACHIN S., McNALLY S. and MEGHIR C. (2010) Resources and standards in urban schoos, *Journal of Human Capital*, 4(4), 365-393

MAYNERIS F., PY L. (2013). The efficiency of enterprise zone programs: some conflicting results?, *Région et Développement*, 38, 209-224.

MINISTERE DE L'EDUCATION NATIONALE (2013) Évaluation de la politique de l'éducation prioritaire, Rapport de diagnostic

OECD (2014) PISA 2012 Results in Focus, What 15-year-old know and what they can do with what they know, [online]. Available: http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

PATACCHINI E. and ZENOU Y. (2012) Juvenile delinquency and conformism, *Journal of Law, Economics and Organization*, 28, 1-31

RUBIN B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, 66, 688-701

VAN DER KLAAUW W. (2008) Breaking the link between poverty and low student achievement: An evaluation of Title I, *Journal of Econometrics*, 142, 731-756