This article was downloaded by: [KU Leuven University Library]
On: 03 March 2014, At: 06:13
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Education Economics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cede20

Football to improve math and reading performance

Chris Van Klaveren[a] & Kristof De Witte[ab]

[a] Top Institute for Evidence Based Education Research (TIER), Maastricht University, PO Box 616, Maastricht 6200 MD, The Netherlands

[b] Faculty of Economics and Business, KU Leuven, Leuven 3000, Belgium

Published online: 27 Feb 2014.

To cite this article: Chris Van Klaveren & Kristof De Witte (2014): Football to improve math and reading performance, Education Economics, DOI: 10.1080/09645292.2014.882293

To link to this article: http://dx.doi.org/10.1080/09645292.2014.882293

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Education Economics, 2014
http://dx.doi.org/10.1080/09645292.2014.882293

Routledge
Taylor & Francis Group

Football to improve math and reading performance

Chris Van Klaveren[a*] and Kristof De Witte[a,b]

[a]*Top Institute for Evidence Based Education Research (TIER), Maastricht University, PO Box 616, Maastricht 6200 MD, The Netherlands;* [b]*Faculty of Economics and Business, KU Leuven, Leuven 3000, Belgium*

(*Received 19 March 2013; accepted 22 December 2013*)

Schools frequently increase the instructional time to improve primary school children's math and reading skills. There is, however, little evidence that math and reading skills are effectively improved by these instruction-time increases. This study evaluates 'Playing for Success' (PfS), an extended school day program for underachieving pupils that uses the football environment as a motivating force. Primary school pupils with low motivation and self-esteem are offered practical and sports-related teaching content for 30 additional hours. The evaluation results of a randomized experiment suggest that PfS does not significantly improve math and reading performance of primary school students.

Keywords: extended school day; increased instructional time; randomized experiment; field experiment; motivation

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

1. Introduction

Programs that intend to improve children's cognitive learning outcomes frequently increase the instructional time that is allocated to core subjects. Well-known examples are the *No Child Left Behind* act in the USA (Bush 2001), the *Future for Education and Care* program in Germany (see section 'Development of All-Day School' in Freitag and Schlicht 2009), and the *Extended School Times* project in the Netherlands (Meyer and Van Klaveren 2013).

These programs generally consider children's school motivation to be an important driver of their cognitive learning outcomes. The positive association between motivation and cognitive learning outcomes is largely supported by the empirical literature (see, for example, Pintrich and Schunk 2002), which also mentions that the interaction between motivation and learning outcomes can create a vicious circle: low learning outcomes negatively influences motivation, which in turn negatively influences learning outcomes.

This study evaluates the Dutch Playing for Success (PfS) program, an example of a program that aims at improving children's math and comprehensive reading skills while emphasizing the important role of motivation. PfS originates from the UK where Premier League football clubs have been collaborating with schools since 1997. Since then it has also been introduced in the Netherlands (since 2008), Japan (since 2010), Spain (since 2010) and Belgium (since 2011). PfS in the Netherlands cooperates with the Dutch Premier League club PSV Eindhoven. Dutch primary school children, aged between 9 and 12, are offered an educational program in

*Corresponding author. Email: cp.vanklaveren@maastrichtuniversity.nl

© 2014 Taylor & Francis

which social learning theory (Bandura 1969) and the concepts of meaningful and cooperative learning (see Slavin 1996; Marzano 2003) are combined to improve children's math and comprehensive reading skills. It is believed that children are motivated by the sports-related education environment in the PSV football stadium, which will in turn improve the learning outcomes of these children.

The PfS program in the UK was evaluated by several empirical studies (see Sharp et al. 2001; Schagen, Kendall, and Sharp 2002; Sharp and National Foundation for Educational Research in England and Wales 2003; Sharp, Kendall, and Schagen 2003). These studies report that PfS significantly improves children's math and reading performance. Sharp et al. (2001) and Sharp and National Foundation for Educational Research in England and Wales 2003), for example, found that the program effect on math performance is equal to 17 months of regular primary school education. This is a substantial program effect given that the program duration for most children was less than 20 hours. The authors therefore conclude that these 'gains in numeracy are particularly impressive, given the relatively short periods of time for which pupils attend' (Sharp and National Foundation for Educational Research in England and Wales 2003, 7).

There are, however, several reasons to suspect that the estimated huge program effects by previous studies are overly optimistic. First of all, it is questionable if the specified control groups in these studies are valid. Sharp and National Foundation for Educational Research in England and Wales (2002), for example, construct a control group in which children are 'selected to be as similar as possible to children attending PfS' (12), but the 'selection' procedure is not accurately explained and the study does not statistically test by a wide range of background characteristics if PfS children differ from control children. Sharp and National Foundation for Educational Research in England and Wales (2003), which is a follow-up evaluation of PfS program in the UK, mention that 'the progress of pupils attending PfS was compared with that of last year's control group' (9), by which they refer to the control group children of their 2002 study. The consequence is that systematic (unobserved) pupil differences between the PfS and control groups (such as parental education, motivation and test scores at the start of the program) can influence and bias the estimated program effect. In Schagen, Kendall, and Sharp (2002), a comparison table of the background characteristics of PfS and control children is shown. This table reveals, among others, that the PfS group is five times larger than the control group and that there are many differences in background characteristics and pre-test scores. Moreover, and worse, the table indicates that the positive PfS estimates for numeracy are to a large extent driven by the fact that children in the control group score lower post-test scores than pre-test scores, while children in the PfS group score higher post-test scores than pre-test scores.

A second drawback of previous studies is that a constant program effect is estimated, while there are many different locations (the so-called Centers) where the program is executed. Although all Centers aim to raise attainments in literacy, numeracy and information and communications technology (ICT), they have their own peculiarities in, for example, the nature of the learning program, the course length, the capacity and resources (e.g. Sharp, Kendall, and Schagen 2003). This study avoids this pitfall by focusing on one implementation in the Netherlands.[1]

Third, earlier work by Sharp and National Foundation for Educational Research in England and Wales (2003) relied on selective small-scale qualitative interviews. Teachers selected pupils for an interview and received instructions to select 'typical'

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

attendees. In total, 39 students from 5 primary and 4 secondary schools were selectively chosen for an interview by 11 teachers. These interviews provide interesting information on the underlying mechanism of the intervention, but because students were not randomly selected it does not provide information on the causal effect of the intervention.

This paper evaluates the effectiveness of PfS by conducting a randomized experiment for Eindhoven, a large Dutch city in the Netherlands. The pedagogical staff and child psychologists of the 24 participating schools provided a list of children who were eligible for the program. These eligible children had low cognitive skills, low motivation levels and/or low self-esteem. The children who were eligible and were willing to enroll in the program were randomly assigned to a control and a PfS groups conditionally on past math and comprehensive reading test scores. The PfS group participated in the program in the first 20 weeks of the school year, while the control group participated in the program in the last 20 weeks of the school year (and after a post-test). The experimental setting with random assignment of children to the program ensures that systematic (unobserved) pupil differences between the PfS and control groups are taken into account and that observed differences in the post-test scores can be attributed to the PfS program.

This study contributes to the empirical literature by evaluating an extended school day program by conducting a randomized experiment. Reviews on the effectiveness of extended school day programs conclude that studies often do not properly control for selection and composition effects, such that the reported estimates may be biased (see Cooper et al. 2000; Scott-Little, Hamann, and Jurs 2002; Lauer et al. 2006, and references therein). A recent review by Patall, Cooper, and Allen (2010) also concludes that rigorous evaluation designs are still very scarce, and that neutral to small positive effects can be expected on cognitive skills. The studies of James-Burdumy et al. (2005), Robin, Frede, and Barnett (2006) and Meyer and Van Klaveren (2013) are worth mentioning because randomized experiments were conducted to evaluate extended school day programs in which instructional time was increased and allocated to math and reading. While Robin, Frede, and Barnett (2006) finds a small positive and significant extended school day effect on math and vocabulary skills, James-Burdumy et al. (2005) and Meyer and Van Klaveren (2013) do not find a positive program effect.

More generally, this study contributes to the empirical literature that evaluates extended school day programs in which learning outcomes and motivation are encouraged through sports and role models (e.g. Perkins, 2000; Skelton, 2000; Bricheno and Thornton, 2007). It also contributes to the literature that evaluates other types of extended school day initiatives, since the didactical mechanism of these initiatives (i.e. meaningful and cooperative learning) that should improve children's cognitive skills tends to be similar to the didactical skills applied in the PfS initiative (see Patall, Cooper, and Allen 2010; Meyer and Van Klaveren 2013). It is worth mentioning that the program content is also interesting for children who are less interested in the football as a sport, even though the theme of the learning program is sports-related (see, for example, Sharp, Kendall, and Schagen 2003). The main emphasis of the program lies on learning math and reading based on meaningful and cooperative learning and not, for example, on playing football.

The paper continues as follows. Section 2 characterizes the PfS program and Section 3 describes the randomized experiment. Section 4 discusses the data and Section 5 empirical results. Finally, Section 6 concludes.

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

2. Playing for success

The PfS program targets underachieving primary school children (aged 9–12 years).[2] A team of experts in each school (pedagogical staff and child psychologists) diagnosed children as underachievers based on their cognitive and social–emotional development. These experts defined seven risk categories and children were considered to be eligible if they could be assigned to at least one of the categories listed below:

(1) Language or math deficit.
(2) Underachievement given the child's potential.
(3) Lack of motivation for schooling.
(4) Low self-esteem.
(5) Quiet and unobtrusive.
(6) Being bullied.
(7) Not sufficiently assertive.

The duration of the program is 10 weeks and children who participate in the program receive three after-school education hours each week in the football stadium of PSV Eindhoven. The program is well guided in the sense that it assigns to each group of four children one teacher and a social worker or a trainee (i.e. students from teacher training programs, social and cultural work, youth helpers and students from sport studies). The professional background differences of the employees make the team multidisciplinary and more experienced. The program costs are about 350 euros per participating child and 25% of these costs are funded by the football club, while the other 75% is funded by public subsidies. One of the reasons why football clubs invest in these programs is because it is regarded as a sign of social commitment of the club.

The theoretical underlying mechanism of PfS originates from social learning theory (Bandura 1969) and from the works of Marzano (2003). Social learning theory argues that improvement in self-efficacy is necessary for school motivation and scholarly success. Therefore, the program works around practical sports-related themes while using the 'meaningful learning' concept of Marzano (2003) to relate abstract subject matter to concrete profession experiences in the real world. Children are, for example, asked to write training programs, communicate with journalists, compute the size of the field, etc. In a real-life sports-related environment both languages and numeracy are explicitly practiced, as well as ICT skills. Cooperative learning is, moreover, adopted as didactical learning strategy within the program. Cooperative learning implies that children perform tasks in small heterogeneous groups with individual responsibilities. There is a large empirical literature that shows that cooperative learning is an effective learning strategy (see for an elaborate review Slavin 1996).

Besides the underlying theoretical mechanisms, local policy-makers and the PfS organization believe that the learning outcomes of children may improve because the traditional school environment is replaced by the more stimulating 'stadium environment', and because children have the opportunity to meet the first-team players. Even though the learning program is sports-related it should be emphasized that the program content is also interesting for those children who are less interested in football (see also Sharp, Kendall, and Schagen 2003). Sharp, Kendall, and Schagen (2003), for example, showed that girls, who are considered to be less interested in football than boys, do not appreciate the program content differently than boys.

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

The Dutch PfS program differs somewhat from the PfS programs in other countries because of the less intensive use of ICT. In the Dutch program, math and comprehensive reading skills are not offered using ICT, even though all children in the Dutch program have their own computer. The Dutch program additionally emphasizes the importance of defining and achieving learning goals. Children are responsible for defining and achieving their learning goals and the employees support children in this process. More generally there are small program differences between countries. However, all programs target similar child populations (i.e. underachieving children) and use a similar selection procedure (i.e. experts within schools label children as underachievers based on their cognitive and social–emotional development). Hence, the results of our study seem representative for PfS programs in other countries.

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

3. Experimental design

Twenty-four Dutch primary schools participated in the randomized experiment. One hundred and seventy-eight children were eligible because they were diagnosed as underachievers by the school experts, were willing to participate in the PfS program and because their parents gave their consent.

The experimental design is illustrated in Figure 1. To avoid ethical issues, all children who were eligible and willing to participate received the PfS treatment. Children in

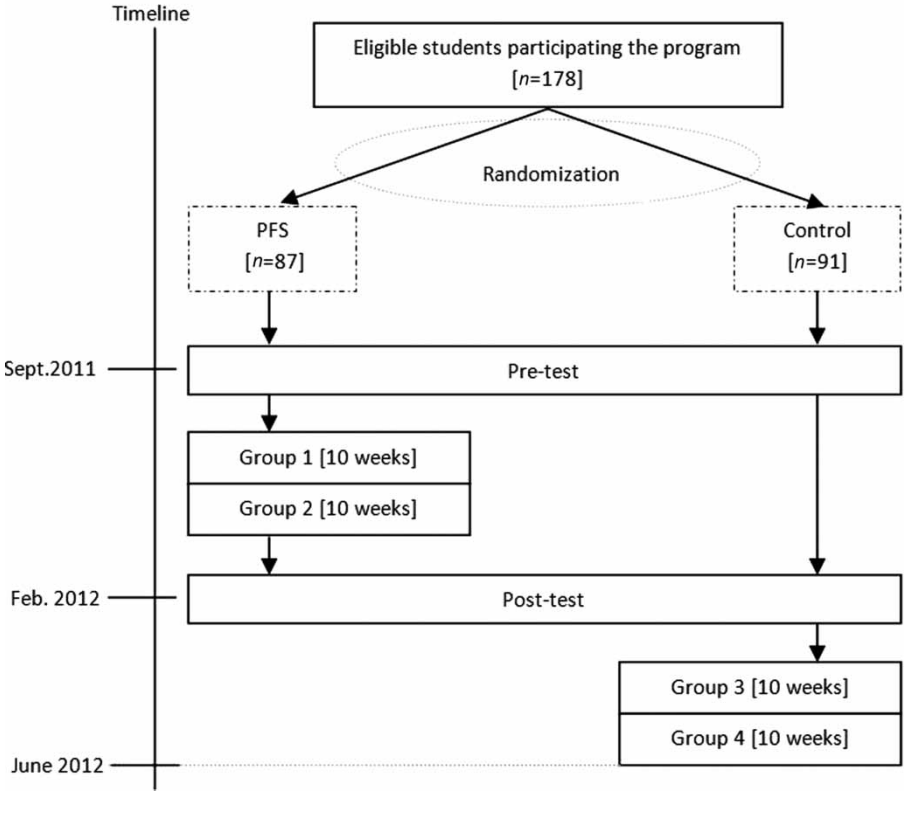

Figure 1. Graphical illustration of the experimental design.

the experimental group (later also referred to as the PfS group) receive the treatment in the first 20 weeks of the school year (before the post-test in February), while students in the control group receive the treatment in the last 20 weeks of the school year (after the post-test in February). In February all Dutch primary school children take a standardized math and comprehensive reading test in February 2012, and therefore, this test conveniently serves as the post-test. Frequently, randomized field experiments are considered as unethical because it withholds children from program participation (see, for example, Rodrik 2008), but schools and parents did not consider the experiment as unethical because all eligible children would eventually participate in the program.

The 178 children who participated in the experiment were paired within schools and classes conditionally on their math and comprehensive reading test scores achieved on a national and standardized test taken in February 2011 (i.e. one year before the post-test). For each child a random number from a uniform distribution was drawn on the interval [0.1]. The child with the highest number within each pair was assigned to the PfS program in the first 20 weeks of the school year (i.e. the experimental group), and the child with the lowest number was assigned to the PfS program after the post-test was taken (i.e. the control group). The figure shows that children were unevenly divided over the experimental and the control group. This occurs because there sometimes were uneven numbers of potential participants within classes such that not all children within a class could be paired. The adopted assignment rule for nonpaired children assigned them to the experimental group if the drawn random number was higher than 0.5 and to the control group otherwise.

The experimental and control groups were randomly divided into two subgroups to increase the intensity of the program and to give children sufficient attention. Groups 1 and 2 represent the experimental group and groups 3 and 4 represent the control group. The learning activities and content were equal for all the groups.

While an experimental design does not necessarily require a pre-test (given the randomization), we took a pre-test to increase the efficiency of the empirical results. This pre-test was taken in September 2011 and designed specifically for this study by CITO (www.cito.com), a testing and assessment company that designs the national standardized tests that children normally make in primary education.

4. Experimental data

Table 1 shows why experts labeled children as underachievers. Usually children have multiple problems and school experts therefore categorized them into more than one

Table 1. Reasons for underachieving indicated by experts.

	N	Freq. (%)
Language or math deficit	59	33.15
Underachievement given child's potential	62	34.83
Insufficient motivation	41	23.03
Low self-esteem	101	56.74
Quiet and unobtrusive	33	18.54
Being bullied	14	7.87
Not assertive enough	33	18.54

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

Table 2. Comparing children in the experimental and control groups.

	Eligible students participating in the experiment				
	Experimental group		Control group		Differences
	Mean	Std Dev.	Mean	Std Dev.	$p > \|z\|$
Girl (=1)	0.35	0.48	0.37	0.49	0.81
Age	10.43	0.98	10.57	0.92	0.56
Dutch (=1)	0.53	0.53	0.59	0.49	0.39
Grade 3 (US terminology)	0.03	0.18	0.02	0.15	0.61
Grade 4	0.39	0.49	0.43	0.50	0.61
Grade 5	0.33	0.47	0.34	0.48	0.92
Grade 6	0.24	0.43	0.21	0.41	0.61
Father's education low	0.21	0.41	0.17	0.38	0.60
Father's education middle	0.38	0.49	0.42	0.50	0.60
Father's education high	0.33	0.47	0.30	0.46	0.60
Mother's education low	0.31	0.47	0.31	0.46	0.97
Mother's education middle	0.36	0.48	0.42	0.50	0.40
Mother's education high	0.23	0.42	0.22	0.42	0.87
Reading February 2011 (assignment test)	26.54	15.95	28.74	14.72	0.34
Reading September 2011 (pre-test)	27.81	18.52	27.58	17.93	0.98
Math February 2011 (assignment test)	75.32	20.39	77.97	18.71	0.53
Math September 2011 (pre-test)	75.44	20.56	77.44	17.54	0.83
Number of observations	87		91		

category. The table indicates that 60% of the children were labeled as children with low self-esteem. The three other most diagnosed deficiencies, that go hand-in-hand with having a low self-esteem, were language or math deficits (33.15), underachievement given the child's potential (34.83) and motivational problems (23.03). For the analysis it is important to recognize that program effects may be different for children in different categories.

Table 2 presents several background characteristics for the experimental and the control group. The table shows the means and standard deviations of observed child characteristics, and the last column shows the *p*-values and indicates if the observed differences in means are significant.

In the Netherlands, girls and boys are usually evenly distributed over classes (see also Table 3). In the PfS experiment, Table 2 indicates a relatively large proportion of boys participating in the PfS program. Because girls aged around 9 typically

Table 3. School characteristics.

	Mean	Std Dev.
Proportion of children in disadvantageous background (weight 0.3)	0.11	0.08
Proportion of children in disadvantageous background (weight 1.2)	0.15	0.17
Proportion of children in disadvantageous areas	0.33	0.47
Proportion of Dutch	0.76	0.08
Proportion of boys	0.51	0.51
Number of students at the school	194.75	143.91
Number of observations	178	

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

perform better on cognitive tests than boys, they might be less eligible for the program. It may also be that girls are less attracted to football and therefore, despite being eligible, do not enroll for the PfS program. The overrepresentation of boys does not affect the internal validity of the experiment (as children are randomly assigned to the treatment).

Children in the program are, on average, 10.5 years old, which is consistent with the observed grade proportions: eligible children mostly came from grades 4 and 5. The observed small and nonsignificant differences in grade proportions between both child groups are the consequence of having an uneven number of children within classes. The proportion of native Dutch children (i.e. children whose parents are both born in the Netherlands) is slightly higher in the control group than the proportion of native Dutch children in the PfS group, even though it is not statistically significant. The proportion of native Dutch children among eligible students participating in the experiment is lower than the proportion of Dutch children on the 24 participating schools (0.76; Table 3). Dutch children are less likely to participate in the PfS experiment, either because they are less often labeled as underachievers or because they are less motivated to participate in the program. Nevertheless, and given the high proportion of Dutch children, we conclude that the focus of PfS is not primarily on non-Dutch students, which differs from many existing extended school day programs that tend to focus on disadvantageous and nonnative students (see, e.g. Patall, Cooper, and Allen 2010; Meyer and Van Klaveren 2013).

Highest achieved education level of the parents is categorized by three education levels: low (i.e. without formal education, primary or lower secondary education), middle (i.e. higher secondary or vocational education) and high (i.e. bachelor or academic education). The descriptive statistics show that these education levels are evenly distributed over the three categories, which again shows that the PfS program does not focus primarily on disadvantageous children who often have lower educated parents (and come from economically disadvantaged neighborhoods). Given that there is a positive causal effect of parent's schooling on children's schooling (Holmlund, Lindahl, and Plug 2011) and that the program welcomes – but is not aimed specifically at – children with a diagnosed disorder, we would expect that children of higher educated parents are more likely to be underachievers. When we compare the proportions of mothers and fathers in the three education categories between the PfS and the control group, we find that the means (and medians) are not statistically different between the two groups.

The lower panel of Table 2 shows the descriptive statistics on achieved math and reading test scores in February 2011 and September 2011. The February test scores were used to assign children to the experimental and the control group. The September test scores are standardized tests developed specifically for this study (see also Section 3) and measure children's math and comprehensive reading ability before the PfS program started. The September test scores serve therefore as pre-test scores. In Appendix 1, we describe in detail how observed test scores of children relate to the achievement levels that they are supposed to have in that particular grade. As expected, we find that the achievement levels of the eligible children are below the appropriate achievement levels.

The descriptive statistics in Table 2 show that the mean test scores in February 2011 and September 2011 are not statistically different between the experimental and the control group. An interesting observation is that the September 2011 test scores are very similar to the February 2011 test scores, which gives the impression that children

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

have not (or only marginally) improved their math and reading skills. There are three explanations for this observation. First of all, children took the math and reading test in September during a test day that was organized at the football stadium of PSV Eindhoven.[3] The test day was organized as a festive day, such that the first experience with the PfS program was a positive and a motivating one. It is therefore likely that the more 'noisy' test day environment influenced children's math and reading test scores negatively. That the test day was a more 'noisy' environment is also reflected by the standard deviation of the reading test in February which is lower than that of the reading test in September. The standard deviations for the math tests appear to be very similar, which might be due to the moment of testing: the math test was taken in the morning, while the reading test was taken in the afternoon. For the analysis it is important that children were randomly assigned to the program, such that the produced noise by the test day environment was randomly distributed as well. So even though the September test scores are downward biased, they still indicate properly that the math and reading skills of the PfS and control group children are not statistically different at the beginning of the PfS program.

The second explanation for the low test scores in September is that the acquired math and reading skills during the school year are partly lost during the summer holidays. This is generally referred to as the summer dip and has led to many summer school initiatives that aim at reducing the negative effects of such a summer dip (see, for example, Cooper et al. 2000; Jacob and Lefgren 2004; Borman and Dowling 2006). A final explanation is that the September test is more difficult than the February test. This is, however, unlikely because both tests are made by CITO, which also designs the regular national math and reading tests that are taken in primary and secondary education. Moreover, children are assigned to secondary education tracks based on the CITO test that children take at the end of grade 6. Therefore, the tests can be considered as valid and reliable. Important for this study is that we randomly assigned children to the treatment and therefore we simply include the February and September pre-test scores to gain efficiency without negative consequences for the validity of our estimates.

Table 3 shows school characteristics for all children participating in the experiment. Because the randomization happened within schools and classes there is no need to present these statistics separately for children in the PfS and the control group, since they are by construction similar for both groups. The first and second rows denote the proportion of children with a disadvantageous background. In the Netherlands, schools receive subsidies for children with higher needs and The Ministry of Education distinguishes between two types (and thus two weights) based on the education level of the parents.

A weight of 0.3 is assigned to a child if both parents finished a lower secondary education level as highest education level (i.e. the lowest pre-vocational track) and a 1.2 weight is assigned to a child if one parent finished a lower secondary education level while the other parent has a lower education level. Table 3 shows that 11% of the children receive a 0.3 weight and that 15% of the children receive a 1.2 weight. Official school-level registration data on all Dutch primary schools indicate that on an average 9% of the children in a Dutch primary school receive a 0.3 weight and 6% of the children receive a 1.2 weight. This indicates that the schools that participated in the experiment have a rather disadvantageous student population.

In a similar fashion we can compare other observed characteristics relatively to the Dutch average. Disadvantageous areas denote neighborhoods with an above-average

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

combination of unemployment, early school leaving, criminality and low income. They receive additional funding by the central government. The proportion of children living in a disadvantageous area is 0.33, which is relatively higher compared to that of the average primary school (0.15). The proportion of native Dutch children is 0.76, which is slightly higher than that of the average primary school (0.71). This means that even though children are more often living in disadvantageous neighborhoods the proportion of non-Dutch students is relatively small. This might be due to selection in schools along ideology (see also De Witte and Van Klaveren 2012). Finally, the proportion of boys is similar to that of the average primary school.

For 13 of the 178 participating children we do not observe post-test scores and these children cannot be considered in the empirical analysis. Table 4 characterizes these children and examines the selective nature of this experimental dropout. The 13 children are evenly distributed over the PfS and the control group which indicates that the incidence of not observing post-test scores was randomly distributed over the two groups. Seven of the 87 children in the PfS group dropped out of the program. Two of these children dropped out because of practical reasons and enrolled again in the second period, one child dropped out of the program because she moved to another school and four children did not want to continue with the program for unknown reasons. For 6 of the 91 children in the control group we do not observe the post-test scores. One child moved to another school, but for the other children in the control group it is unclear why they did not take the post-test. Because of the small number of children dropping out of the program, in Table 4, we cannot determine (parametrically nor nonparametrically) if the means are significantly different between the two groups. Consequently, we do not show the standard deviations. Table 4 reveals that parental education of dropouts in the experimental group is, on average, higher than parental education of dropouts in the control group. On the other hand, children who drop out of the control group are more often native Dutch and higher achieving.

Table 4. Comparing dropout children in the PfS and control groups.

	Dropouts	
	PfS Mean	Control Mean
Girl (boy = 0)	0.29	0.33
Age	10.33	10.70
Dutch (=1)	0.29	0.83
Grade 3 (US terminology)		0.17
Grade 4	0.43	0.33
Grade 5	0.43	0.33
Grade 6	0.14	0.17
Father's education average[a]	2.50	1.5
Mother's education average	2.00	1.67
Reading February 2011 (assignment test)	28.00	36.17
Reading September 2011 (pre-test)	32.77	25.17
Math February 2011 (assignment test)	73.57	80.00
Math September 2011 (pre-test)	76.27	79.33
Number dropouts	7	6
Number of assigned children	87	91

[a]Father's education has two missing observation.

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

Nevertheless, considering that the pre-test scores of children assigned to the control group are somewhat (but not significantly) higher than those of children assigned to the PfS group, and given the small percentage of experimental dropout, it is unlikely that the observed differences of children who dropped out of the program will bias the estimated program effect.

5. Estimation results

The evaluation of the results proceeds in three steps. First, we examine the test score differences using simple *t*-tests. The upper panel of Table 5 presents the reading estimates, while the lower panel presents the math estimates. The first two rows of each panel show the mean pre- and post-test scores, while the last column shows if the post-test scores differ significantly from the pre-test scores. The third row of each panel shows if the mean pre- and post-test scores and the difference between the two significantly differ between the PfS group and the control group.

The results indicate that the reading performance of children in the PfS group increased with 9.8 points, while the reading performance of children in the control group increased with 10.9 points. To put those figures in perspective, we relate them to the expected performance increase as indicated by the testing institute CITO. Table 6 shows for each grade how much children should improve their math and reading skills per year, according to the standards that are set by the testing institute CITO. As younger children learn faster, Table 6 indicates that children in lower grades should improve their achievement levels more than children in higher grades. The average improvement over all grades amounts to 12.75 points for math and 10.10 points for reading. Unfortunately, these improvement points cannot directly be related to the increase in math and reading performance of children in the PfS and control groups, as these children are unevenly distributed over grades. We avoid this drawback by constructing weighted averages indicating how much children in the PfS and control groups should have improved their math and reading performance conditional on the numbers of children who were enrolled in each grade. This is presented on the right-hand side of Table 6. The weighted averages reveal that children in the control group should obtain a higher performance increase on numeracy and reading than students in the PfS group.

Table 5. Mean pre- and post-test scores for PfS and control groups.

	Pre-test scores	Post-test scores	Pre-post differences
Reading			
PfS group	26.23 (2.10)	36.04 (1.92)	9.81*** (1.80)
Control group	28.14 (2.10)	39.01 (1.69)	10.88*** (1.49)
PfS-control difference	−1.90 (2.98)	−2.97 (2.55)	−1.07 (2.32)
Math			
PfS group	75.42 (2.43)	88.64 (2.14)	13.21*** (1.84)
Control group	77.80 (1.95)	91.52 (1.63)	13.72*** (1.31)
PfS-control difference	−2.37 (3.11)	−2.88 (2.68)	−0.51 (2.25)

Note: Standard errors are shown in parentheses.
*Statistically significant at the 10% level.
**Statistically significant at the 5% level.
***Statistically significant at the 1% level.

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

Table 6. Average expected performance increase according to testing institute CITO.

	Assumed increase			Proportion	
	Math	Reading		PfS (87)	Controls (91)
Grade 3	19.75	16.34		0.03	0.02
Grade 4	13.5	8.71		0.39	0.43
Grade 5	12.25	8.84		0.33	0.34
Grade 6	5.50	6.50		0.24	0.21
Average	12.75	10.10	Weighted average reading	8.36	8.44
			Weighted average math	11.22	11.52

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

Relating the theoretical average performance increase of Table 6 to the observed performance increase of Table 5, one could conclude that the observed average improvement of the children in the experimental group is more than sufficient, considering the standards. It is, however, important to recognize that the pre-test scores of these students are lower than the pre-test scores that children should achieve according to the standard. Therefore, it is only natural that the performance increase is more than sufficient.

Test score improvements of children in the control group are somewhat higher than those of children in the PfS group, but this is mainly due to a small, and statistically insignificant, difference in the pre-test scores.

As a second evaluation procedure, we estimate the treatment effect more precisely by an ordinary least squares regression. Light, Singer, and Willett (1990) showed that the statistical power is maintained at half the sample size if a regression analysis is performed with a set of covariates that predicts about half of the variation in the outcome jointly. The following model is estimated:

$$Y^s_{\text{post},i} = \alpha_0 + \alpha_1 X_i + \alpha_2 \text{PFS}_i + \alpha_3 Y^s_{\text{Sept},i} + \alpha_4 Y^s_{\text{Feb},i} + \varepsilon_i, \quad (1)$$

where $Y^s_{\text{post},i}$ denotes the standardized post-test score of student i for subject s (s=math, reading), X_i represents a vector of student and school characteristics, and PFS indicates if the student participated in the PfS program. $Y^s_{\text{Sept},i}$ denotes the September pre-test score and $Y^s_{\text{Feb},i}$ represents the February test score which is included to account for the difference between the September test score and the February test score. As usual, the error term, $\in_i$, is assumed to be normally distributed with mean zero and variance $\sigma^2_{\in}$ and all explanatory variables are assumed independent of the error term. The estimation results for math and reading are presented in Table 7.

The left panel of Table 7 shows the results for math, while the right panel presents the results for comprehensive reading. For each subject two model specifications are estimated that control for achieved September and February test scores as well as for child control variables. The latter include gender, age, ethnicity, a set of grade dummies and a set of dummy variables that indicate the education level of the father and mother (i.e. those mentioned in Table 2). Models (1) and (3) show the estimation results when we do not include school-level controls, while school-level control variables are included in models (2) and (4). The included school-level variables are similar to those in Table 3. Including school-level control variables does not improve the explained variation in the outcome variables nor does it influence the estimated effect of the treatment.

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

Table 7. Regression estimates of the PfS effect for the treated.

	Math				Reading			
	(1)		(2)		(3)		(4)	
	Coef.	Std Err.	Coef.	Std Err.	Coef.	Std Err.	Coef.	Std Err.
PfS participation (1=yes)	−0.054	(0.098)	−0.042	(0.098)	−0.058	(0.101)	−0.065	(0.103)
$Y^s_{\text{Septe},i}$	0.016***	(0.004)	0.016***	(0.004)	0.019***	(0.003)	0.020***	(0.004)
$Y^s_{\text{Feb},i}$	0.026***	(0.004)	0.027***	(0.004)	−0.026***	(0.005)	−0.027***	(0.005)
Constant	−3.940***	(0.967)	−3.013*	(1.595)	−1.585*	(1.778)	−1.980	(1.788)
Child controls	Yes		Yes		Yes		Yes	
School controls	No		Yes		No		Yes	
N	165		165		165		165	
Adj. R^2	0.66		0.66		0.61		0.60	

Note: Standard errors are printed in parentheses.
*Statistically significant at the 10% level.
**Statistically significant at the 5% level.
***Statistically significant at the 1% level.

To be more precise, in estimating the model with only child-level background characteristics, an adjusted-R^2 of 0.30 for math and 0.26 for reading is observed. By adding the pre-test score variables, the adjusted-R^2 increases to 0.66 and 0.61 for, respectively, math and reading. Adding additional school-level variables does not increase or even decreases (for reading) the adjusted-R^2. The estimation results become thus less efficient by including school-level variables. In the discussion below, we therefore focus on the estimation results of models (1) and (3). Note that the estimation results are similar when the standard errors are clustered at the class or the school level. Moreover, we performed quantile regressions to take into account possible nonlinear effects of the PfS program. However, the estimation results were similar to those reported in Table 7.

The results further indicate that children have higher post-test scores if they have higher test scores in September and February 2011. The estimation outcomes reveal that the February 2011 test scores correlate better with the post-test scores than the September test scores. In Section 4, we explained that this is likely caused by the noise that was produced by the festive test day environment in September. The pre- and post-test correlation tables are shown in Appendix 2.

Finally, and most importantly, the estimation results show that children in the experimental group do not perform significantly better than children in the control group. The estimate is even negative, which is likely due to the higher pre-test scores of the control group children (note that these were not significantly different from the experimental group children). The effect size of about -0.06 for both math and reading is close to zero. Both the difference-in-differences in the simple *t*-test in Table 5 and the linear regression estimates therefore indicate the PfS program did not effectively increase the math and reading performance of the program participants.

A potential problem is that children give up certain activities when participating in PfS and the effectiveness of the program may depend on the type of activities that children give up. Unfortunately, as the intervention is an extended school day program, it is unknown which activities children gave up when participating in PfS, and therefore, we assume that these activities correlate strongly with the education level of the mother and the nationality of the child. The intuition behind this assumption is that socio-economic background variables may determine what children do after school, and children of higher educated parents may, for example, read more in their leisure time which can make the program less effective for them.

The estimation results of (1) and (3) in Table 8 show the estimated interaction effects when the treatment indicator is interacted with the mother's education level (high, middle and low (=ref)) and with being Dutch (1= Dutch) for, respectively, math and reading. We conveniently show the estimated interaction effects only as these effects indicate if the estimated program effect varies with the interacted background characteristics. Columns (2) and (4) show the estimation results, for, respectively, math and reading, when interactions between the treatment indicator and the indications of the experts (i.e. Table 1) are included. The estimation results show that none of the included interaction effects are statistically significant. Under the assumption that the type of after-school activities are strongly related to the indications and background characteristics shown in Table 8, we conclude that the effectiveness of PfS does not depend on the type of activities that children give up.

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

Table 8. Estimation results when the PfS indicator is interacted with socio-economic variables and the school expert's indications.

	Math				Reading			
	(1)		(2)		(3)		(4)	
	Coef.	Std Err.	Coef.	Std Err.	Coef.	Std Err.	Coef.	Std Err.
PfS constant	0.065	(0.179)	0.185	(0.199)	0.214	(0.217)	0.073	(0.245)
PfS*(edu mother miss)	−0.204	(0.363)			−0.108	(0.453)		
PfS*(edu mother middle)	0.175	(0.221)			−0.018	(0.270)		
PfS*(edu mother high)	−0.057	(0.255)			−0.312	(0.311)		
PfS*(child is Dutch)	−0.293	(0.193)			−0.322	(0.237)		
PfS*(language or math deficit)			−0.031	(0.195)			−0.001	(0.240)
PfS*(cognitive underachievement)			−0.393	(0.208)			−0.118	(0.251)
PfS*(insufficient motivation)			0.020	(0.225)			0.222	(0.276)
PfS*(low self-esteem)			−0.315	(0.184)			−0.266	(0.227)
PfS*(quiet and unobtrusive)			−0.040	(0.236)			0.151	(0.298)
PfS*(being bullied)			−0.178	(0.345)			0.568	(0.424)
PfS*(not assertive enough)			0.356	(0.230)			0.200	(0.281)
Expert indications	No		Yes		No		Yes	
Pre-test controls	Yes		Yes		Yes		Yes	
Child controls	Yes		Yes		Yes		Yes	
School controls	No		No		No		No	
N	165		165		165		165	
Log-likelihood	−136.540		−128.774		−170.423		−163.095	

Note: Standard errors are printed in parentheses.
*Statistically significant at the 10% level.
**Statistically significant at the 5% level.
***Statistically significant at the 1% level.

6. Conclusion and discussion

This paper reports the effects of an extended school day program. The program targets underachieving students (i.e. students with low motivation, difficulties at home, socio-economic problems or low self-esteem) and aims to improve their learning outcomes on math and reading by working in a football stadium around role models and by practical exercises. The program called 'PfS' is popular in the UK, Belgium, Japan, Spain and the Netherlands. This paper evaluates the effectiveness of the Dutch program. For the first time, a randomized field experiment is conducted to determine the causal impact of the intervention. The randomization of eligible students avoids any selection biases on observed and unobserved characteristics.

The empirical results suggest that the PfS program did not improve the math or reading performance of the participating children. We see three possible explanations for the absence of an effect. First of all, the program duration of 10 weeks might be too short. Second, we did not test for the effects on motivation but only for educational attainments. Besides reading and numeracy outcomes, the program explicitly targets motivation and self-esteem of students. While the former might not have been changed, the latter might have changed. Third, PfS may have increased the math and reading performance of all eligible children if control group children were contaminated by children in the experimental group. Such a contamination effect means that the estimation results indicate that the PfS program is not effective, while in reality the program does have a positive effect.

Nevertheless, the zero outcome of the intervention is interesting. First, it contrasts to earlier (positive) findings in the literature. As argued in the literature review, this earlier work uses different methodologies which focus on correlations rather than on causal evidence. The revised paper is the first paper to examine the effect of the intervention using experimental evidence. The opposite finding sheds a new light on the effectiveness of the intervention. Second, given the scarce resources which are drawn to this intervention, we think that policy-makers and schools should know that this intervention does not increase learning outcomes.

As a note for a broader discussion, we observe that studies on the effectiveness of extended school day programs find more modest program effects if they focus more on the internal validity. With respect to the PfS program, previous less well-designed studies find (huge) positive program effects, while our randomized experiment finds no PfS effect on learning outcomes. More generally, review studies on the effectiveness of extended school day programs (i.e. Scott-Little, Hamann, and Jurs 2002; Lauer et al. 2006; Patall, Cooper, and Allen 2010) indicate that studies using more rigorous evaluation designs find smaller (and often insignificance) program effects. These review studies clearly show that the reviewed studies often do not attempt to properly control for selection and composition effects or do not even use a control group to identify the program effect, which may result in biased estimates. This study, for examples, also finds great improvements of the math and reading skills for the treatment group, but these improvements obviously cannot be attributed to the evaluated PfS program. The finding that studies focusing more on internal validity find more modest outcomes may not be so surprising given that it is more likely that selection effects are absent. Nevertheless, there is often a pressure by policy-makers, stakeholders and even academic journals to 'find' and report significant outcomes (which results in a publication bias).

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

Acknowledgements

We would like to thank Playing for Success Eindhoven and Charlotte Jacobs of CITO for providing data. The usual disclaimer holds.

Notes

1. Schagen et al. (2002) apply a multilevel model to account for the fact that children are clustered in classes and locations.
2. We note that the PfS program often also targets children between 12 and 14 in lower secondary education, but this is not the focus of the program that we evaluate.
3. To be precise, the September test was taken on 30 August 2011.

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

References

Bandura, A. (1969). "Social-Learning Theory of Indemnificatory Processes." In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research*, 213–262. Chicago: Rand McNally.

Borman, G. D., and N. M. Dowling. 2006. "Longitudinal Achievement Effects of Multiyear Summer School: Evidence from the Teach Baltimore Randomized Field Trial." *Educational Evaluation and Policy Analysis* 28 (1): 25–48.

Bricheno, P., and M. Thornton. 2007. "Role Model, Hero or Champion? Children's Views Concerning Role Models." *Educational Research* 49 (4): 383–396.

Bush, G. W. 2001. *No Child Left Behind*. Washington, DC: Office of the President of the United States.

Cooper, H., K. Charlton, J. Valentine, and L. Muhlenbruck. 2000. "Making the Most of Summer School: A Meta-analytic and Narrative Review." *Monographs of the Society for Research in Child Development* 65 (1): 1–130.

De Witte, K., and C. Van Klaveren. 2012. "Comparing Students by a Matching Analysis – On Early School Leaving in Dutch Cities." *Applied Economics* 44 (28): 3679–3690.

Freitag, M., and R. Schlicht. 2009. "Educational Federalism in Germany: Foundations of Social Inequality in Education." *Governance* 22 (1): 47–72.

Holmlund, H., M. Lindahl, and E. Plug. 2011. "The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods." *Journal of Economic Literature* 49 (3): 615–651.

Jacob, B. A., and L. Lefgren. 2004. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *Review of Economics and Statistics* 86 (1): 226–244.

James-Burdumy, S., M. Dynarski, M. Moore, J. Deke, W. Mansfield, and C. Pistorino. 2005. *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program*, Final Report. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://www.ed.gov/ies/ncee

Lauer, P. A., M. Akiba, S. B. Wilkerson, H. S. Apthorp, D. Snow, and M. L. Martin-Glenn. 2006. "Out-of-School-Time Programs: A Meta-analysis of Effects for At-Risk Students." *Review of Educational Research* 76 (2): 275–313.

Light, R. J., J. D. Singer, and J. B. Willett. 1990. *By Design: Planning Research on Higher Education*. Cambridge, MA: Harvard University Press.

Marzano, R. J. 2003. *What Works in Schools: Translating Research into Action*. Alexandria, VA: Association for Supervision & Curriculum Development.

Meyer, E., and C. Van Klaveren. 2013. "Evaluation of an Extended Day Program in the Netherlands: A Randomized Field Experiment." *Economics of Education Review* 36: 1–11.

Patall, E. A., H. Cooper, and A. B. Allen. 2010. "Extending the School Day or School Year: A Systematic Review of Research (1985–2009)." *Review of Educational Research* 80 (3): 401–436.

Perkins, S. 2000. "Exploring Future Relationships Between Football Clubs and Local Government." *Soccer and Society* 1 (1): 102–113.

Pintrich, P. R., and D. H. Schunk. 2002. *Motivation in Education: Theory, Research, and Applications*. Upper Saddle River, NJ: Merrill Prentice Hall.

Robin, K. B., E. C. Frede, and W. S. Barnett. 2006. *Is More Better? The Effects of Full-Day vs. Half-Day Preschool on Early School Achievement*. NIEER Working Paper. National Institute for Early Education Research. http://nieer.org/docs/index.php?DocID=144
Rodrik, D. 2008. *The New Development Economics: We Shall Experiment, But How Shall We Learn?* Working Paper Series 08-055. Harvard University, John F. Kennedy School of Government.
Schagen, I., L. Kendall, and C. Sharp. 2002. "Measuring the Success of Playing for Success." *Educational Research* 44 (3): 255–267.
Scott-Little, C., M. S. Hamann, and S. G. Jurs. 2002. "Evaluations of After-School Programs: A Meta-evaluation of Methodologies and Narrative Synthesis of Findings." *American Journal of Evaluation* 23 (4): 387–419.
Sharp, C., L. Kendall, S. Bhabra, I. Schagen, and J. Duff. 2001. *Playing for Success: An Evaluation of the Second Year* (DfES Research Report 291). London: DfES.
Sharp, C., L. Kendall, and I. Schagen. 2003. "Different for Girls? An Exploration of the Impact of Playing for Success." *Educational Research* 45 (3): 309–324.
Sharp, C., and National Foundation for Educational Research in England and Wales. 2002. *Playing for Success: An Evaluation of the Third Year*. London: DfES Publications.
Sharp, C., and National Foundation for Educational Research in England and Wales. 2003. *Playing for Success: An Evaluation of the Fourth Year*. London: DfES Publications.
Skelton, C. 2000. "'A passion for Football': Dominant Masculinities and Primary Schooling." *Sport, Education and Society* 5 (1): 5–18.
Slavin, R. E. 1996. "Research on Cooperative Learning and Achievement: What we Know, What we Need to Know." *Contemporary Educational Psychology* 21 (1): 43–69.

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014

Appendix 1

In Dutch primary schools, children take a national standardized math and reading test (usually twice a year). This appendix, and in particular Table A1, explains how children's test scores are linked to achievement levels within primary school grades. The second column of Table A1 shows to which level a certain test score belongs. The numbers in this column refer to the primary school grade, Mid refers to the achievement levels that children should have in the middle of the school year and End refers to the achievement levels that children should have at the end of the school year.

Table A1. Relation between achieved test scores and achievement level.

		Math		Reading	
Nr.	Level	Lower bound	Upper bound	Lower bound	Upper bound
1	<Mid1	−99.00	38.99	−99.00	−5.24
2	Mid1	39.00	44.00	−5.23	0.43
3	Mid1–End1	44.01	48.99	0.44	6.07
4	End1	49.00	54.25	6.08	9.98
5	End1–Mid2	54.26	59.74	9.99	12.10
6	Mid2	59.75	64.50	12.11	15.46
7	Mid2–End2	64.51	68.49	15.47	20.02
8	End2	68.50	73.50	20.03	24.37
9	End2–Mid3	73.51	79.49	24.38	28.44
10	Mid3	79.50	84.00	28.45	31.16
11	Mid3–End3	84.01	86.99	31.17	32.46
12	End3	87.00	90.00	32.47	34.47
13	End3–Mid4	90.01	92.99	34.48	37.15
14	Mid4	93.00	96.25	37.16	40.00
15	Mid4–End4	96.26	99.74	40.01	42.99

(*Continued*)

Table A1 (*Continued*)

		Math		Reading	
Nr.	Level	Lower bound	Upper bound	Lower bound	Upper bound
16	End4	99.75	102.75	43.00	45.00
17	End4–Mid5	102.76	105.24	45.01	45.99
18	Mid5	105.25	107.50	46.00	47.25
19	Mid5–End5	107.51	109.49	47.26	48.74
20	End5	109.50	112.00	48.75	50.50
21	End5–Mid6	112.01	114.99	50.51	52.49
22	Mid6	115.00	118.00	52.50	54.50
23	>Mid6	118.01	999.00	54.51	999.00

Note: We are grateful to Charlotte Jacobs (CITO) for providing us with these statistics.

Let us consider the average math test score in February 2011 of PfS children who were enrolled in grade 4 in school year 2011–2012. Table 2 only shows an average math test score of 75.32, which represents the test scores of all PfS children who were enrolled in grades 3–6 in school year 2011–2012. The average math test scores achieved in February 2011 by PfS children who were enrolled in grade 4 is 69.58. It is important to recognize that these children were still in grade 3 in February 2011, and so the achievement levels that these children are supposed to have are shown in Table A1 in row number 10, where column 2 indicates Mid3. The lower and upper bound math score interval for Mid3 is [79.5;84], and therefore, we can conclude that the math level that PfS children were supposed to have was below level. The average test score of 69.58 indicates that these children had an achievement level that resembles the math level that children should have at the end of grade 2 (see Table A1, rows 7 and 8).

In a similar fashion, we can also examine if test scores of children who were enrolled in other grades were at the appropriate level. On average, we find for the eligible students that the achievement levels are below the appropriate achievement levels, which is in line with the program design.

In the main text we have not elaborately discussed if children in the experimental group are at the appropriate achievement levels, because this is not of importance for the evaluation of the PfS program, given that children are randomly assigned to the PfS group and the control group. We do, however, emphasize in Section 4 that the achievement levels of children in the experimental group are below the appropriate achievement levels and that the focus of the program on children who are labeled as underachievers can influence the size of program effect that we will find.

Appendix 2

In this appendix, we show the correlations between the reading and math post-test scores and the pre-test scores in February and September in Table A2. The table shows for each subject that both pre-tests correlate highly with the post-test. Moreover, it shows that the pre-test in February 2011, which is used to assign children to the PfS program, correlates better with the post-test scores than the September pre-tests.

Table A2. Pre- and post-test correlations.

	Math			Reading		
	Post-test	Pre-test February	Pre-test September	Post-test	Pre-test February	Pre-test September
Post-test	1.00			1.00		
Pre-test February	0.77	1.00		0.72	1.00	
Pre-test September	0.70	0.70	1.00	0.65	0.59	1.00

Downloaded by [KU Leuven University Library] at 06:13 03 March 2014