Evaluating the “Programa Mais Sucesso Escolar”: Lessons learned from evaluating the impact of a Portuguese national educational policy for compulsory education.

M. Clara Barata
Ed.D
ISCTE-Lisbon University Institute, CIS-IUL
M. Manuela Calheiros
Ph.D
ISCTE-Lisbon University Institute, CIS-IUL
Joana Nunes Patrício
M.A. in Psychology
ISCTE-Lisbon University Institute, CIS-IUL
João Graça
M.A. in Psychology
ISCTE-Lisbon University Institute, CIS-IUL
M. Luísa Lima
Ph.D
ISCTE-Lisbon University Institute, CIS-IUL

Abstract

Portugal is in a climate of increasing economic austerity. Student retention is common in Portugal, although it is a very expensive educational strategy, which may not have positive benefits on learning. Our paper provides evidence regarding the impact of the Portuguese national policy “Programa Mais Sucesso Escolar” (PMSE) created in 2009 to increase school achievement and reduce student retention. Our goal was to design an evaluation that would provide the necessary empirical evidence for policy-makers to make informed choices regarding the program. To do so, we evaluated the first two years of impact of PMSE on a varied list of indicators, using multilevel modeling and an empirically-matched control group of schools that applied for the program and did not get it. We found that PMSE schools had higher percentages of transitions and higher success in Mathematics, Portuguese and English in the first and second years of the program, with effect sizes varying between -0.17 and 0.59. We also observed negative effects of PMSE namely in students’ individual performance on the ninth-grade high-stakes exam in both years of implementation (ES varying between -0.33 and 0.23). We found no significant effects on school-level success measures in sixth-grade exams and ninth-grade high-stakes exams, average age of graduation at the end of each cycle, and recourse to alternative education paths. In the discussion we analyze the implications of these results for policy, practice and program evaluation research.

Introduction

While the most recent PISA results place Portugal within the OECD average on reading, and progressing rapidly toward the average in mathematics and science (Organisation for Economic Cooperation and Development 2010), Portugal still has one of the highest retention rates in compulsory school among OECD countries (Organisation for Economic Cooperation and Development 2011). This "culture of grade retention" (Organisation for Economic Cooperation and Development 2011, 60) has prevailed despite the lack of scientific evidence that supports the efficacy of retention as a pedagogical strategy (e.g. Jacob and Lefgren 2004; Jimerson 2001), and in spite of changes in national policy and regulation to minimize it (e.g. Lei n.º 46/86, D.R. nº 237, Série I, de 1986-10-14). In the Portuguese case, changes in culture required alternative approaches to managing children's learning difficulties, while promoting achievement for all students.

To facilitate changes in the retention culture, the Ministry of Education (ME) designed two important policies. First, in 2005, the ME legislated that teachers create an individualized plan of recuperation, supervision and development for all students failing their grade (Despacho Normativo nº 50/2005, D.R. nº 215, Série I, de 2005-11-09). Second, in 2009 and within the scope of this new law, the ministry launched PMSE, a school-based four-year initiative to reduce retention in compulsory education.

At the core of the program were a set of organizational strategies affecting class size, class composition, and teacher professional development to increase differentiated instruction. Despite mixed evidence of its efficacy (Dobbelsteen, Levin and Oosterbeek 2002; Finn and Achilles 1999; Hanushek 1999; Hoxby 2000; Krueger 1999), class size interventions are still one of the most common educational policies used internationally to increase educational achievement; differentiated instruction is one of the mechanisms to address class heterogeneity in achievement and maximize learning for all students (Heacox 2006). The program also aimed at maximizing compulsory school completing rates by investing in multi-year or cycle-level educational strategies[1].

In this paper we present the results of the first and second year of impact of PMSE on school success, success in high-stakes exams, cohort and cycle survival rates, and recourse to alternative education paths. Using multilevel modeling and an empirically-matched control group of schools that applied for the program and did not get it, we found that PMSE schools generally had significant positive effects on school success and significant negative effects on students' performance in the ninth-grade high-stake exam. We found no significant effects on school-level success measures in sixth and ninth-grade high-stakes exams, average age of graduation at the end of each cycle, and recourse to alternative education paths.

The application process

To participate in PMSE, schools had to apply to the ME by presenting a detailed plan of strategies to improve the outcomes of students likely to be retained. In this plan, schools also had to commit to lowering retention rates by one third each year, for four years. To make the plan feasible, the ME would pay for additional teaching time for the implementation of the organizational measures and would give schools autonomy to manage their resources each year. This increased autonomy was well-aligned with recommendations from the OECD for Portuguese schools in a generally centralized system (Organisation for Economic Cooperation and Development 2007). Finally, the ME also committed to providing a team of technical and academic experts to advise on aspects of implementation.

[1] In Portugal, at the time, there were 3 cycles of compulsory education: the first cycle included the 1st through 4th grade levels; the second cycle included de 5th and 6th grade levels; and the 3rd cycle included de 7th through 9th grade levels.

The offer from the ME was met with incredible enthusiasm from schools. The ME expected to enroll approximately 30 schools in the program, but received 375 applications to join the PMSE (Barata, Calheiros, Patrício, Graça and Lima 2012). Of the 375 schools that applied, only 123 were selected to join the program; the remainder constituted our control group. To select schools the ME examined the plans of recuperation and development for failing students presented by each school, and selected those with coherence and consistency. These plans contained innovative strategies to organize students in dynamic groups depending on achievement, coupled with differentiated instruction approaches. Most of the plans presented were inspired by two intervention approaches recently developed in Portuguese schools, the "TurmaMais" or MoreClassroom (Fialho and Salgueiro 2011; Verdasca 2008; Verdasca 2010; Verdasca and Cruz 2006), and the "Fénix" or Phoenix (Alves and Moreira 2011; Azevedo and Alves 2010; Moreira 2009). Sixty-seven and 46 schools were closely aligned with TurmaMais or Fénix, respectively. Another 10 schools presented consistent models of intervention that did not closely resemble either intervention model and were placed in a group called "Híbridas" or Hybrids (Direcção-Geral de Inovação e de Desenvolvimento Curricular 2010; Direcção-Geral de Inovação e de Desenvolvimento Curricular 2012). After being accepted into PMSE, all schools were directed to choose up to four courses and two grade levels in which to invest their credits.

The "Programa Mais Sucesso Escolar" (PMSE): Class size, class composition and differentiated instruction

Despite some diversity in the application of intervention approaches, PMSE schools shared a set of organizational strategies affecting class size and composition, and invested in teacher professional development to increase differentiated instruction.

Class size reduction is one of the most common policies used to address low academic achievement internationally (e.g. Hoxby 2000; Dobbelsteen, Levin and Oosterbeek 2002) and also one of the most popular policies among parents, teachers, teachers' unions and management teams. However, it is a very expensive educational investment (Hoxby 2000) because it often requires a considerable increase in the teaching time awarded to schools, and consequently the hiring of new teachers.

Class size reduction is often associated with class composition policies. After determining that a teacher will be given fewer students to teach, oftentimes a new criterion is used to reorganize students in classes. The criteria most often used to guide class composition include achievement level, gender, and race/ethnicity (this last one being more frequent in the United States, Hoxby 1998). These criteria can be used to maximize or minimize homogeneity in the classroom. In PMSE, the majority of schools employed extra teaching credits to divide students into smaller units, and used prior achievement level to create more homogeneous groups of students.

Because class size and class composition policies are such a costly investment, it is essential that we know the true impact of this policy. However, the class size policies are hard to study and the available research offers mostly contradictory results. In a summary of all empirical evidence available until 1994 including 277 estimates from 59 rigorous studies of the impact of class size on academic achievement, Hanushek (1999) reported that only 15% of all studies presented significant positive estimates of impact, while 13% of studies presented significant negative estimates; a pattern most likely to represent a null effect of the policy.

It is likely that some of the studies on class size were biased by self-selection. For example, it is likely that parents concerned with their children's achievement will choose schools with higher quality and lower class sizes. In this situation, when we compare the results of smaller and larger classes, we are simultaneously comparing the effect of class size and parental investment. In this example, the effect of the class size policy will be overestimated, i.e. the estimated effect will be larger than the effect of the true impact of class size. Schools also frequently create classes of different sizes, and use tracking mechanisms to place students with lower achievement in smaller classes and students with higher achievement in larger classes. In this case, it is likely that students in the smaller class will perform worse than students in the larger class, but the difference in performance will be much more an effect of class composition than class size; and the estimated effect of class size will be underestimated, i.e. the estimated effect will be smaller than the effect of the true impact of class size.

The best solution to problems of self-selection is to employ cluster randomized-control trials, i.e. randomly assign schools to a control or intervention group, and then evaluate their outcomes. That was the case of the Student/Teacher Achievement Ratio or STAR experiment (Finn and Achilles 1999; Hanushek 1999; Krueger 1999). In contrast to the previous results presented by Hanushek, the STAR experiment indicated that performance on standardized tests increased significantly for students in small classes compared to students in regular classes, particularly after just one year of the class size intervention (Krueger 1999). The experiment also indicated that the provision of a teacher aid had only a modest effect of student achievement (Krueger 1999). Finally, small classes had a larger effect for minority students and students receiving free lunch. For PMSE, these results indicated that class size reduction could impact student achievement in the absence of additional resources, such as teacher professional development directed toward small class instruction. Moreover, the results also indicated that larger effects could be expected after just one year of intervention.

Because randomized trials present considerable limitations, such as cost, ethical dilemmas (Diaz and Handa 2006) and generalization issues (Hoxby 1998), other studies have applied econometric techniques and take advantage of the natural variation in educational processes to achieve estimates of true impact of class size policies. Using regression discontinuity and instrumental variables, Hoxby (1998, 2000) demonstrated that natural class size reductions of 30 to 15 students did not have a significant impact on student performance in state tests. Hoxby (2000) attributed the lack of significant impact of class size to the fact that teachers were not equipped to deal with natural variations in class size. When arbitrarily given a smaller class, teachers were likely to continue using the same instructional strategies they employed for larger class sizes; and it was only under the atypical conditions of an experiment that such strategies made no difference.

A more recent study of fourth, sixth and eighth-grade Dutch children added to Hoxby's evidence by examining effects of class composition in addition to class size (Dobbelsteen, Levin, & Oosterbeek 2002). Dobbelsteen and colleagues (2002) initially found similar results to Hoxby, i.e. students in smaller classes did not present higher academic achievement than students in larger class. However, when they examined the changes in IQ class composition as an effect of the class size change, they found that more homogeneous IQ classes presented higher academic achievement; and furthermore, when accounting for class composition, students in smaller classes did perform better than students in larger classes.

For PMSE, this evidence indicated that the use of organizational strategies to reduce class size, associated with specific class composition strategies to maximize homogeneity could potentially lead to positive results in academic achievement. Moreover, PMSE's stakeholders effort to include professional teacher development in strategies to take advantage of these class size and class composition changes should help maximize those positive results.

In the majority of PMSE schools this meant teacher professional development directed at increasing differentiated instruction. Differentiated instruction generally consists of changes to teaching and learning routines in order to address the academic diversity usually found in classrooms (Heacox 2006; Morgado 2005). It involves a set of strategies to modify curricula, resources, teaching methods and learning tasks in a planned way so that the classroom environment is closer to the zone of proximal development of each student or group of students (Vygotsky 1978) and learning is optimal. In PMSE schools, professional development in differentiated instruction focused on many strategies, such as: approaches to student group work (such as peer tutoring); formative assessments and student evaluation focused on individual progress; curriculum management focused on interests of small groups; use of teaching materials of more concrete or abstract level based on level of small-group achievement; and flexibility in timing and opportunities for learning based on specific-group needs, among others (Barata et al. 2012).

Finally, because of PMSE's goal to maximize compulsory school completion for all students, PMSE schools focused on maximizing success for students not only at each grade level, but also at cycle completion. This involved a set of strategies to evaluate students' progress in cycles of grade levels, rather than yearly; and designing multi-year recuperation plans for failing students.

From Program Evaluation to Policy: Evaluating PMSE

In the past few decades there has been substantial development in the design of program evaluation research, as well as methods for statistical data analysis of program impact. In

Portugal, the ME has made considerable investments to gather reliable student data (Decreto-Lei nº 88/2007), and in 2007 created an office, Gabinete de Estatística e Planeamento da Educação (GEPE), whose mission was to produce and analyze the statistics of education, in order to inform educational policy and practice (Decreto-Regulamentar n.º 25/2007). Occasionally, the ME also collaborates with Universities and other experts to analyze the impact of public policies. The two models of intervention employed in the PMSE are examples of local initiatives brought to scale after careful evaluation by Universities (e.g., Verdasca 2006), with funding from external entities (e.g. Fundação Calouste Gulbenkian).

Despite considerable advances in the quality of the information available pertaining to education outcomes, most of the empirical evidence developed by the ME and the external consultants does not answer the question “What is the true impact (or cause-effect relationship) of an educational program in students’ achievement?” This empirical problem is due to the fact that the majority of the current evaluation studies of Portuguese policy in education use only qualitative data and small samples, which makes these causal inferences about the effect of the policy unlikely. The problem is further amplified by the lack of a control group, and the fact that participation in most educational programs depends on choices made by parents, teachers, legislators or other stakeholders. These choices make the participation in educational programs the product of a process of self-selection, instead of being randomly determined (Murnane and Willett 2010; Shadish, Campbell and Cook 2002). As such, the variation in educational programs is potentially correlated with other determinants of school success, such as teachers’ expectations or parents’ investment, producing biased results of the effects of the program.

This is the case for the initial study of the impact of PMSE, on educational achievement. The program was initially evaluated comparing each schools performance with their historical level of retention (i.e. the average retention of the same grade level in the four years prior to the beginning of the program). A report from a ministry office concluded that "Considering that 186 projects of PMSE schools demonstrated an average historical level of retention of 84.65%, per grade, in reference to the 4-year period of 2005/2009, at the end of the academic year there was a global increase of 7.46% in school achievement" (Direcção-Geral de Inovação e de Desenvolvimento Curricular 2010, 32). Unfortunately, this average reduction of 7.46% in student retention cannot be interpreted as the true impact of the program for many reasons; first, because schools received financial incentives to stay in the program by lowering retention rates; second, because without comparison to a set of control schools, this reduction may just reflect a national policy change in all Portuguese schools.

The best way to solve this empirical problem would be to randomly assign schools that applied to the program to PMSE or a control group, such as in the STAR experiment, and then evaluate their outcomes. Because this option is not applicable to programs at scale, and because of the aforementioned limitations of experiments, we chose to take advantage of the variation created by the selection process into PMSE, comparing schools chosen to enter the program to schools that applied but were not chosen to enter the program (i.e. control schools). Because these schools are likely to be different, we used Propensity Scores Estimation (PSE) to reduce the differences in educational achievement between intervention and control groups that are not attributable to PMSE (Rosenbaum and Rubin 1983). Diaz and Handa (2006) demonstrated that when used properly, PSE estimates can be reliable and can approach experimental estimates.

The present study

The goal of this study was to design an evaluation that would provide the necessary empirical evidence for policy-makers to make informed choices regarding the program. To do so we evaluated the first two years of impact of PMSE on a varied list of indicators, using multilevel modeling and an empirically-matched control group of schools that applied for the program and did not get it. Indicators included school success, success in high-stakes exams at the school and student level, cohort and cycle survival, and alternative education paths. In Portugal, indicators of school success regulate students' academic life, including students' year-to-year transitions and performance on specific disciplines. Because most PMSE schools invested their extra teaching credits in Mathematics, Portuguese and English (of all courses, approximately 32% were Portuguese, 29% were Mathematics and 23% were English in the first and second years), we examined the impact on these three courses only. We also included measures of success in the ninth-grade high-stakes exams for three reasons. First, the ninth-grade high-stakes exams complement the school's evaluation with a national standardized measure of performance. Second, performance on the ninth-grade high-stakes exams counts 30 to 100% toward the final evaluation in compulsory education. Third, data on ninth-grade high-stakes exams was available at the student level and therefore supplemented school-level performance data. Data on the sixth-grade exam offered an earlier standardized measure of performance, but was made available only at the school level. Cohort and cycle survival indicators were included given the program's focus on minimizing retention and maximizing the achievement of low-performing student who would otherwise not complete each cycle of compulsory education. Finally, data on alternative education paths allowed us to check whether the improvement on the remaining indicators was a function of moving the low-performing students out of the regular paths into alternative education paths that are not included in these indicators.

The following research question was addressed:

What is the impact of the first two years of PMSE on school success, success in high-stakes exams at the school and student level, cohort and cycle survival, and alternative education paths?

Methods

Study Sample

The study sample was comprised of 375 schools; data were collected on 123 intervention schools and 252 control schools in the first year (2009/2010), and on 115 intervention schools and 248 control schools in the second year (2010/2011). Intervention and control schools were compared on varied indicators pertaining to the year of application (2008/2009), namely: 1) at the district level, geographic distribution, and urban-rural qualification; 2) at the school level, inclusion in large administrative units, including in the same building preschool and/or secondary school; total number of students, total number of faculty, percentage of female faculty, average age of faculty, total number of staff, average years of teaching experience of faculty, number of faculty with masters or higher; and 3) at the student level, percentage of students receiving government support (SASE A, SASE B and SASE C), percentage of students with computer at home, percentage of students with internet at home, and number of school violence events. Intervention and control schools were significantly different on 7 out of 23 indicators in the first year; and 10 out of 23 indicators in the second year (Table 1). All indicators were used in the matching strategy outlined below.

Table 1. Demographic Data on Schools and School Districts in the 1st Year (First Panel), and in the 2nd Year (Second Panel) Of PMSE.

	Control Schools		Intervention Schools		Diff.	Sig.
	N	*Mean*	*N*	*Mean*		
	First Year of PMSE					
Geographic Distribution						
DREN	252	0.31	123	0.24	0.07	
DREC	252	0.23	123	0.17	0.06	
DRELVT	252	0.33	123	0.23	0.11	*
DREA	252	0.05	123	0.29	-0.24	***

DREALG	252	0.07	123	0.07	0.01	
Urban-Rural qualification						
Predominantly rural district	252	0.08	123	0.07	0.01	
Moderately urban district	252	0.20	123	0.28	-0.07	
Predominantly urban district	252	0.72	123	0.65	0.07	
Schools in large administrative unit	252	0.83	123	0.74	0.09	*
Includes preschool	252	0.04	123	0.07	-0.03	
Includes secondary school	252	0.37	123	0.56	-0.19	**
Total # of students	252	699.57	123	663.88	35.69	
Total # of faculty	252	89.30	123	85.94	3.36	
Percentage female faculty	252	73.72	123	72.74	0.98	
Average age of faculty	252	42.10	123	41.42	0.69	*
Total # of staff	252	33.71	123	34.68	-0.96	
Average teaching experience of faculty	252	17.46	123	16.71	0.75	*
# of faculty with masters or higher	252	4.56	123	4.39	0.17	
Percentage students receiving government support						
SASE A	252	26.13	123	24.72	1.41	
SASE B	252	18.31	123	16.36	1.95	*
SASE C	252	2.01	123	1.89	0.12	
Percentage students with computer at home	252	57.89	123	57.93	-0.04	
Percentage students with internet at home	252	38.16	123	38.70	-0.54	
# of school violence events	251	3.79	122	2.12	1.67	
Second Year of PMSE						
Geographic Distribution						
DREN	248	0.32	115	0.25	0.07	
DREC	248	0.23	115	0.16	0.08	~
DRELVT	248	0.34	115	0.22	0.12	*
DREA	248	0.05	115	0.29	-0.24	***
DREALG	248	0.06	115	0.09	-0.03	
Urban-Rural qualification						
Predominantly rural district	245	0.08	113	0.08	0.00	
Moderately urban school district area	245	0.20	113	0.28	-0.08	~
Predominantly urban school district area	245	0.72	113	0.64	0.09	
Schools in large administrative unit	248	0.83	115	0.75	0.08	~
Includes preschool	248	0.04	115	0.08	-0.03	
Includes secondary school	248	0.38	115	0.53	-0.15	**
Total # of students	248	701.78	115	655.62	46.17	
Total # of faculty	248	89.49	115	84.51	4.98	
Percentage female faculty	248	73.74	115	72.75	0.99	
Average age of faculty	248	42.13	115	41.45	0.68	*
Total # of staff	248	33.72	115	34.49	-0.77	
Average teaching experience of faculty	248	17.47	115	16.71	0.76	*
# of faculty with masters or higher	248	4.57	115	4.59	-0.02	
Percentage students receiving government support						
SASE A	248	26.08	115	24.64	1.44	
SASE B	248	18.35	115	16.51	1.84	*
SASE C	248	2.01	115	1.86	0.16	
Percentage students with computer at home	248	57.70	115	58.17	-0.47	
Percentage students with internet at home	248	37.92	115	39.20	-1.28	
# of school violence events	248	3.84	113	1.59	2.24	*

Note. ~ p<.10, * p<.05, ** p<.01, *** p<.001.

Measures and Data Sources

Educational achievement

Data on educational achievement for the years 2008/09 through 2010/2011 was obtained to create a varied list of indicators measuring school success, success in high-stakes exams at the school and student level, cohort and cycle survival, and alternative education paths.

School success. Students are evaluated on a scale of 1 (worst performance) to 5 (best performance) in the 5^{th} to 9^{th} grade in compulsory education. A performance lower than 3 is equivalent to failing a course or grade level. Using this scale, indicators for school success included percentage of grade-level transitions (i.e. number of students who transitioned per year divided by number students enrolled in same year, by school); and percentage success in Mathematics, Portuguese, and English (i.e. number of students with final performance of 3, 4 or 5 in the respective discipline, divided by number of students enrolled in the discipline, by year and by school). To quantify the impact on high-performance school success, we also included percentage of fours and fives in Mathematics, Portuguese, and English (i.e. number of students with final performance of 4 or 5 in the respective discipline, divided by number of students enrolled in the discipline, by year and by school).

Success in high-stakes exams, school level. Students are evaluated on a scale of 1 (worst performance) to 5 (best performance) in the 6^{th} and ninth-grade high-stakes exams. A performance lower than 3 is equivalent to failing the high-stakes exams. Using this scale, indicators for success in the high-stakes ninth grade exam at the school level included: percentage of exams per school (i.e. total number of ninth-grade high-stakes exams, divided by number of students enrolled in ninth-grade level, per school), percentage of success in the Mathematics and Portuguese exam (i.e. number of students with performance of 3 or above in the ninth-grade high-stakes exams of the respective discipline, divided by number of students enrolled in ninth-grade level, by school), and percentage of fours and fives in the Mathematics and Portuguese exam (i.e. number of students with performance of 4 or 5 in the ninth-grade high-stakes exams of the respective discipline, divided by number of students enrolled in 9^{th} grade level, by school). Indicators for success in the sixth-grade high-stakes exam at the school level included: percentage of success in Mathematics and Portuguese exam (i.e. number of students who transitioned in sixth-grade exams of the respective discipline, divided by number of students enrolled in 6^{th} grade level, by school); and percentage of fours and fives in the Mathematics and Portuguese exam (i.e. number of students with final performance of 4 or 5 in the sixth-grade exams of the respective discipline, divided by number of students enrolled in 6^{th} grade level, by school).

Success in high-stakes exams, student level. Indicators for success in high-stakes exams at the student level included: performance in the Mathematics and Portuguese ninth-grade exam, per student (i.e. performance in the Mathematics and Portuguese ninth-grade exam); and difference between school performance and high-stakes performance in Mathematics and Portuguese (i.e. school performance minus ninth-grade high-stakes exam performance, of the respective discipline), per student.

Cohort and cycle survival. To measure cohort and cycle survival, we included cohort survival rates on the 2^{nd} and 3^{rd} cycle (i.e. number of students who transitioned in the 6^{th} and 9^{th} grade level, divided by number of students enrolled in 5^{th} and in the 7^{th} grade level, two and three years before, respectively, by school), and average age of graduation at the end of each cycle (i.e., average age of the students in the 4th, 6th and 9th grade levels, minus average age of the students on enrollment in the same cycle, four, two or three years before, respectively, minus expected number of years in the cycle, as in four, two or three years, respectively). This last indicator takes on negative values when students complete a cycle before the normal time; takes a value close to zero when students complete the cycle in regular time, and takes positive values when students are retained and complete the cycle after the regular time.

Alternative education paths. To measure the program's impact on the recourse to alternative education paths, we included percentage of enrollments in alternative education and training courses on the 2nd and 3rd cycle (i.e. number of students enrolled in alternative education and training courses on the 2nd and 3rd cycle, divided by total number of students enrolled on the 2nd and 3rd cycle, by school).

School demographics

Demographic data on schools and school districts for the year 2008/09 was also provided by the ME and included all district, school and student-level indicators described in the Study Sample section.

Procedures

In collaboration with the ME, and after extensive interviews with PMSE stakeholders at the ministry, regional, school district, and school level, the research team defined a list of indicators necessary to measure the impact of PMSE and to empirically match intervention and control schools. All indicators, except those pertaining to the high-stakes ninth grade exam, were provided by the ME at the school level in Excel format, between 17 October 2011 and 13 March 2012. Indicators pertaining to the high-stakes ninth grade exam were downloaded from the *Júri Nacional de Exames* website at the student level in Excel format, and then school-level indicators were created when necessary. The databases were cleaned, stored in files .dta and then analyzed in Stata (Version 12).

Data Analytic Strategy

After deciding to compare schools chosen to enter the program to schools that applied but were not chosen to enter the program, the team used several data analytic strategies designed to utilize all data available, and also reduce bias. First, because PMSE schools were directed by the ME to choose up to four courses and two grade levels in which to invest their credits, each PMSE school contributed with a predetermined set of data units for the analysis of impact. In contrast, control schools could potentially contribute with information from all grade levels available. To make the two groups more equivalent, data on intervention schools was restricted to the cycles and grade levels in which the PMSE was implemented. In the control schools, the cycles and grade levels were determined randomly based on the percentage of cycles and grade levels available in the intervention schools in the first year (approximately 2% at 1^{st} grade, 8% at 2^{nd} grade, 2% at 3^{rd} grade, 1% at 4^{th} grade, 18% at 5^{th} grade, 5% at 6^{th} grade, 48% at 7^{th} grade, 8% at 8^{th} grade and 7% at 9^{th} grade).

Second, because the implementation design involved giving schools the power to decide in which (of up to four) courses and which (of up to two) grade levels the implementation would take place, each PMSE school contributed with more than one datapoint for most of the indicators. For example, to study the impact of PMSE on the percentage of grade-level transitions, each intervention (or control) school could contribute with up to two percentages for each grade level in the program. Therefore, all analyzes with multiple data points were adjusted according to the contributions of each grade level and course in the program by including multilevel models with random effects for schools. In the control schools, the analyses were adjusted according to the contributions of each grade level randomly determined through the process described above. A similar process was used for the student-level data available for success on the ninth-grade high-stakes exam. We used multivariate analysis to compare PMSE and control schools on educational achievement when only one point of data was available per school for a particular indicator.

Third, because schools chosen to enter PMSE are likely to be different from schools not chosen to enter the program (i.e. control schools) we used Propensity Scores Estimation (PSE) to reduce the differences in educational achievement between intervention and control groups that are not attributable to PMSE (Rosenbaum and Rubin 1983). Diaz and Handa (2006) demonstrated that when used properly, PSE estimates can be reliable and can approach experimental estimates (ex. High-stakes exams indicators).

We implemented a version of PSE, *inverse propensity score weighting* (Murnane and Willett 2010), by using the *pscore* procedure in Stata 12. Pscore estimated the propensity score of the treatment (i.e. PMSE) on a list of variables using a logit model and stratified schools in blocks according to the pscore. The list included all demographic data on schools and school districts for the year prior to the first year of implementation of PMSE (2008/09) described above. Then the procedure checked if PMSE and control blocks were balanced (i.e. balancing property). When the balancing property was satisfied, the pscores were inverted for the intervention group, and subtracted from 1 and inverted for the control group. These new pscores were used as analytic weights (*aweights* in Stata) to adjust the estimates from the impact models.

Finally, we also adjusted all estimates from the multivariate and multilevel impact models for school and school district characteristics, as well as baseline school averages.

In the results section below we discuss the results of the final model for each indicator, in most cases a multilevel model, adjusted for *aweights*, school and school district characteristics, as well as baseline school averages. In the tables we present the adjusted means of the control and intervention schools, the estimated difference between these groups, the level of significance of this difference, and the effect size based on these final models. The effect size indicates the magnitude of the PMSE effect regardless of the instrument or method used. It was computed by dividing the estimated adjusted difference between groups by the standard deviation of the indicator for the control group (Gormley, Gayer, Phillips and Dawson 2005; Wong, Cook, Barnett and Jung 2008).

Results

Impact on school success

In Table 2 we present the impact results on school success in the 1st and 2nd years of PMSE for the indicators percentage of grade-level transitions, and percentage of success in Mathematics, Portuguese, and English. Accounting for the multilevel structure of the data, and adjusting for the *pscore* weights, school and school district characteristics, in the first year of implementation, the intervention schools had significantly higher percentages of grade-level transitions from the second to the ninth grade level than the control schools, in 4.35% (s.e. = 0.69, $p<.001$). This difference was equal to 5.22% (s.e. = 1.27, $p <.001$) in the percentage of success in Mathematics, 6.04% (s.e. = 0.9, $p <.001$) in Portuguese, and 4.71% (s.e. = 1.17, $p <.001$) in English.

In the second year of implementation, the intervention schools had a significantly lower (at trend level) percentage of grade-level transitions from the second to the ninth grade level than the control schools, in 0.88% (s.e. = 0.47, $p<.1$); and a significantly higher percentage of success than the control schools in the following outcomes: in Mathematics equal to 7.06% (s.e. = 1.45, $p<.001$); in Portuguese to 5.29% (s.e. = 0.88, $p<.001$); and in English to 6.46% (s.e. = 1.29, $p<.001$).

Table 2. Impact on School Success in the 1st and 2nd Years of PMSE.

Variables	*N* Grade-levels	*N* Schools	Adjusted mean of control schools	Adjusted mean of intervention schools	Diff.	Sig.	Effect size
Percentage of grade-level transitions							
1st year	464	359	86.66	91.01	4.35	***	0.48
2nd year	459	345	89.84	88.96	-0.88	~	-0.17
Percentage of success in Mathematics							
1st year	396	308	72.77	77.98	5.22	***	0.44
2nd year	385	296	70.17	77.23	7.06	***	0.59
Percentage of success in Portuguese							
1st year	401	314	81.85	87.89	6.04	***	0.56
2nd year	394	305	84.36	89.64	5.29	***	0.54
Percentage of success in English							
1st year	369	289	79.51	84.22	4.71	***	0.44
2nd year	362	278	80.37	86.83	6.46	***	0.59

Note. ~ p<.10, * p<.05, ** p<.01, *** p<.001. In all the indicators a positive difference between the intervention and the control schools and a positive effect size indicates a positive impact of PMSE.

In Table 3 we present the impact results on school success in the 1st and 2nd years of PMSE for the indicators percentage of fours and fives in Mathematics, Portuguese, and English. In the first year of the program, the intervention schools had a significantly higher percentage of fours and fives in Portuguese, in 3.72% (s.e. = 1.23, p <.01), and had a marginally higher percentage of fours and fives in English, in 2.37% (s.e. = 1.37, p <.1), than the control schools. There was no significant effect on the percentage of fours and fives in Mathematics in the first year.

In the second year of the program, the intervention schools had significantly higher percentage of fours and fives in Mathematics, in 2.91% (s.e. = 1.16, p <.05), and in Portuguese, in 3.98% (s.e. = 1.37, p<.01), than the control schools. There was no significant effect on the percentage of fours and fives in English in the second year.

Table 3. Impact on High-Performance School Success for the 1st and 2nd Years of PMSE.

Variables	*N* Grade-levels	*N* Schools	Adjusted mean of control schools	Adjusted mean of intervention schools	Diff.	Sig.	Effect size
Percentage of fours and fives in Mathematics							
1st year	396	308	28.32	29.80	1.49		0.14
2nd year	385	296	25.65	28.56	2.91	*	0.28

Percentage of fours and fives in Portuguese							
1st year	401	314	27.44	31.16	3.72	**	0.32
2nd year	394	305	27.81	31.79	3.98	**	0.35
Percentage of fours and fives in English							
1st year	369	289	33.69	36.06	2.37	~	0.21
2nd year	362	278	32.77	34.67	1.90		0.16

Note. ~ p<.10, * p<.05, ** p<.01, *** p<.001. In all the indicators a positive difference between the intervention and the control schools and a positive effect size indicates a positive impact of PMSE.

Impact on High-Stakes Exams, School Level

In Table 4, Panel A, we present the impact results on the ninth grade high-stakes exams (school-level outcomes) for the indicators percentage of exams per school, percentage of success in the Mathematics and Portuguese exam, and percentage of fours and fives in the Mathematics and Portuguese exam. In Table 4, Panel B, we present the impact results on the sixth grade high-stakes exams (school-level outcomes) for the indicators percentage of success in Mathematics and Portuguese exam; and percentage of fours and fives in the Mathematics and Portuguese exam.

As can be seen in Table 4, there is no statistical evidence of a significant impact of PMSE in the ninth-grade high-stakes exams at the school level, in the first or in the second year of PMSE.

Table 4. Impact on Ninth and Sixth-Grade Exams (School Level Outcomes) in the 1st and 2nd Years of PMSE.

Variables	*N* Schools	Adjusted mean of control schools	Adjusted mean of intervention schools	Diff.	Sig.	Effect size
	9th grade					
Percentage of exams per school						
1st year	33	80.61	87.83	7.22		0.51
2nd year	35	82.26	86.51	4.25		0.42
Percentage of success in the Mathematics exam						
1st year	33	48.07	50.10	2.03		0.10
2nd year	35	36.59	37.11	0.52		0.03
Percentage of success in the Portuguese exam						
1st year	33	72.86	61.81	-11.05		-0.79
2nd year	35	59.50	46.62	-12.88		-0.87

Percentage of fours and fives in the Mathematics exam						
1st year	33	22.77	21.57	-1.20		-0.08
2nd year	35	14.66	15.49	0.83		0.08
Percentage of fours and fives in the Portuguese exam						
1st year	33	18.60	20.11	1.51		0.10
2nd year	35	18.53	7.46	-11.07		-1.37
	6th grade					
Percentage of success in the Mathematics exam						
1st year	25	76.74	72.94	-3.80		-0.30
2nd year	70	60.24	64.90	4.66	~	0.30
Percentage of success in the Portuguese exam						
1st year	25	84.81	91.08	6.28		0.61
2nd year	71	81.11	81.29	0.18		0.02
Percentage of fours and fives in the Mathematics exam						
1st year	25	24.61	31.81	7.20		0.65
2nd year	70	28.30	31.58	3.28		0.27
Percentage of fours and fives in the Portuguese exam						
1st year	25	26.21	27.35	1.14		0.10
2nd year	71	39.23	39.33	0.09		0.01

Note. ~ $p<.10$, * $p<.05$, ** $p<.01$, *** $p<.001$. In all the indicators at the school level, a positive difference between the intervention and the control schools and a positive effect size indicates a positive impact of PMSE.

In the sixth-grade exams, there is no evidence of a significant impact of PMSE, in the first year of PMSE. However, in the second year of PMSE, the intervention schools had marginally higher percentages of success in the mathematics exam than the control schools, equal to 4.66% (s.e. = 2.75, $p < .1$, see Table 4 Panel B).

Impact on High-Stakes Exams, Individual Level

In Table 5, we present the impact results on the ninth-grade high-stakes exams (student-level outcomes) for the indicators performance in the Mathematics and Portuguese ninth-grade exam, per student; and difference between school performance and high-stakes performance in Mathematics and Portuguese, per student.

There is a negative and significant impact of PMSE in the ninth-grade high-stakes exam at the individual level. In the first year of PMSE, the students of the intervention schools had a marginally lower performance in the Mathematics exam than the students of the control schools, in 0.22 points (s.e. = 0.12, $p < .01$); and in the second year of PMSE the students of the intervention schools had a significantly lower performance in the Portuguese exam than the students of the control schools in 0.19 points (s.e. = 0.09, $p <.05$).

For the indicator of difference between school performance and performance in the high-stakes exam, students of the intervention schools had a significantly higher difference in this grades (i.e., had higher school performance than performance in the exam) than the students of the control schools in Mathematics, in the first year, at 0.20 (s.e. = 0.1, $p <.05$), and in Portuguese, in the first year at 0.15 points (s.e. = 0.07, $p <.05$), and in the second year at 0.22 points (s.e. = 0.13, $p <.1$). Conversely, the students of the intervention schools had a significantly lower difference in performance than the students of the control schools in Mathematics, in the second year of PMSE at 0.16 points (s.e.= 0.08, $p <.05$).

In the performance in the Mathematics exam, there is no evidence of a significant impact of PMSE, in the second year of PMSE; and in the performance in the Portuguese exam, there is no evidence of a significant impact of PMSE, in the first year of PMSE.

Table 5. Impact on Ninth-Grade High-Stakes Exam at the Student Level, in the 1st and 2nd Years of PMSE.

Variables	*N* Students	*N* Schools	Adjusted mean of control schools	Adjusted mean of intervention schools	Diff.	Sig.	Effect size
Performance in the Mathematics exam							
1st year	1955	33	2.75	2.53	-0.22	~	-0.22
2nd year	2570	35	2.32	2.41	0.09		0.09
Performance in the Portuguese exam							
1st year	1942	33	2.97	2.83	-0.14		-0.18
2nd year	2605	35	2.77	2.58	-0.19	*	-0.25
Difference between school performance and high-stakes performance in Mathematics							
1st year	1955	33	0.41	0.62	0.20	*	0.30
2nd year	2570	35	0.71	0.55	-0.16	*	-0.24
Difference between school performance and high-stakes performance in Portuguese							
1st year	1942	33	0.29	0.44	0.15	*	0.22
2nd year	2605	35	0.49	0.71	0.22	~	0.33

Note. ~ p<.10, * p<.05, ** p<.01, *** p<.001. In all the indicators at the individual level, a positive difference between the intervention and the control schools and a positive effect size indicates a positive impact of PMSE, except in the Difference between school performance and high-stakes performance in which a negative difference between the intervention and the control schools and a negative effect size indicates a positive impact of PMSE.

Impact on Cohort and Cycle Survival

In Table 6, we present the impact results on cohort and cycle survival in the 1st and 2nd years of PMSE for the indicators cohort survival rates on the 2nd and 3rd cycles, and average age of graduation at the end of each cycle. In the first year of the program, the intervention schools had marginally lower cohort survival rates in the 2nd and 3rd cycles than the control schools, in 8.05% (s.e. = 4.16, p <.05). There was no statistically significant impact of PMSE in the Cohort Survival Rates in the second year of PMSE or in the Average Age of Graduation in both years.

Table 6. Impact on Cohort and Cycle Survival in the 1st and 2nd Years of PMSE.

Variables	*N* Grade-levels	*N* Schools	Adjusted mean of control schools	Adjusted mean of intervention schools	Diff.	Sig.	Effect size
Average age of graduation at the end of each cycle							
1st year	61	60	0.04	0.05	0.01		0.03
2nd year	118	117	0.09	0.12	0.04		0.17
Cohort survival rates on the 2nd and 3rd cycles							
1st year	58	57	83.47	75.42	-8.05	~	-0.44
2nd year	106	106	85.72	84.91	-0.82		-0.04

Note. ~ p<.10, * p<.05, ** p<.01, *** p<.001. In the Cohort Survival Rates a positive difference between the intervention and the control schools and a positive effect size indicates a positive impact of PMSE. In the Average Age of Graduation higher and positive values indicate a higher retention rate, i.e., a positive difference between the intervention and the control schools and a positive effect size indicates a negative impact of PMSE.

Impact on Alternative Education Paths

In Table 7, we present the impact results on recourse to alternative education paths for the indicator percentage of enrollments in alternative education and training courses on the 2nd and 3rd cycle. There is no statistical evidence of a significant impact of PMSE in the percentage of enrollments in alternative education and training courses per cycle, in both years of the program implementation.

Table 7. Impact on Alternative Education Paths in the 1st and 2nd Years of PMSE.

Variables	N Grade levels	N Schools	Adjusted mean of control schools	Adjusted mean of intervention schools	Diff.	Sig.	Effect size
Percentage of enrollments in alternative education and training courses per cycle							
1st year	421	329	7.32	7.67	0.35		0.05
2nd year	257	189	9.23	8.19	-1.04		-0.16

Note. ~ p<.10, * p<.05, ** p<.01, *** p<.001. In this indicator higher values indicate a higher percentage of enrollments in alternative education and training courses per cycle, so a negative difference between the intervention and the control schools and a negative effect size indicates a positive impact of PMSE.

Discussion

The goal of this study was to evaluate the impact of the first two years of PMSE on a varied list of educational achievement indicators, using multilevel modeling and an empirically-matched control group of schools that applied for the program and did not get it. The results show that the PMSE had a positive and significant effect on educational achievement, especially on measures of school success. Specifically, we found that PMSE schools generally presented higher percentages of grade-level transitions, and higher percentages of success in Mathematics, Portuguese and English than control schools, and observed effect sizes were moderate to large (significant ES between -.17 and 0.59; McCartney and Rosenthal 2000). The results also indicated that PMSE had a significant positive impact on high-performance school success (fours and fives) in these three courses, and that effect sizes in this domain were medium (significant ES between 0.20 and 0.35). We can then conclude that the main goal of PMSE to increase school achievement and reduce student retention was achieved.

However, while our study indicated that PMSE generally had a significant positive impact in measures of school success, there was a non-significant to negative impact on the ninth-grade high-stakes Mathematics and Portuguese exams. Specifically, we found no evidence of a significant impact of PMSE in the percentage of exams per school, percentage of success in the Mathematics and Portuguese exam, and percentage of fours and fives in the Mathematics and Portuguese ninth-grade high-stakes exams at the school level, in the first or in the second year of

PMSE. We also found that students of the intervention schools tended to demonstrate lower performance in the exams, and there was a significantly higher difference between school performance and performance in the exams for students in the intervention schools when compared to students in the control schools (significant ES between -.33 and .23).

We hypothesize that in an attempt to lower retention and maximize cohort survival rates, PMSE schools seem to be moving a larger number of students with lower achievement through compulsory school, and getting them to final exams. This strategy has the undesirable effect of lowering PMSE schools' performance on exams. In fact a higher (though not yet statistically significant) number of students in PMSE schools are taking the exam (7% in the first year, and 4% in the second year), when compared to control schools. The average national percentage of exams per school in 2009/2010 was 84%, while in PMSE schools it was 87% in 2009/2010 and 86% in 2010/2011. It is also important to note that the performance of PMSE students on the high-stakes exams, though lower than control schools, it was not very distant from national averages. In fact, the average national performance on the exams in 2009/2010 was 2.70% and 2.72%, while in PMSE schools it was 2.53% and 2.58%, respectively in Mathematics and Portuguese. Finally, these estimates were based on a much smaller sample than the analysis of impact on school success, and need to be interpreted with caution.

Nevertheless, these negative results on high-stakes exams may make PMSE schools vulnerable to accountability policies based solely on standardized national evaluations, as well as attacks by the media based on national school rankings based on exams. This reflection demands the design of new indicators of school efficiency, which combine results on external evaluations with other indicators, such as cohort survival rates, in order to reinforce the effort of public schools to bring more students to compulsory school completion.

In contrast to the effects on the ninth-grade exams, we found a positive trend in the sixth-grade high-stakes Mathematics exam, possibly indicating that an intervention in the earlier grade levels may be more efficient. We also found a trend toward marginally lower cohort survival rates for intervention schools in the first year of PMSE and no effects on average age of graduation, which was unexpected given the program's strong philosophy to have each student complete compulsory education. However, these results need to be confirmed with additional data because of limited sample size. Finally, we saw no effects on recourse to alternative Education and Training Courses.

Impact on Policy and Practice

In terms of our ultimate goal of influencing policy-making in Portuguese education, presentations of results to ME and PMSE representatives, as well as the general public, in April through June 2012 have been well-received, and the ministry has so far allowed PMSE to continue in schools as predicted in 2009. These presentations have also allowed us to reflect with stakeholders regarding the unexpected results of the impact evaluation, namely the impacts on high-stakes exams and cohort survival rates.

We have also proceeded to share the results with teachers and principals in study schools in an attempt to directly influence practice, but the impact of these presentations is not yet clear. In a complete evaluation of the implementation of PMSE in the first two years comparing intervention and control schools, we found that class size, class composition and differentiated instruction strategies were the most distinctive features of PMSE schools (Barata et al. 2012). We hope that PMSE stakeholders and teachers continue to invest in professional development in innovative differentiated instruction practices, which seem to be the key to the success of class size policies (Hoxby 2000), and possibly one important mechanism to the positive impact of

PMSE on school success. We also hope PMSE schools and students can focus their attention on the high-stakes exams, so as to not expose the program to undeserved criticism.

Finally, regarding the field of policy evaluation in Portugal, we hope to have contributed to the national discussion of what constitutes appropriate evidence of the true impact of an educational program in students' achievement, by examining the impact of PMSE on a varied list of indicators, and using rigorous methods. We hope that new opportunities will come forth to examine the full impact of PMSE after four years of implementation; and design randomized trials that will improve causal inferences about current and future policies and facilitate the decisions of policy-makers regarding the Portuguese education system.

References

Alves, José, and Luísa Moreira. 2011. *Projecto Fénix – Relatos que contam o sucesso.* Porto: Edição Faculdade de Educação e Psicologia da Universidade Católica Portuguesa.

Azevedo, Joaquim, and José Alves. 2010. *Projecto Fénix - Mais Sucesso Para Todos - Memórias e dinâmicas da construção do sucesso escolar*. Porto: Faculdade de Educação e Psicologia.

Barata, Maria C., Maria M. Calheiros, Joana N. Patrício, João Graça and Luísa Lima. 2012. *Avaliação do Programa Mais Sucesso Escolar.* Lisboa: Direção-Geral de Estatísticas de Educação e Ciência – Ministério da Educação e Ciência.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences.* Hillsdale NJ: Lawrence Erlbaum.

Decreto-Lei nº 88/2007

Decreto-Regulamentar n.º 25/2007

Despacho Normativo nº 50/2005, D.R. nº 215, Série I, de 2005-11-09

Diaz, Juan, and Sudhanshu Handa. 2006. An assessment of propensity score matching as a nonexperimental impact estimator. Evidence from Mexico´s PROGRESA Program. *The Journal of Human Resources,* XLI: 319-345.

Direcção-Geral de Inovação e de Desenvolvimento Curricular. 2010. *Programa Mais Sucesso Escolar: Relatório final*. http://www.dgidc.minedu.pt/outrosprojetos/index.php?s= directorio&pid=122#i

Direcção-Geral de Inovação e de Desenvolvimento Curricular. 2012. *Relatório de avaliação anual do Programa Mais Sucesso Escolar 2010/2011.* Relatório não publicado.

Dobbelsteen, Simone, Jesse Levin and Hessel Oosterbeek. 2002. The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economics and Statistics*, 64: 17-38.

Fialho, Isabel, and Hélio Salgueiro. 2011. *TurmaMais e sucesso escolar: Contributos teóricos e práticos.* Évora: Universidade de Évora.

Finn, Jeremy, and Charles Achilles. 1999. Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis,* 21: 97-109.

Gormley, William, Ted Gayer, Deborah Phillips and Brittany Dawson. 2005.The effects of universal pre-K on cognitive development. *Developmental Psychology,* 41: 872-884.

Hanushek, Eric. 1999. Some findings from an independent investigation of the Tennessee STAR Experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis,* 21: 143-163.

Heacox, Diane. 2006. *Diferenciação curricular na sala de aula.* Translated by João Almeida. Porto: Porto Editora.

Hoxby, Caroline. 1998. *The effects of class size and composition on student achievement: New evidence from natural population variation.* NBER Working Papers 6869, National Bureau of Economic Research, Inc.

Hoxby, Caroline. 2000. The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics, MIT Press,* 115: 1239-1285.

Jacob, Brian, and Lars Lefgren. 2004. Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics, MIT Press,* 86: 226-244.

Jimerson, Shane. 2001. Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30: 420.

Krueger, Alan. 1999. Experimental estimates of education production functions. *The Quarterly Journal of Economics,* 114: 497-532.

Lei n.º 46/86, D.R. nº 237, Série I, de 1986-10-14

McCartney, Kathleen, and Robert Rosenthal. 2000. Effect size, practical importance, and social policy for children. *Child Development*, 71:173-180.

Moreira, Luísa. 2009. *Projecto fénix – um projecto de inovação pedagógica: operacionalização e impacto no agrupamento Campo Aberto – Beiriz.* Tese de Mestrado não publicada, Porto: Universidade Católica Portuguesa.

Murnane, Richard, and John Willett. 2010. *Methods matter.* Oxford: Oxford University Press.

Organisation for Economic Cooperation and Development. 2007. *Improving school leadership*. http://www.oecd.org/dataoecd/44/56/40710632.pdf

Organisation for Economic Cooperation and Development. 2010. *PISA 2009 results: What students know and can do – student performance in reading, mathematics and science*. http://dx.doi.org/10.1787/9789264091450-en

Organisation for Economic Cooperation and Development. 2011. *Grade retention during compulsory education in Europe: Regulations and statistics.* http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports /126EN.pdf

Rosenbaum, Paul, and Donald Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70: 41-55.

Shadish, William, Donald Campbell, and Thomas Cook. 2002. *Experimental and quasiexperimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Verdasca, José, and Teodolinda Cruz. 2006. O projecto TurmaMais: dialogando em torno de uma experiência no combate ao insucesso e abandono escolares. *Revista Portuguesa de Investigação Educacional,* 4: 113-128.

Verdasca, José. 2006. Desempenho escolar, dinâmicas de evolução e elementos configuracionais estruturantes: Os casos do 2º e 3º ciclos do básico nos municípios de Évora e de Portel. *Educação: Temas e Problemas,* 1: 125-134.

Verdasca, José. 2008. "TurmaMais: uma tecnologia organizacional para a promoção do sucesso escolar". In *Conferência Internacional sucesso e insucesso: escola, economia e sociedade*, organized by Manuel V. Cabral, 139-176. Lisboa: Fundação Calouste Gulbenkian.

Verdasca, José. 2010. "Programa Mais Sucesso Escolar: Um desafio na afirmação da autonomia da escola". In *Projecto Fénix – Mais sucesso para todos,* edited by Joaquim Azevedo and José M. Alves, 31-35. Porto: Universidade Católica Portuguesa.

Vygotsky, Lev. 1978. *Mind and society: the development of higher mental processes*. Cambridge, MA: Harvard University Press.

Wong, Vivian, Thomas Cook, Steven Barnett and Kwanghee Jung. 2008. An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, 27: 122-154.