Learning about the Enforcement of Conditional Welfare Programs: Evidence from the Bolsa Familia Program in Brazil*

Katja Maria Kaufmann
Bocconi University and IGIER

Eliana La Ferrara
Bocconi University and IGIER

Fernanda Brollo
University of Alicante

June 2013

Abstract

We study the implementation of a large-scale government program, "Bolsa Familia" in Brazil, which conditions transfers to poor families on children's school attendance. In particular, we attempt to disentangle the role of formal rules from the *enforcement* of these rules by analyzing whether and how people learn about the quality of enforcement and how this affects their behavior. We find that individuals finetune their behavior in response to signals about the quality of enforcement. Children's attendance increases after the family receives a warning or punishment for failure to comply in the past. The attendance response, on the other hand, is attenuated when punishments are delayed or waived with ad hoc justifications. We show that families learn both from *own* signals, observing the consequences of own noncompliance, and from *peers'* signals, that is, observing the consequences of peers' noncompliance. Enforcement thus has important spillover effects on other families, who learn from the experiences of their children's peers.

*We thank Josh Angrist, Oriana Bandiera, Esther Duflo, Eric Edmonds, Erika Field, Maitreesh Ghatak, Caroline Hoxby, Dan Keniston, Asim Khwaja, David Lam, Leigh Linden, Karthik Muralidharan, Rohini Pande, Imran Rasul, Paul Schultz, Jeff Smith, Chris Udry, Gerard Van den Berg, Eric Weese, conference and seminar participants at Dartmouth, EUDN, Harvard, Michigan, MIT, NBER, Namur, Paris School of Economics, Toulouse School of Economics, Uppsala, Universitat Pompeu Fabra and Yale for helpful comments. E-mail: katja.kaufmann@unibocconi.it; eliana.laferrara@unibocconi.it; fernanda.brollo@gmail.com.

1 Introduction

Evaluations of government policies or programs provide evidence on the combined effect of the formal rules of a policy (or program) and of the quality of implementation of these rules. To improve the design of policies or to make predictions about the success of a policy in a different context, it is crucial to be able to disentangle the role of formal rules from the role of enforcement. The aspect of implementation (such as the quality of enforcement) has received very little attention in the literature and this paper aims to provide evidence on the importance of this dimension.

In this paper, we analyze whether people learn about the enforcement of program rules, from own experience or from the experiences of their peers, and whether they adjust their behavior accordingly. In particular, we study the implementation of the large-scale government program "Bolsa Familia" in Brazil, which conditions transfers to the poor on school attendance of their children. The decision that these poor households are facing is whether or not to comply with the program conditions in each of the periods. On the one hand, not complying has some benefits, but on the other hand it is costly in that that there is some chance that the government detects the misdeed and punishes the household (e.g. in terms of withholding transfers). How households behave depends on their beliefs about the government's behavior.

In the first part of the paper, in which we analyze the importance of learning from own experience, we pursue two goals: After illustrating that the enforcement of "Bolsa Familia" has an important random component, we first show that this randomness is perceived and taken into account by program recipients. We demonstrate this by showing that individuals react to the arrival of warnings. If individuals thought that the implementation of the program was deterministic, the arrival of a warning would be perfectly anticipated (at the time of noncompliance) and should thus not trigger any changes in behavior.

Our second (and main) goal is to establish that individuals' reaction to the arrival of a warning is –at least in part– due to the fact that they learn about the implementation of the policy. In particular, we need to show that households not only react to warnings because the arrival of the warning means that they learn the realization of the government policy, but also because they update their beliefs about the behavior of the government. To put it differently, we want to show that people do not only learn the "outcome of a lottery in which they participate (realization effect)", but they also learn about the "distribution of the lottery (updating effect)". For this purpose, we conduct the following test: If the households' reaction is only due to the "realization effect", then the reaction should not depend on *when* the warning has arrived (conditional on the time left in the program). The timing of the arrival of the warning should instead matter, if the households learn about the implementation of the program. We find important evidence of the latter, i.e. households respond less if they receive a warning late, which suggests that they update their beliefs about the implementation (i.e. about the delay of warnings) and adjust their behavior. To provide further evidence on the fact that people learn about the enforcement of the program, and not –for example– about nonattendance of the children, we show that we find the same patterns if we focus on schools that inform parents about children's nonattendance directly (i.e. independently of Bolsa Familia).

In the second part of the paper, we analyze whether people learn from experiences of their peers. We start by testing whether, controlling for a child's own warning stage, warnings received by her peers (BFP recipients in the same school, grade or class) increase her attendance. We find that they do, and we provide evidence that this is not driven by shocks or conventional peer effects. In particular, while warnings arrive with a large variation in delays, we can show that people respond exactly in the period in which their peers receive their warning, and not –for example– one period earlier. At the same time, there is no reason to assume that shocks should affect individuals always exactly at the time when their peers get warned, but not one period earlier. We address the concern of conventional peer effects driving these results by controlling directly for peers' attendance behavior. We also exploit the fact that a child's siblings are typically in a different grade, and sometimes in a different school (depending on age difference), to test whether warnings received by a sibling's peers also affect a child's attendance. Thus we can control for the warnings of peers in the child's own class and test whether warnings of siblings' peers have an additional effect. We find that signals received by siblings' peers matter, which lends further support to the hypothesis that families learn about strictness of enforcement.

To summarize, our findings suggest that the quality of enforcement of government policies is important. People learn about the enforcement, not only from own experience but also from their peers, and adjust their behavior accordingly. We find evidence of important spillover effects in the learning process about quality of enforcement.

In our analysis we make use of administrative data on BFP from the Brazilian Ministry of Social Development (MDS) and the Ministry of Education (MEC). We have information about child and family characteristics, as well as monthly data on school attendance of each child, monthly information on warnings –from which we create the complete warning history of each family– and monthly information on the benefit the family received (or should have received) and whether the benefit was suspended in a given month. To study how families learn from others about the enforcement of program conditions, we merge the administrative data described above with School Census data, so that for each child in the family we can identify their BFP peers who are in the same school, grade and class.

In the light of the fact that enforcement of government policies is likely to often be imperfect which can have important effects on the effectiveness of such policies, it is even more surprising that we know very little about the implementation side of policies and on whether individuals' are aware of the quality of enforcement and adjust their behavior according.

One of the few papers that discusses the implications of imperfect enforcement is Banerjee, Glennerster, and Duflo (2008). They analyze the effects of providing incentives for the presence of nurses in government public health facilities, to address the problems of high staff absence in the public Indian health care system. The presence of government nurses in the health facilities was recorded by an NGO, and the government took steps to punish the worst delinquents. Banerjee, Glennerster, and Duflo (2008) find that, initially, the monitoring system was extremely effective, which shows that nurses are responsive to financial incentives. But after a few months, the local health administration appears to have undermined the scheme from the inside by letting the nurses claim an increasing number of "exempt days". They conclude that eighteen months after the inception, the program had become completely ineffective.

In this paper, we aim to shed light on the process by which individuals' learn about quality of enforcement, leading them to adjust their behavior.

The few papers in the literature studying the effect of enforcement on people's behavior are in the areas of unemployment insurance, compliance with TV license fees and crime deterrance.

Van den Berg, van der Klaauw, and van Ours (2004) analyze a novel set of policy tools in the area of unemployment insurance, namely monitoring the job search effort of unemployed workers and punishing them financially if they do not meet the effort requirements. In particular, the authors study the causal effect of a punitive sanction (in the form of a benefit reduction) on the transition rate from unemployment to employment using a unique data set on welfare recipients in the Netherlands. They find that the imposition of sanctions substantially increases the individual transition rate from welfare to work (see also Van den Berg and van der Klaauw (2006)).

Apart from the fact that this paper analyzes the behavioral response to punishment in a very different context –very poor families in Brazil and whether they send their children to school to comply with the conditions of a conditional cash transfer program–, we investigate whether people learn about the likelihood of enforcement. In particular, we show that people anticipate punishments and update their perceptions about the probability of being punished by learning from own experience and from peers' experiences.[1]

One of the very few other papers that is interested in the implications of enforcement is Rincke and Traxler (2011), who analyze the importance of spillovers from enforcement of compliance with TV license fees. Microdata on compliance with TV license fees allow them to distinguish between households that were subject to enforcement and those that were not. Using snowfall as an instrument for local inspections, they find a striking response of households to increased enforcement in their municipality: on average, three detections make one additional household comply with the law.

In contrast to this paper, our paper focuses on analyzing the learning process with which individuals update their beliefs about the strictness of enforcement using a large panel data set with monthly data on attendance (compliance) and warnings (enforcement). We analyze the importance of learning from own experience as well as from individuals' peers who are in direct contact (attend the same class).

Another empirical paper that are interested in whether and how people learn about enforcement are in the areas of crime deterrence: Lochner (2007) investigates the effect of learning about arrest probabilities on criminal behavior. In particular, he analyzes the effect of own and siblings' criminal history and arrests on perceptions about arrest probability (using data on people's subjective beliefs) and estimates the effect of perceptions about arrest probabilities on criminal behavior in a structural model.

Since we analyze the implementation of a (large-scale) conditional cash transfer program (CCT), the "Bolsa Familia" program in Brazil, our paper is related to a large literature on CCT programs: for papers on Progresa/Oportunidades see, e.g., Attanasio, Meghir, and Santiago (2011), Attanasio, Meghir, and Szekely (2003), DeBrauw and Hoddinott (2010), De Janvry and Sadoulet (2006), Schultz (2004) and Todd and Wolpin (2006), on Bolsa Escola (the predecessor program of "Bolsa Familia") see, e.g., Bourguignon, Ferreira, and Leite (2003), Bursztyn and Coffman (2010), De Janvry, Finan, and Sadoulet (2009). There is very little evidence on the Bolsa Familia Program (one notable exception is Bastagli (2008)), most likely due to the fact that BFP was not implemented in a randomized fashion and is therefore more difficult to evaluate. At the same time the BFP has

[1] Related to the aspect of anticipating "treatment", Van den Berg, Bergemann, and Caliendo (2009) show that if people perceive a high probability of a treatment, they modify their behavior. Crepon, Ferracci, Jolivet, and van den Berg (2012) test for anticipation of future treatment upon receiving information, in a general dynamic evaluation framework. In particular, the authors study effects of notifications of future participation in training programs and find that notification has a significant effect on the treatment probability (positive) and on the probability to leave unemployment (negative).

several very interesting program features that are distinct from other well-known CCTs, such as different warning stages and conditions being linked to joint attendance of all school-aged children.[2]

In contrast to these papers, we are one of the very few papers addressing the enforcement aspect of conditional cash transfer programs. There are a few papers that analyze the enforcement of eligibility criteria (i.e. errors of type one and two in terms of families included in or excluded from the program, see, e.g., Cameron and Shah (2011) who analyze the effects of mistargeting on social capital and crime), but –to the best of our knowledge– not on enforcement of program conditions for the families in the program and how families adjust their behavior in response to updating their beliefs about the strictness of enforcement.

Our paper is also related to the literature on peer effects and social interactions, even though our focus is not to identify conventional or "direct" peer effects (such as, for example, the effect of non-attendance of peers on own attendance decisions). Instead we are interested in identifying the effect of peers receiving warnings (as a signal about the quality of enforcement) on an individual's attendance decisions. This has the advantage that we can identify learning more directly and can avoid several problems that the conventional peer effect literature has to face (such as the reflection problem etc). In particular, in the case of "conventional" peer effects, it is often difficult to disentangle whether peer effects are due to the fact that an individuals learns from peers' behavior (for example about the value of schooling) or due to the fact that peers' behavior enters an individual's utility function (for example, a child likes going to school more if her friends attend as well).

Another paper interested in the diffusion of information is Banerjee, Chandrasekha, Duflo, and Jackson (2012), who analyze the diffusion of participation in a microfinance program through social networks. They exploit exogenous variation in the importance (in a network sense) of the people who were first informed about the program and estimate structural models of diffusion in order to (i) determine the relative roles of basic information transmission versus other forms of peer influence, and (ii) distinguish information passing by participants and non-participants. They find that participants are significantly more likely to pass on information than informed non-participants, but that information passing by non-participants is still substantial and significant. Lastly, they find that, conditioned on being informed, an individual's decision is not significantly affected by the participation of her acquaintances.

Our paper analyzes learning in a very different context, that is we are interested in whether people learn about the enforcement of government policies and thereby update their beliefs about the "costs of non-compliance". This is important from a policy-perspective to understand what are the costs involved with imperfect ("weak") enforcement, which governments trade-off with the benefits of weak enforcement (such as lower costs in terms of monitoring and enforcing, lower welfare costs in terms of taking away/reducing poor households' transfers and –possibly– political benefits if lack of enforcement is used to win over voters).

Other examples of paper that are interested in the role of information and social interactions, but in the context of retirement plan decisions, is Duflo and Saez (2003). Another related paper on learning from peers in the context of technical change in agriculture is Foster and Rosenzweig (1995). Some examples of the related literature on peer effects and social interactions more generally are Aizer and Currie (2004), Ding and Lehrer (2006), Duflo, Dupas, and Kremer (2010),

[2] In terms of literature on schooling in Latin America more generally, see, for example, Harbison and Hanushek (1992) who discuss programs to improve school quality of the poor in the Northeast of Brazil.

Hanushek, Kain, Markman, and Rivkin (2003), Hoxby (2000), Hoxby and Weingarth (2005) and Sacerdote (2001). Two survey articles that provide an overview and synthesis of the literature on how social networks influence behaviors are Jackson and Yariv (2010) and Blume, Brock, Durlauf, and Ioannides (2010).

The following papers analyze peer effects in the context of conditional cash transfer programs in Colombia and Mexico: Barrera-Osorio, Bertrand, Linden, and Perez-Calle (2011) analyze a randomized CCT in Colombia with three different treatments based on school attendance and randomize on the student level, which allows them to generate intra-family and peer-network variation. They find evidence that siblings (particularly sisters) of treated students work more and attend school less than students in families that received no treatment. The authors also find that indirect peer influences are relatively strong in attendance decisions with the average magnitude similar to that of the direct effect. Other papers that analyze peer effects and social interactions in the context of Progresa/Oportunidades are Angelucci, Giorgi, Rangel, and Rasul (2010) on family networks, Angelucci and Giorgi (2009) on consumption spillovers and Bobonis and Finan (2009) on peer effects in schooling.

The outline of the paper is as follows: In Section 2 we provide background information on the Bolsa Familia program, in particular on program conditions and the mechanisms of enforcement. In Section 3 we present the data and descriptive statistics. In Section 4 we discuss our empirical strategies, firstly for analyzing whether individuals learn about enforcement from own experience and secondly for analyzing whether individuals learn from their peers. In Section 5 we discuss our results on learning from own experience and peers. We discuss robustness of these results in Section 6 and Section 7 concludes.

2 Background Information on the Bolsa Familia Program

The Bolsa Familia Program reaches around 11 million Brazilian families, that is 46 million poor people (equivalent to 25% of the Brazilian population) with a budget of over 12 billion reais (USD 6 billion). Thus Bolsa Familia reaches nearly three times as many people and is about three times as large in terms of budget as the well-known conditional cash transfer program Progresa/Oportunidades.

The Bolsa Familia Program (BFP) results from consolidating four different programs (Federal Bolsa Escola Program, Bolsa Alimentacao, Auxilio Gas, Fome Zero) into one single program (see, for example, Lindert, Linder, Hobbs, and de la Brire (2007)). It was launched by Lula in 2003. The election-free year 2005 was used as year of consolidation to strengthen the core architecture of the program (20 legal and operational instruments were issued, institutionalizing various aspects of the program and its decentralized implementation) and to strengthen the program's registry of families (Cadastro Unico) and the monitoring of conditionalities. In 2006, the Ministry of Social Development embarked on initiatives to promote further vertical integration with sub-national CCTs and integrated the conditional transfers paid under the Child Labor Eradication Program (PETI) into the BFP. The Ministry started providing incentives for quality management and rewarding innovations in the decentralized management of the BFP. The agenda for 2007 and beyond was to further strengthen the basic architecture of the program, particularly in the areas of monitoring and verification of conditionalities, to strengthen the oversight and control mechanisms and continued improvements on the program's targeting system.

The targeting of the program was conducted in two steps. The first step consisted of geographic targeting on federal and municipality level. The federal government allocated BFP quotas to municipalities according to estimates of poverty. [3] Within municipalities spatial maps of poverty were used to identify and target geographic concentrations of the poor. The second step was to determine eligibility at the household level. Eligibility is determined centrally by the Ministry of Social Development (MDS) based on household registry data that is collected locally and transmitted into a central database known as the Cadastro Unico.

In the following we will discuss three of the key aspects of the program that are relevant to our analysis. The first is related to the magnitude and importance of the transfer amounts for the eligible families, the second to the conditions that families have to comply with to avoid "punishment" and the third concerns the different degrees of "punishments" that families have to incur in case of noncompliance.

Firstly, BFP provides two types of benefits (transfers): basic and variable benefits which depend on family composition and income. Families with a monthly per capita family income of up to R$60 (US$30) are classified as "extremely poor", families with between R$60 and R$120 are classified as "moderately poor". The base benefit is provided only to families in extreme poverty, regardless of their demographic composition. Both extremely poor and moderately poor families receive a variable benefit which depends on the number of children in the family (capped at three to avoid promoting fertility) and on whether the mother is pregnant or breast-feeding.

The benefit amounts are as follows in the period of analysis, 2008 and 2009: From January 2008 to June 2008, the base benefit amounts to R$ 58 (approximately US$ 30, which is the per capita income threshold for the extremely poor), the variable benefit to R$ 18. Between June 2008 and July 2009, the amounts are R$ 62 and R$ 20 for base benefit and variable benefit, respectively, and from August 2009 to March 2011, the base benefit is R$ 68, the variable benefit R$ 22.

To illustrate the magnitude of the program transfers, think –for example– of a family with three children that earns is right at the threshold to be classified as extremely poor (i.e. they have a monthly per capita income of R$60 and a family income of R$ 300). This family will receive monthly transfers of around R$120, which amounts to 40% of their total family income. For poorer families the transfer amounts make up an even larger fraction of their total family income.

Secondly, the BFP cash transfers are conditional on all age-relevant family members complying with school attendance "conditionalities". Each school-aged child has to attend at least 85% of school hours each month (absence due to health reasons is justified and does not count towards the number of absent days). This element of "joint responsibility" for children in the same family is a quite unique feature of BFP (compared to other well-known CCTs such as Progresa/Oportunidades).

Thirdly, the consequences of non-compliance in the different stages of conditionality are as follows. In the first case of non-compliance the family receives a warning without financial repercussion. With the second warning, benefits are blocked for 30 days after which the family receives the accumulated benefit of the previous and the current month. The third and fourth warning lead to a loss of benefits for 60 days each time. After the fifth warning, the benefit is canceled and the

[3] Original municipal level allocations were established by comparing eligibility criteria (per capita income thresholds of R$ 100) to the 2001 national HH survey (PNAD) combined with 2000 Census. In 2006 MDS adjusted overall targets as well as specific program quotas for municipalities using PNAD 2004. Revisions resulted in adjustments of specific municipal quotas with some municipalities facing a reduction in BFP allocation (only few municipalities had beneficiary totals that were above the newly established quotas, but nobody was cut out of the program).

family loses eligibility (according to the general rules, the family can reapply to the program 18 months later, for exceptions see description below).

Families are well informed about transfer amounts, conditionalities and punishments in the case of non-compliance: Bolsa Familia transfer amounts and conditionalities are being widely and regularly publicized on TV, radio and newspapers in Brazil, and are spelled out in a booklet issued to each beneficiary family (Agenda de Compromissos). In case of noncompliance, a family receives a warning message when withdrawing their transfer money at the bank. At the same time, they are reminded about the warning stage they are in and the punishment that they may have to incur in the next case of noncompliance (for example, one possible warning message can be translated as "if you fail to comply again, your money might be suspended").

The implementation of this conditionality scheme involves different actors. First of all, children's hourly attendance is recorded by school teachers. The school sends the attendance lists of students to the municipality: they report the exact fraction of school hours attended, in case attendance was below 85%, otherwise they only report that the student complied. Each municipality collects the lists and sends them to the Ministry of Education (MEC), which determines whether a family complied or not in a given month, i.e. whether all the children between 6 and 15 have attended at least 85% of days. They send detailed reports to the Ministry of Social Development (MDS) which establishes which warning the family should receive in case of noncompliance and whether the family is entitled to the transfer for that month (based on the warning stage reached). The MDS sends this information to the Caixa, which transfers the benefit amount to the bank account of the family if the family is entitled to receiving the transfer for that month.[4]

The family can withdraw their transfer money starting at a pre-specified date each month with a Bolsa Familia "electronic benefit cards". At the time of withdrawal, the family receives a message, which declares the warning stage, the month of failure to which the warning refers, the names of the child(ren) who failed and which type of warning they might receive in the next instance of non-compliance.

These different steps of the process explained above involve some time and can lead to a significant delay in terms of the month in which the warning is received compared to the month of failure (for an illustration, see Figure 1, which displays the Brazilian school calendar that coincides with the calendar year). The median time of delay is five months, but the delay ranges between two and ten months.

The following exemptions from the standard procedure can be observed. Firstly, the school can justify insufficient attendance in a given month ex-post (which we observe in the data). In that case the child will count as having complied with the rules in that month. Secondly, while according to the general rule a family who receives the fifth warning has to leave the program and can only be admitted back after 18 months, the municipality has the right to readmit the family more quickly in case they start complying after the fifth warning.

[4]The Caixa Econmica Federal has been contracted as BFPs operating agent. The Caixa is a federal savings/credit union organization, which –apart from banking services– has traditionally provided payments issuance services for federal assistance programs, because of its' broad network which guarantee its' presence in all Brazilian municipalities (it operates over 2,000 agencies nationwide, and is linked with close to 9,000 lottery points and over 2,000 banking correspondents). The Caixa consolidates and manages the national registry database for social programs, the Cadastro nico, assigns registered individuals the unique Social Identification Number (NIS), and makes payments directly, crediting beneficiaries electronic benefit cards (EBCs) on a monthly basis through its extensive banking network. The Caixa also designed and operates the software currently used by the Ministry of Education (MEC) for consolidating the information resulting from the monitoring of compliance with conditionalities.

3 Data

We make use of administrative data on the Bolsa Familia program from the Brazilian Ministry of Social Development (MDS), which contains the following information: The household registry data (Cadastro Unico) contains extensive information on families' background characteristics and on each individual member of the household, such as age, gender, race, marital status, education, employment status and occupation of each adult household member, per capita expenditures, ownership of durable goods, schooling history of each child and so forth.

We have information on monthly school attendance of each child for 2008 and 2009 (that is we know the exact fraction of hours failed below the threshold of 85% and otherwise we only know that the condition was met), monthly information on warnings (so we can create the complete warning history of each family) and monthly information on benefits that the family received (or should have received) and whether the benefit was blocked or suspended in a given month.

A third source of data is the School Census, which contains information on all children who are enrolled in a given school. We merge the administrative data described above with School Census data (merge based on social security number of the child –if available– and otherwise based on area code, school code, grade, full name and birth date of child), so that we can identify the (BFP) peers who are in the same school, grade and class for each child in the family. Also this data provide information on whether a child passed or failed a given grade.

Table 1 presents summary statistics of the variables used in our analysis.

4 Empirical strategy

4.1 Response to Own Experience of Enforcement

In this first part of the paper, in which we analyze the importance of learning from own experience, we pursue two goals: After illustrating that the enforcement of "Bolsa Familia" has an important random component, we first show that this randomness is perceived and taken into account by program recipients. Our second (and main) goal is to establish that individuals' reaction to the arrival of a warning is –at least in part– due to the fact that they learn about the enforcement of program conditions.

To give a brief overview over our empirical strategies before discussing them in more detail below: First, we demonstrate that people perceive the randomness of enforcement by showing that individuals react to the arrival of warnings. If individuals thought that the implementation of the program was deterministic, they should anticipate the arrival of a warning at the time when they decide not to comply in a given month. The actual arrival of the warning (several months later) should thus not trigger any (further) changes in behavior. Below we will discuss in more detail how we establish that individuals respond to the arrival of warnings (and thus perceive the randomness of enforcement).

Second, we provide evidence that individuals' reaction to the arrival of a warning is –at least in part– due to the fact that they learn about the implementation of the policy. In particular, we need to show that households do not only react to warnings because the arrival of the warning means that they learn the realization of the government policy, but also because they update their beliefs about the behavior of the government. To put it differently, we want to show that people do not only learn the "outcome of a lottery in which they participate (realization effect)", but they also learn about the "distribution of the lottery (updating effect)".

Lastly, to provide further evidence for the hypothesis that people learn about the enforcement of the program and not –for example– about nonattendance of the children, we analyze if results differ for schools that inform parents about children's nonattendance directly (i.e. independently of Bolsa Familia).

The outcome variable we use in the following analysis is whether a child fails to comply in a given month (i.e. attends less than 85% of school hours in a given month) or not. For ease of exposition, we refer to this variable as "failure" and we study the probability to "fail", but it should be clear that this term is not used in the traditional meaning of grade repetition, but rather it refers to failure to comply with the conditionality embedded in BFP.

In the following, we discuss the empirical strategy we use to show that individuals react to the arrival of warnings (and to thereby establish that they perceive the randomness of the program implementation). We analyze a situation in which a household fails to comply in a given month and then receives a warning several months later. We are interested in whether (i.) the only response of the household happens (immediately) after the month of noncompliance (but before the arrival of the warning), e.g. the likelihood to fail goes down after the month of noncompliance because of mean reversion or because the household anticipates the arrival of a warning and wants to comply from the moment of failure onwards to avoid progressing further in warning stage. Or whether (ii.) there is an additional response of the household upon arrival of the warning (i.e. a further reduction in the likelihood of failure), suggesting that the warning was not fully anticipated and thus showing that households do perceive randomness in terms of program implementation.

First, we show graphically how households' compliance behavior (i.e. school attendance) changes over time before and after the arrival of a warning.

Second, we conduct an analogous analysis in a regression framework to control for other changes and to test for significance. In particular, we regress our outcome variable "failure to comply" on dummies for $\tau = 0, ..., T$ months before and after the month of warning, $D_{W\tau}$, and estimate the following specification:

$$Y_{ht} = \sum_{\tau=-T}^{T} \beta_\tau D_{W\tau} + \gamma X_{ht} + D_t + D_h + \epsilon_{ht}, \tag{1}$$

where h denotes the household, t the month, Y is a dummy equal to 1 when the child attends less than 85 percent of the school hours in a given month ("failure"); X_{ht} is a vector of household level controls including number of boys and girls in different age brackets (0-5, 6-10, 11-15, 16-18); D_t denotes month and year fixed effects to control for example for seasonality effects; D_h denotes household fixed effects and ϵ is the error term. We estimate Equation (1) using a linear probability model and show that the coefficients β_τ are positive before the arrival of the warning, i.e. for $\tau = -T, ..., 0$, and negative after the arrival of the warning, i.e. for $\tau = 1, ..., T$. Thereby we show that the likelihood of failure goes down significantly upon arrival of a warning.

Third, we include dummies for months after failure, $D_{F\theta}$ for $\theta = 1, ..., T$, to control for possible mean reversion effects and the possibility that the household fully anticipates the arrival of the warning at the time of failure and decreases the likelihood of noncompliance immediately after the month of failure. In particular, we estimate the following specification:

$$Y_{ht} = \sum_{\tau=-T}^{T} \beta_\tau D_{W\tau} + \sum_{\theta=1}^{T} \beta_\theta D_{F\theta} + \gamma X_{ht} + D_t + D_h + \epsilon_{ht}. \quad (2)$$

As above we want to show that the coefficients β_τ are positive before the arrival of the warning, i.e. for $\tau = -T, ..., 0$, and negative after the arrival of the warning, i.e. for $\tau = 1, ..., T$, also after controlling for possible mean reversion effect and thus to show that households' decrease the likelihood of failure upon receiving a warning.

Fourth, we regress our outcome variable "failure to comply" on the warning stage that households are in, since the warning stage determines the cost of noncompliance. One challenge in the identification of the effects of warnings stems from the fact that some families have a much higher propensity to fail and these families are the ones who reach high warning stages. Using cross-sectional variation in the data, this would bias the (negative) coefficient on warnings towards zero or even induce a positive correlation between "receiving warnings" and "failure to comply".

We address this problem using family fixed effects in the rest of the analysis to control for time-invariant differences in the propensity to fail. This implies that we use variation within family over time, for example in terms of receiving warnings, to identify their effect on attendance behavior each month.

Our baseline specification for estimating the effects of own warnings is:

$$Y_{iht} = \sum_{k=1}^{5} \alpha_k WS_{ht}^k + \gamma X_{iht} + D_t + D_h + \epsilon_{iht} \quad (3)$$

where i denotes the child, h the household, t the month, Y is a dummy equal to 1 when the child attends less than 85 percent of the school hours in a given month ("failure"); WS^k denotes a dummy equal to one if the household in in warning stage k (with $k = 1, ..., 5$), X is a vector of child level controls including gender, age, number of brothers and sisters in different age brackets (0-5, 6-10, 11-15, 16-18); D_t denotes month and year fixed effects to control for example for seasonality effects; D_h denotes household fixed effects and ϵ is the error term. We estimate Equation (3) using a linear probability model and show that the coefficients α_k are negative, i.e. warnings and failure to comply are negatively correlated as expected.

Note that we do not plan on comparing the effects of different warning stages in this paper, since the comparison of the effect of different warning stages poses the following challenge: When estimating the coefficients on different warning stages α_k ($k = 1, ..., 5$) for the full sample, those coefficients are estimated by averaging over different subsets of families. For example, the coefficient on warning stage 5 is only estimated for those families who ever get to warning stage 5 in our two-year period of observation, while the coefficient on warning stage 1 is estimated over the set of families who receive at least one warning. Since the comparison of the effects of different warning stages is not important for the analysis of learning, we will address this issue in a companion paper.

To show that we really identify a causal effect of warnings on attendance behavior, we use exogenous variation in the timing of the warning. In particular, we make use of the fact that the exact day of the receipt of warnings depends on the last digit of the social security number of the legally responsible adult of the family, which is basically randomly assigned.[5]

[5] As explained in Section 4.2.1, BFP benefits are not transferred to all families on the same day to avoid large numbers of BFP recipients come to the bank (or other cash points) on exactly the same day and to avoid stock outs

We test the following hypothesis: If families' responses are due to the warnings they have received and not due to something else happening in that month (such as shocks), then those families who receive warnings earlier in the months should show a larger behavioral response than families who receive warnings later. Imagine, for example, that person A receives a warning at the beginning of the month, so she can easily adjust her behavior in that month ("early warning"). Person B on the other hand receives the warning on the last day of the month ("late warning"), so she only adjust her failure behavior if her failure in that month up to that day was just below the threshold.

In practice, we interact warning stage with dummies for "early warning and "late warning" and use two different definitions: Definition 1 splits the period in which benefits are transferred (and thus warnings received) into halves, while definition 2 is constructed using a split of 70:30 (to allow for an even lower possibility to adjust for "late" recipients).

Once we have shown that there is a causal effect of warnings on subsequent attendance behavior –i.e. people perceive uncertainty in terms of the strictness of enforcement–, we still need to disentangle different possible mechanisms leading to this result. In particular, we want to understand if the response is due to a "realization" effect, due to "updating" beliefs about strictness of enforcement or learning about non-attendance of the children.

To exclude the possibility that the response to warnings is entirely due to learning about nonattendance of the children, we analyze the response to warnings of households who are informed directly about nonattendance of their children by school. For that purpose, we merge our administrative data with a dataset on schools called PROVA. Since PROVA was conducted only in grades five to nine of public schools, we can merge less than half of our sample. The variables we make use off to capture whether parents receive information about nonattendance of their children from schools (i.e. independently of BFP warnings) are as follows: Schools inform parents (a) in writing, (b) in individual meetings, and (c) by visiting parents at home.

Lastly, we want to provide evidence that individuals' reaction to the arrival of a warning is –at least in part– due to the fact that they learn about the implementation of the policy. In particular, we need to show that households not only react to warnings because the arrival of the warning means that they learn the realization of the government policy, but also because they update their beliefs about the behavior of the government. For this purpose, we conduct the following test based on household's reaction to warnings that arrive with different delays (conditional on the time remaining in the program): If the households' reaction is only due to the "realization effect", then households should react *more* strongly to late warnings for the following reason: Given a distribution over delays and a certain probability of not receiving a warning at all, the longer the delay, the more likely it is that the household will not be warned at all. Thus the household becomes less careful in terms of compliance. The (late) arrival of a warning comes as a surprise and households react strongly by decreasing the likelihood to fail. In the case of learning or 'updating' the beliefs about the strictness of enforcement on the other hand, households will react *less* strongly to late warnings, since the long delay is a signal of weak enforcement (which is relevant for future non-compliance).

---

and price increases due to spending being concentrated on that day. Instead, the last digit of the social security number of the person that is legally responsible for the children (and who receives the BFP-card to withdraw money) determines, on which exact day within a month the benefit is transferred to a families' bank account, and thus on which it can be withdrawn. At the time of withdrawing money, the family also receives the warning message.

We estimate:

$$Y_{iht} = \sum_{k=1}^{5} \alpha_k WS_{ht}^k + \sum_{k=1}^{5} \beta_k WS_{ht}^k * DEL_h^k + \gamma X_{iht} + D_t + D_h + \epsilon_{iht} \tag{4}$$

where DEL_h^k is the number of months (in deviation from the median of 5 months) that passed between the month in which the attendance condition was not met and the month in which household h was notified of the punishment in the form of transitioning to warning stage k. We expect $\beta_k > 0$, i.e. delays in enforcement should be associated with a smaller reduction in failure for a given warning, compared to timely enforcement. In other words, warnings received with longer delay are less effective.

Before moving on to discussing strategies of identifying the importance of learning from peers, let us also consider an "extreme"form of lax enforcement, i.e. the ex post justification given by the school when a child failed to attend, so that the household does not advance in warning stage. We estimate:

$$Y_{iht} = \sum_{k=1}^{5} \alpha_k WS_{ht}^k + \sum_{n=1}^{3} \beta_n JUST_{h,t-n} + \gamma X_{iht} + D_t + D_h + \epsilon_{iht} \tag{5}$$

where $JUST_{h,t-n}$ is a dummy equal to 1 if household h has been justified in month $t-n$, with n going from 1 to 3, i.e. we consider three lags of this variable. Again, we expect $\beta_n > 0$, i.e. to the extent that justifications are interpreted as signals that failure to comply with conditionality is not punished, justifications should be associated with a relatively higher probability of failing to attend school.

At the same time, serially correlated shocks (e.g. health shock to a parent) could lead to the same pattern. For that reason, we only want to analyze if the correlation is positive as expected and instead of interpreting this correlation as causal, we will postpone the analysis of justifications as signals of weak enforcement to the next section on learning from peers.

4.2 Response to Peers' Experiences of Enforcement

In this paper we want to analyze the learning process of households with respect to their beliefs about the strictness of enforcement of program conditions. We aim to understand the importance of different channels, such as learning from own experience and learning from peers. In this section we will discuss the empirical strategies to identify the latter effect. It is likely that families do not only learn from their own private signals of enforcement (e.g., if they receive warnings), but also from "public signals," that is from the experience of other children who receive warnings, lose their transfer or even have to leave the program. In particular, children who are in the same class, grade or school as the family's own children are likely to be key sources of information regarding the implementation of the program. We would expect the family to decide about attendance in each month based on this information. To discuss our empirical strategy it is useful to start from a simple benchmark specification, which is not exactly the one we estimate (see below), but which helps us highlight identification challenges:

$$Y_{iht} = \sum_{k=1}^{5} \alpha_k WS_{ht}^k + \beta \ PEERWARN_{iht} + \gamma X_{iht} + D_t + D_h + \epsilon_{iht} \tag{6}$$

where $PEERWARN_{iht}$ denotes several possible indicators for the presence of peers who receive a warning in month t. As a preview, we find a significant negative correlation between own failure to attend in a given month and the fraction of peers who are warned ($\widehat{\beta} < 0$), that is, the likelihood to fail decreases when peers receive warnings. Also we find a significant positive correlation between own failure to attend and the fraction of peers whose failure is justified ex-post.

We need to address several important identification challenges before we can interpret these correlations as learning. In what follows we group these identification challenges under two headings: correlated shocks and "conventional" peer effects.

After discussing the identification challenges and our strategies in some detail, we will discuss the interpretation of the results as learning about enforcement from peers. In particular, we will analyze how the effect of peer warnings of different severity vary depending on a family's own current warning stage. For example, we will test if the effects of peers receiving higher warnings than oneself are stronger than of peers receiving the same or a lower level warning (since the former warning contains more information, while the family already knew that the government enforced the warning stage that they had reached themselves).

4.2.1 Identification Challenges

Correlated shocks

The first threat to identification are correlated shocks that may directly affect a student and her peers, thus inducing a correlation between $PEERWARN_{iht}$ and ϵ_{iht} in Equation (6). Consider for example an economic shock leading to an increase in the opportunity cost of schooling in the area where individual i lives. In response to such a shock, both individual i and her peers would be more likely to fail and thus more likely to receive a warning. If the shock is persistent, the shock would still induce individual i and her peers to fail more several months later when the warning is received. However, this type of mechanism would generate a *positive*, not negative, correlation between one's failure and peers' warnings ($\widehat{\beta} > 0$), while what we find in the data is a negative correlation ($\widehat{\beta} < 0$). This negative correlation is consistent with a "scare"effect, i.e. student i decreases her failure to avoid the sanctions that her peers experienced, but not with positively correlated economic shocks. The shock could of course be mean reverting, which is a concern that we address below.

A different type of correlated shock may be linked to changes in school policies, e.g. the teachers or headmaster become stricter in registering students' non-attendance. This would lead to a higher number of peers receiving warnings. At the same time, everyone would respond by attending more (failing less), because it is known that the teacher or headmaster have become stricter, not because they extract signals on enforcement from peers' warnings. This type of correlated shock would thus generate an estimate with the same sign as we find, i.e. ($\widehat{\beta} < 0$).

To address this concern, we include the peer warnings as a lead variable as a falsification test, i.e. we include $PEERWARN_{ih,t+1}$ in Equation (6). Our reasoning exploits the time lag from the moment in which a failure in attendance occurs and the moment in which the student gets warned, which has a median value of 5 months. If children decrease their failure because of greater strictness, they will start attending more soon after they experience this greater strictness, which would be before the time in which warnings are received (e.g., before 5 months). This means that changes in attendance would precede warnings, so we should find a negative and significant coefficient on $PEERWARN_{ih,t+1}$. If not, then it is unlikely that the greater strictness that started 5 months earlier (median delay between failure and warning), has an effect exactly 5 months afterwards (when peers receive the warning) but not, say, 4 months afterwards. This strategy can also address the concern of mean reverting shocks: Again it is highly unlikely that a shock that increased failure 5 months ago, suddenly decreases failure with the exact same delay as the warning happens (but has no increasing effect on failure 4 months after the initial shock).

To summarize, the augmented version of Equation (6) that we estimate is:

$$Y_{iht} = \sum_{k=1}^{5} \alpha_k WS_{ht}^{k} + \beta \sum_{n=-1}^{+1} PEERWARN_{ih,t+n} + \zeta \sum_{n=-1}^{+1} PEERFAIL_{ih,t+n} + \gamma X_{iht} + D_t + D_h + \epsilon_{iht} \quad (7)$$

Conventional peer effects

The second concern is related to "conventional" peer effects. The mechanism would be as follows: A shock of some sort (for example a local economic shock or school level shock) increases children's likelihood to fail. Individual i's peers receive warnings and in response to their own warnings they fail less. As a consequence also individual i fails less because she observes her peers failing less and not because she extracts signals on enforcement from peers' warnings. This is what we call a "direct" or "conventional" peer effect as opposed to learning about the strictness enforcement from peers' warnings.[6] To address this concern, we proceed with four strategies.

First, we can control directly for the fraction of peers (in i's class) who fail, to analyze if warnings of i's peers have an independent effect on individual i's likelihood to fail. This makes it difficult for us to find any effect of peer warnings for the following reason. This strategy implies that we compare two classes with the same fraction of students failing, while in one class the fraction of children warned is larger. If people respond to own warnings, then children in the class with more warnings need to have a general tendency to fail *more* than the children in the class with fewer warnings. To see a learning effect, the individual should fail *less* in a class with more warnings, while –according to this argument– she is in a class where people are prone to fail *more*.

Second, we exploit warnings received not by individual i's own peers, but by the peers of i's siblings. In particular, we will show that a child's likelihood to fail responds not only to own peers' warnings, but also to siblings' peers who get warned, while controlling for own peers' warnings. If own peers' warnings were driven by school or grade-specific shocks, this strategies allows to control for such shocks (i.e. holding own peers' warnings constant) to see if siblings' peers' warnings still have an effect. We consider different specifications in which we take into account all siblings or, for example, focus only on siblings going to different schools (to rule out both grade and school-specific shocks). If families learn about enforcement, then each child's attendance should be influenced also by siblings' peers' warnings which contain information about the government's quality of enforcement.

Therefore, in addition to the fraction of individual i's own peers who are warned, we also include

[6] The direct peer effect could be due to preferences for school attendance that depend on the presence of peers, but also to interpreting peers' attendance as a signal that attendance is in some way beneficial (for example because of high returns to schooling, or strict enforcement and possible loss of benefits in case of non-attendance. To identify learning effects in a clean way, we control for direct peer effects in the first set of results, which of course might lead to an underestimate of the "true" learning effects, if part of the learning about enforcement happens through observing ones' peers attend more.

the maximum of the fractions of i's siblings' peers who got warned, as in the following specification

$$Y_{iht} = \sum_{k=1}^{5} \alpha_k WS_{ht}^k + \beta_1 PEERWARN_{iht} + \beta_2 MaxPEERWARN_{-i,ht} + \gamma X_{iht} + D_t + D_h + \epsilon_{iht} \quad (8)$$

where the subscript $-i$ indicates that the Max is taken at the household level excluding individual i's own peers.

In principle, we could include the variable capturing siblings' peers' warnings in different ways. The goal is simply to show that siblings' peers also matter even after controlling for warnings of one's own peers. We include siblings' peers' warnings in this way, since we believe that information transmission (about the quality of enforcement) is likely to be happening as follows: As discussed in Section families receive their warning message when they withdraw their transfer money at the bank. These warning messages –and even more so loss of transfers– are likely to be objects of discussion between children at school and between their parents. For example, one child might try to convince the other to skip school together. If the family of the second child has just received a warning (or even lost their transfer money), she might respond that she cannot skip school again, because her parents would be really angry if she is responsible for them losing further benefits. Parents are also likely to talk, among other things exactly for the purpose of understanding the risk of warnings and thus to warn each other about the strictness of enforcement. At the same time, there might of course be a certain level of stigma related to receiving warnings. Thus if a family is the only one who received a warning, children and parents might be less likely to mention the occurrence of a warning. If on the other hand, several families are in the same situation, warnings are much more likely to become a general topic of discussion and thus it is more likely that families receive a more accurate signal of the frequency of warnings, while a small fraction might not have been discussed and noticed at all. In other words, the higher the fraction warned, the more likely will this signal be close to the "true" fraction of warnings. Therefore, we will be analyzing the effect of the "toughest" signal of enforcement on children's attendance, as described below.

As discussed above, this approach is used not only to disentangle learning from direct peer effects, but also to rule out grade- or school-specific shocks, since we show that siblings' peers matter while controlling for own peers' warnings (which should indirectly control for grade/school-specific shocks, if those are important and driving the results).

Third, we might still be concerned about peer effects between siblings that could lead to the pattern discussed in the second approach above, as an alternative explanation to learning effects. In particular, if a large fraction of the peers of i's siblings receive warnings, then i's siblings might attend more (either because of the warning or due to peer effects) and then i might decide to attend more because her siblings attend more. The mechanism underlying this direct peer influence among siblings may be one of role models, or it may be related to time use (for example, when two brothers usually play soccer together and if one goes to school, the other one loses his soccer partner). It seems plausible that this type of peer effect between siblings should be stronger, the closer the siblings are in age and more likely if the two siblings are of the same gender.

To analyze the importance of this alternative explanation, we repeat the last test comparing the effect of fraction of own peers warned and of the maximum fraction of siblings' peers warned who are in schools *different* from i's school. In other words, we estimate a modified version of Equation (8) where the variable $MaxPEERWARN_{-i,ht}$ is constructed from siblings who attend a different school than i. Siblings usually go to different schools only if they differ enough in age and thus have to go to different types of schools (such as elementary, junior high school etc). If the learning story is true, then the "maximum fraction of siblings' peers warned who are in schools different to i's school" should have an effect that is as strong as the "maximum fraction of all siblings", since both contain information about enforcement which takes place at the federal level. If on the other hand, peer effects between siblings are driving the results, then we would expect the former to have a smaller effect, because in that case the siblings are further apart in age and thus less likely to be playmates.

This leads us to a closely related test, based on the idea that children of the same gender are also more likely to be playmates. Thus, we test if there is a differential effect of the "maximum fraction among the siblings" depending on whether the maximum is taken among siblings of the same gender or the opposite gender. If the learning story is true, the maximum should have the same effect independent on whether it is based on siblings of same or opposite sex. If the sibling peer effect story is true on the other hand, we would expect the maximum among opposite-sex siblings to have a smaller effect (as these are less likely to be playmates).

Fourth, to lend further credibility to the hypothesis that people learn from their peers, we use an additional strategy to address the concern of peer effects between siblings, that is we make use of the following feature of the implementation of BFP: Warnings and ex-post justifications of non-compliance are implemented at different institutional levels. The implementation of warnings (and loss of benefits) is in the hand of the Ministry of Social Development, i.e. done at the federal level. The ex-post justification of a child that failed to comply in a given month is in the hand of the school (teacher or headmaster). If the effects that we find are due to learning effects, then the attendance behavior of an individual should be affected by warnings of her siblings' peers who are in different schools (because it is a signal of enforcement quality at the federal level), but should not be affected by justifications of her siblings' peers who are in different schools (or –if justifications are correlated across schools– at least less affected). Instead both warnings and justifications should affect an individual's behavior, if they happen to peers who are in the same school (class) as the individual. If on the other hand, the effects we find are peer effects among siblings, then in both cases (with warnings and justifications) the peers of the siblings would have an effect on the individual via "sibling" peer effects, i.e. effects of warnings and justifications should be symmetric.

This last point is of course not only interesting to support our hypothesis of learning, but also because it shows that families learn not only from signals of *strict* enforcement, but also from signals of weak enforcement, and respond as expected.

4.2.2 Interpretation and Informational Content

In this Section we discuss our approach of comparing the effects of peers receiving warnings of different severity. In particular, we will analyze how the effects of different peer warnings vary depending on a family's own current warning stage. Our hypothesis is that, for example, for families who are currently in warning stage one, receiving information about peers who also received warning one contains less information than peers who receive higher warnings, since the family already received a direct signal that the government enforces in terms of sending first warnings.

For that reason we will test if the effects of peers receiving warnings higher than oneself are stronger than of peers receiving the same or a lower-level warning. More specifically, we will estimate four different regression equations, conditional on whether a family is currently in warning stage

1, 2, 3 or 4. In the equation we include two different peer variables, firstly the "fraction of peers with a warning stage *lower or equal* to family i's" and secondly the "fraction of peers with warning stage *higher* than family i's".

In terms of learning we expect that the second variable should have a larger coefficient than the first, since the second contains new information, that is that the government also implements warning stages higher than the one the family is currently in. On the other hand, observing peers receiving the same warning that the family already received in the past contains less information, since the family already knows that the government implements up to that warning.

5 Analysis

5.1 Response to Own Experience of Enforcement

5.1.1 Own Experience of Strict Enforcement

In this Section we analyze whether families perceive uncertainty with respect to the enforcement of program conditions and whether they update their beliefs about the quality of enforcement. For this purpose we first show that people respond to the arrival of warnings.

Our outcome variable, "failure", is a dummy equal to one if a child fails to comply in a given month, i.e. attends less than 85 percent of school hours in a given month, and 0 otherwise.

We start with presenting graphs of the evolution of "failure to comply" over time conditional on having received a warning in $t = 0$ (normalization). In addition we condition on having received the warning with a certain delay, where the delay can be between 2 and 10 months. Graph 2 illustrates the likelihood to fail for the case that the warning takes place with a delay of five months (the median and modal time of delay). Thus in $t = -5$, the probability of failure is one (since we have conditioned on being warned in $t = 0$ with a delay of five months). Between $t = -4$ and $t = 0$, the likelihood to fail decreases to a "normal" level given the current warning stage. Once the warning takes place, the likelihood to fail jumps down and stays at this lower level given the new higher warning stage. The graphs look similar in the case of shorter delays (of 4, 3 or 2 months), that is the likelihood to fail jumps down once the warning takes place and stays at the new lower level. If the warning is received with a delay of more than 5 months on the other hand, the likelihood to fail does not change upon warning receipt. We will discuss this finding in more detail below (when analyzing it in a regression framework).

In the following we analyze how the likelihood of "failure to comply" changes over time in a regression analysis. In particular, we regress "failure" on dummies for each period before and after the receipt of a warning while controlling for individual and time fixed effects and time-varying characteristics. In Table 2 we present estimates by warning stage and pooling the different warning stages, while pooling different times of delay (in contrast to the graphical illustration). The main finding is that the coefficients on the dummies up to the period of the warning receipt (for $t = -5, ..., 0$) are significantly positive, while the coefficients for the periods after the warning (for $t = 1, ..., 5$) are significantly negative. Thus the likelihood of failure to comply jumps down upon receipt of a warning, i.e. when a new warning stage is reached, and stays at this new lower level (given the higher costs of failure associated to a higher warning stage), as was already illustrated in the graphs (see previous paragraph).

As discussed in the previous section, estimating the effects of warnings on attendance is complicated by the fact that some families have a much higher propensity to fail and they are the ones who reach high warning stages, which leads to a downward bias of a negative coefficient on warnings received towards zero or even a positive correlation between "failure" to comply and warnings received when using cross-sectional data (see Specification (1) in Table 4).

[Insert Table 4]

We address this problem using family fixed effects in the rest of the analysis to control for differences in the propensity to fail and thus we make use of variation in terms of warnings (or justifications etc) within family over time to identify their effect on attendance behavior each month. Table 4 shows that using family fixed effects, the likelihood of non-attendance (that is failure to comply in a given month) decreases when the family receives a warning. This negative effect is of increasing magnitude when the warning stage increases, but as discussed in Section 4.1, one should not compare the coefficients on the different warning stages when using the full sample for estimation, since the coefficients are estimated over different subsets of families that reach the relevant warning stage of interest (for further results see companion paper).

To analyze if the negative correlation we see in Table 4 between warnings and failure is causal, we make use of the fact that the exact day of the receipt of warning is exogenous in that it depends on the last digit of the social security number (see detailed discussion in Section 4.1). We then test the following prediction: If families' responses are really due to the warnings they have received and not due to something else happening in that month (such as shocks), then those families who receive warnings earlier in the month should show a larger behavioral response than families who receive warnings later. In practice, we interact warning stage with dummies for "early warning" and "late warning" and use two different definitions: Definition 1 splits the period in which benefits are transferred (and thus warnings received) into halves, while definition 2 is constructed using a split of 70:30 (to allow for an even lower possibility to adjust for "late" recipients).

Table 6 is based on a fixed-effect regression equation that includes an interaction term between warning stage and "early" versus "late" warning. In particular, we include an interaction term between warning stage and a dummy for "early" and "late" and another interaction term between a lagged variable of warning stage and the "early/late" dummy. The lagged variable becomes one in the month following the warning, which implies that both "early" and "late" recipients of warnings should have the same chance of adjusting their behavior. The main variable "warning stage" on the other hand becomes one in the month, in which the warning is received. This is to test if families who receive an early warning show a larger response to their warning in that same month, than those with late warnings (who have less of an opportunity to adjust).

Columns (1) and (2) show results from estimating an equation that uses definition (1) of early and late (50:50 split of the transfer period within the month). Coefficients on interaction with the dummy for "early" are displayed in Column (1) and those with "late" in (2), while they are both estimated in the same equation. Columns (4) and (5) instead are coefficient estimates from estimating the same equation but based on definition (2) of early/late (70:30 split). Columns (3) and (6) show p-values for whether there are significant differences between coefficients on the effects of early versus late warnings.

Table 6 shows that families who receive their warning early in the month respond more strongly to warnings in that same month than families who receive the warning late. Using definition (1), coefficients are larger for those families who receive the warning earlier, though differences are only statistically significant for warning stages 4 and 5. Using definition (2), which leaves even less room for recipients of "late" warnings to respond, differences are statistically significant for all five warning stages. The effect of the lagged increase in warning stage is basically the same for families warned early or late, with exception of the effect of the first warning which is stronger for families who were warned late in the previous month.

To summarize, the evidence of Table 6 is consistent with families responding to the actual receipt of the warning (which was "assigned" randomly in terms of timing) and its' informational content (and not to some shock that occurs in that month when warnings happen).

Having shown a causal effect of warnings on attendance, we now move to the analysis of whether this effect can be interpreted as families learning about the strictness of enforcement. As discussed above, if families had perfect knowledge about the nonattendance of their children and knew with certainty that noncompliance is punished with warnings and when these warnings take place, then families can anticipate warnings as soon as they fail. Since in that case warnings would not contain any new information, families should not show any behavioral response to warnings.

Since we have shown that families do respond to warnings, families either learn about strictness of enforcement or parents learn about nonattendance of their children. To disentangle the two possible stories of learning, we compare schools that inform parents directly about children's nonattendance with those that do not. Under the hypothesis that warnings have an effect because of parents learning about nonattendance of their children, then the "warning effect" should be smaller in schools where parents are informed about nonattendance directly (i.e. independently of BFP). If on the other hand, the effect of warnings is due to families learning about enforcement, there should not be a significant difference in the effect of warnings between the two types of schools. For the purpose of this test, we merge our data with another data set, PROVA, which contains –among other modules– a module with detailed information on schools, such as, or example, information on school practices like informing parents about nonattendance of children. Since PROVA was conducted only in schools grades 5 to 9, we merge less than half of our sample.[7]

Table 7 presents results from regressing failure on warning stages and an interaction term between warning stage and a dummy for whether parents receive information from schools about nonattendance of their children (in addition to the standard set of controls such as, for example, household fixed effects). Columns (1) to (3) use different dummies depending on the source of information from schools. In particular, Column (1) shows results based on a dummy for whether the child is in a school which informs parents in writing about nonattendance of their children (according to director 81% of schools conduct this practice). Column (2) uses a dummy for whether the school informs parents in individual meetings (94% of schools), and in Column (3) the dummy is based on whether parents are informed by the school via home visits (68% of schools).

Table 7 shows that there is basically no difference in the effect of warnings between schools which inform parents about nonattendance of their children versus those that do not. There is only one instance in which "informed" parents respond less strongly, that is the response to warning stage 1 in the case of individual meetings. Nevertheless, also in this instance the warning still has an effect on informed parents, so it contains additional information. In the case of "information

[7]Note that the schools we can merge are not very different from the schools in our BFP data that we cannot merge. For example, overall in our BFP data the fraction of Bolsa Familia children in schools is 47%, while the fraction of BFP children in the schools we can merge is 43%.

via writing", informed parents appear to respond even more strongly to warning stage 2.

These results suggest that warnings contain information on and above informing parents about children's nonattendance. In fact, the "enforcement" aspect of the information appears to be the more relevant one, since the effect of warnings is not smaller for "informed" parents.

To summarize, the receipt of warnings reduces families' likelihood to fail. Our results show that this effect is causal, and thus families perceive uncertainty with respect to the quality of enforcement.

We now move to an analysis of how households response depends on the delays with which a warning arrives (i.e. how many months after the month of non-compliance). The median delay time is 5 months, but can vary between 2 and 10 months. Table 8 shows that the effect of a warning is significantly smaller, if the warning was received with a longer delay. This is true when we interact warnings with delay in months rescaled by the median (Column 1) and also when we interact warnings with a dummy for whether the length of delay was above the 80th percentile (Column 2).

[Insert Table 8]

This effect is consistent with households updating their beliefs about the strictness of enforcement (while the realization effect would lead to a larger response in case of a late warning since in that case the warning comes as a surprise at a moment when households had perceived a higher likelihood of not being warned at all and had therefore become less careful in terms of compliance).

We also want to investigate a second measure of the quality of enforcement that we use is whether a child's non-attendance in a given month gets justified ex-post by the school (see background section for details), in which case the non-attendance is not counted as non-compliance. Specification (3) in Table 8 shows that the likelihood of failure is higher, if a child's non-attendance was justified ex-post in the previous month. Of course, this could be due to serially correlated shocks that affect the attendance behavior in two consecutive months. For this reason, we investigate further whether there is a signal of quality of enforcement in these "justifications" in the next section, when analyzing if families learn from justifications of the children's peer and respond in their attendance behavior.[8]

5.2 Response to Peers' Experiences of Enforcement

In this Section we examine how families react to the signals of enforcement received not directly by them but by the "peers"of their children. We use the following peer group definitions: school, grade and class of each child and we analyze two types of signals: Peers receiving warnings and peers getting justified ex post for their failure. The former is a signal of (strict) enforcement, while the latter is a signal of non-enforcement (or weak enforcement).

We construct two types of peer group variables: one is the fraction of peers in a class who were warned; the other is a dummy equal to one if at least one peer was warned. We also compare the effect of own peers being warned versus siblings' peers being warned. Furthermore, we compare the effects of peers receiving different levels of warnings and how these effects depend on a family's own current warning stage.

[8] In the context of justifications of peers affecting a child's attendance behavior, we do not face the same identification challenge of serially correlated shocks as in the case of "own" justifications. We discuss identification challenges and strategies in the context of peer learning in Sections 4.2 and in the following section.

First we discuss the results related to peers receiving warnings (Tables 9 and 10). Second, we discuss the results related to peers who receive justifications (Tables 14 and 13).

5.2.1 Peers Receiving Warnings

In Column (1) of Table 9 we include the "fraction of own peers who receive any warning" as a main effect and a lagged effect in addition to own warnings. We include a lagged variable of peer warnings, because it is ex ante difficult to establish whether children learn about their peers' warnings in time to adjust their attendance in the current month or in the following month. The coefficient on the fraction of peers warned in column (1) is -0.014, significant at the 5 percent level, indicating that there is a significant negative correlation between own peers receiving a warning in a month and own attendance in the same month. The coefficient on the lag is instead zero.

To address the concern that this result might be due to direct peer effects (i.e. peers get warned, fail less and the individual responds to her peers' change in attendance behavior), we control for the fraction of peers who fail to attend in that month (see Column (2), Table 9). We find that the fraction of peers who fail is significantly positively correlated with own failure (which could be due to peer effects or common shocks). At the same time we still find a highly significant and even larger negative effect of peers' warnings on own failure after controlling for peers' failure (coefficient -0.026), which is consistent with the interpretation that people learn about stricter enforcement (and reduce their failure in response) when observing peers being warned.

In Column (3) of Table 9, we include the fraction of peers warned as lead variable to rule out the possibility that our result is driven by changes in school policies, e.g. the teacher or headmaster becoming stricter in registering students' non-attendance (and also to address the concern of mean reverting economic shocks). As explained in Section 4.2, if children decrease their failure because of greater strictness, one would expect them to start increasing their attending soon after they experience this greater strictness (and not on average five months later, which is the median lag with which warnings occur). Certainly there is no reason to believe that people change their behavior due to greater strictness always exactly in the month in which the warning happens (which can be between two and ten months after the initial month of failure), but not one month before the warning occurs.

This means that if the reason for people's change in behavior is greater strictness of teacher/school, then we should also find a negative and significant coefficient on the lead variable. If instead people respond to peers' receiving warnings and learning about enforcement, then the change in behavior should only happen in the month of the warnings and not one month earlier. In Column (3), we find that the coefficient on the lead variable is negative but smaller than the main effect and not significant, which is consistent with our learning interpretation.

In the last specification of Table 9, we use an alternative variable to capture the effect of peer warnings. In principle it is not clear that there should be a linear relationship between the fraction of peers warned and own attendance. It may be the case that what matters is whether at least one peer receives a warning, which is what we test in Specification (4). We again find a negative coefficient on this variable, but the significance level is reduced to 10 percent. So the number or fraction of peers who are warned affects the quality of the signal, which was to be expected.

[Insert Table 10]

In Table 10 we provide further evidence for our learning hypothesis and analyze additional specifications to rule out further alternative stories, in particular related to classroom and sibling peer effects. If results were driven by classroom conventional peer effects, that is children warned in individual i's class decrease their own failure (i.e. increase their attendance) and therefore i decreases her failure, then the warnings received by the peers of i's siblings (who are in a different class) should not have an effect on i's behavior (or at least a substantially smaller effect than the one of i's own peers). For this reason we include two "peer warning" variables: one is the fraction of own peers warned and the other is the "maximum of the fractions of peers warned among the siblings of individual i".

As explained in Section 4.2, we use the maximum since it is likely to be closer to the fraction of warnings that people actually learn about (closer to the "true" signal), since families are more likely to mention having been warned, when a critical mass of other families have received warnings as well, so that BFP warnings and enforcement become a topic of wider debate. In this case the families are likely to get a signal that is close to the fraction of actual warnings. If on the other hand, few families receive warnings, the topic is less likely to come up at all. Therefore the maximum of the fraction among siblings is likely to be the most reliable signal.

Table 10 shows that not only an individual's own peers matter for her attendance decision, but also her siblings' peers matter. This finding lends further support to the learning hypothesis, since in the case of conventional peer effects one would expect the effect of peers in one's own class to always be stronger than peers of siblings in different classes, grades or even schools. On the other hand, warnings of people in other grades or schools contain the same amount of information about enforcement at the federal level and thus should be equally important.

While this helps to address the concern that the results could be driven by classroom peer effects (or by grade/school-specific shocks), we might still be concerned that our result are driven by peer effects between siblings, as discussed in the list of identification challenges. In particular, if a large fraction of the peers of i's siblings receive warnings, then i's siblings might attend school more and then i might decide to attend more because her siblings attend more. We hypothesize that this type of peer effect between siblings should be stronger, the closer the siblings are in age and for siblings of the same gender.

To analyze the importance of this alternative explanation, we repeat the last test using the variable "maximum fraction of siblings' peers warned" but firstly focusing solely on siblings at different schools and secondly on siblings of the opposite sex. Siblings usually go to different schools only if they are different in age and thus have to go to different types of schools (such as elementary, junior high school etc). If the learning story is true, then the "maximum fraction of siblings' peers warned who are in schools different to i's school" (or using only siblings of opposite gender) should have an effect that is as strong as the maximum fraction of all siblings. If on the other hand, peer effects between siblings are driving the results, then we would expect the former to have a smaller effect, because in that case the siblings are further apart in age (or of opposite sex) and thus less likely to be playmates.

Specification (4) in Table 10 shows that the coefficient on "maximum fraction of siblings' peers warned in different schools" is significantly negative and of similar magnitude as the coefficient on the "maximum fraction among all siblings". The same is true for Specification (6), which shows that focusing on the peers of siblings of opposite gender has basically the same effect as regarding the maximum of peers warned among all siblings.

This provides evidence against a story of peer effects between siblings (or peer effects between children in different classes and grades) and in favor of a learning story, where families learn about the quality of enforcement from their children's peers. We provide further supportive evidence of learning in Section 5.2.2, when analyzing the effect of peer justifications. But first we want to compare effects of peers receiving different levels of warnings.

Interpretation and Informational Content

We will now analyze how the effect of peer warnings of different severity vary depending on a family's own current warning stage. Our hypothesis is that, for example, for families who are currently in warning stage one, receiving information about peers who also received warning one contains less information than peers who receive higher warnings, since the family already received a direct signal that the government enforces in terms of sending first warnings.

For that reason we will test if the effects of peers receiving higher warnings than oneself are stronger than of peers receiving the same or a lower-level warning. More specifically, we will estimate four different regression equations, conditional on whether a family is currently in warning stage 1, 2, 3 or 4. In the equation we include two different peer variables, firstly the "fraction of peers with a warning stage *lower or equal* to family i's" and secondly the "fraction of peers with warning stage *higher* than family i's".

In terms of learning we expect that the second variable should have a larger coefficient than the first, since the second contains new information, that is that the government also implements warning stages higher than the one the family is currently in. On the other hand, observing peers receiving the same warning that they already received in the past contains less information, since the family already knows that the government implements up to that warning.

Table 12 suggests that the effect of having peers with higher warnings than oneself has a stronger effect on attendance than having peers with warnings of the same or lower level than oneself. For example, for families in warning stages one or two, the coefficient on "peers with higher warnings" is about twice to four times as large as the coefficient on "peers with same or lower warnings".

5.2.2 Peers Receiving Justification of Absence

In this Section, we analyze the effect of having peers who receive an ex-post justification for their failure. It is interesting to see if people learn not only from positive signals of the quality of enforcement (i.e. seeing people receiving warnings), but also from negative signals, or signals of weak enforcement. Furthermore, we use this section to lend further credibility to the hypothesis that people learn from their peers.

In particular, we perform the following test. Warnings and ex-post justifications of non-compliance depend on different institutions: The implementation of warnings is in the hands of the Ministry of Social Development, i.e. done at the federal level. The ex-post justification of a child that failed to comply in a given month is in the hands of the school. If the effects that we find are learning effects, then the attendance behavior of an individual should be affected by warnings of her siblings' peers who are in different schools (signal of enforcement quality at the federal level), but should not be affected by justifications of her siblings' peers who are in different schools (or –if justifications are correlated across schools– they should at least be affected to a lesser extent). Instead both warnings and justifications should affect an individual's behavior, if they happen to peers who are in the same school (class) as the individual. If on the other hand, the effects we find are peer effects among siblings, then in both cases (with warnings and justifications) the peers of the siblings would have an effect on the individual via her sibling's attendance.

[Insert Table 14]

Table 14 shows that in the case of justifications, only justifications of the individual's own peers significantly increase own failure (independently of whether the maximum of peers justified happened in the "own" class or in the siblings' class). If instead a sibling –who attends a different school– has peers who get justified, this even decreases the individual's failure (coefficient on second lag is significant), while according to a sibling peer effect there should be positive effect (individual A's sibling has peers who get justified, so the sibling fails more and -if both siblings play with each other- also individual i should fail more).

While the negative coefficient on the maximum fraction of siblings' peers warned is prima facie difficult to rationalize, one may think that there could be a strategic game between siblings. Imagine that the parents punish those children who fail and if a single child is responsible for the family getting warned (and possibly losing money) the punishment is particularly harsh. If the individual learns that her siblings' peers' failure gets justified, this is an indication that also her sibling might get justified ex-post in case of non-attendance. In that case, her own failure would be noticed more (as discussed in the background section, parents receive a warning message when withdrawing transfer money and the message lists which child failed to attend and when), so the individual decides to attend. For siblings who attend the same school as the individual, there are two competing effects: on the one hand, the effect discussed above of not wanting to be the one responsible for the family getting warned. On the other hand, having peers' who get justified in the same school (but different class) might contain some information about the likelihood to get justified oneself. This should lead to a coefficient that is less negative for the siblings' peers' justification, than for the siblings' peers' justifications for siblings at different schools. Comparing Specification (2) and (4) shows that indeed the coefficient on "maximum fraction of siblings' peers in other school warned" is larger than the coefficient on "maximum fraction of all siblings".

[Insert Table 13]

Table 13 shows an alternative way to provide evidence of the existence of peer learning by controlling for the fraction of peers who fail (to control for direct peer effects or shocks). While without controlling for the fraction of peers who fail, the fraction of peers who are justified is significantly positively correlated with own failure, the coefficients lose significance when adding this control variable. If we use the alternative measure "any of own peers justified" (instead of the exact fraction), then the coefficient remains positive and there is a significant positive effect of having any peer who was justified on own attendance, which is consistent with the results in Table 14.

6 Robustness: Attendance Response or Bribing of Teachers

In this Section we aim to address the concern that our results are not actual responses in terms of children's attendance, but that they can be explained by parents bribing/convincing teachers to be less strict in registering children's nonattendance or by teachers wanting to safe the families' transfers by being more lenient after families received warnings.

Before addressing this concern, it is important to stress that our conclusion of "families learning about the strictness of enforcement" is not affected for the following reason: For example, we find that individual i's failure goes down when *peers* of i's siblings get warned. Since the teacher of child i would not know that (in particular if the siblings –and their peers– attend different schools), this result has to be explained by parents learning about the quality of enforcement and adjusting their behavior, either in terms of child's attendance or in terms of an increased effort in bribing/convincing teachers. For that reason, teachers wanting to be lenient to families cannot explain our results without families learning from their children's peers.

Furthermore, our evidence suggests that there is a clear response in actual attendance with respect to observing "justifications". In particular, when an individual has peers (or has siblings who have peers) who get justified, then she fails more in response (i.e. attends less). It is difficult to conceive that this result can be driven by teachers responding by becoming stricter in terms of registration of non-attendance, for the following reasons:

In this context parents obviously do not have an incentive to inform the teacher, while at the same time a teacher would (a) not necessarily know which children received justifications (not even in his own school, since the director would decide) and (b) it is entirely unlikely that the teacher knows whether siblings of a child in his class have peers who are justified (in particular if those siblings attend different schools). Even in cases where a teacher learns about justifications from the school director, it is unlikely that the teacher would then go against the directive of the school, that is to be tougher in registering non-attendance while the director justifies ex-post, and to risk unpopularity among student (in particular if children then get justified anyways).

Lastly, children have several different teachers during the day and on different days in a month and each of them has to register attendance. Thus, to have many days of unregistered nonattendance, parents need to bribe (or convince) a relatively large number of different teachers. Teachers earn a relatively high salary, so it is unclear if parents can bribe several different teachers with a sufficiently large amount to induce them to risk their job.

Based on these arguments, we conduct the following robustness checks: We analyze schools separately based on their fraction of BFP children. Schools with low BFP fraction are likely to be higher quality and also BFP children in these schools are likely to be a selected group, in that they are, for example, of higher ability, parents care more about good schools, families live in richer areas with better schools etc. The incentives to go to school should be higher for children at those schools (because of preference for school, higher returns etc). If we nevertheless find, that the level of failure is higher in schools with low BFP fractions, then this clearly suggests stricter enforcement and that teacher bribing is less of an issue. Thus we will check if the results we find still hold up for these types of schools.

Related to this point, the incentive to fail should be higher in rural areas, where children are used to work in agriculture. Nevertheless, we find that registered failure is actually lower in rural areas. This could be due to the fact that it is easier for parents to convince teachers that they need their children to help them (likely that parents and teacher know each other well) and/or to bribe them. For that reason, we will check if results hold up once we drop rural areas.

7 Conclusion

In this paper we study the implementation of a large-scale conditional cash transfer program "Bolsa Familia" in Brazil, which conditions transfers to poor families on children's school attendance. We analyze how people learn about the quality of enforcement and how this affects their behavior and –as a consequence– the effectiveness of the program.

We find that individuals respond to incentives and finetune their behavior in response to private as well as public signals about the quality of enforcement of the program. (Non-)Enforcement does not only have a direct effect on the family affected, but also important spillover effects on other families, who learn from the experiences of their children's peers.

Our finding that people adjust their behavior to the strictness of enforcement implies that not only formal rules but actual enforcement of the program conditions can be crucial for the program effectiveness. This aspect might be particularly important for developing countries, as they might lack administrative capacity (or political will) to strictly enforce the rules. Thus the design of conditional welfare program should take into account this important dimension and be such that proper enforcement can be guaranteed or at least facilitated.

References

Aizer, A. and J. Currie (2004). Networks of neighborhoods? correlations in the use of publicly-funded maternity care in california. *Journal of Public Economics*.

Angelucci, M. and G. D. Giorgi (2009). Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review 99(1)*, 486–508.

Angelucci, M., G. D. Giorgi, M. Rangel, and I. Rasul (2010). Family networks and school enrollment: Evidence from a randomized social experiment. *Journal of Public Economics 94(3-4)*, 197–221.

Attanasio, O. P., C. Meghir, and A. Santiago (2011). Education choices in mexico: Using a structural model and a randomized experiment to evaluate progresa. *Review of Economic Studies*.

Attanasio, O. P., C. Meghir, and M. Szekely (2003). Using randomized experiments and structural models for scaling-up: Evidence from the progresa evaluation. *IFS Working Paper, EWP03/05.*.

Banerjee, A., A. Chandrasekha, E. Duflo, and M. O. Jackson (2012). The diffusion of microfinance. *NBER Working Paper No. 17743*.

Banerjee, A. V., R. Glennerster, and E. Duflo (2008). Putting a band-aid on a corpse: Incentives for nurses in the indian public health care system. *Journal of the European Economic Association 6(2-3)*, 487–500.

Barrera-Osorio, F., M. Bertrand, L. L. Linden, and F. Perez-Calle (2011). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in colombia. *American Economic Journal: Applied Economics 3(2)*, 167–195.

Bastagli, F. (2008). Conditionality in public policy targeted to the poor: Promoting resilience?. *Social Policy & Society 8 (1).*

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2010). *Identification of Social Interactions.* Handbook of Social Economics.

Bobonis, G. J. and F. Finan (2009). Neighborhood peer effects in secondary school enrollment decisions. *The Review of Economics and Statistics 91(4)*, 695716.

Bourguignon, F., F. Ferreira, and P. G. Leite (2003). Conditional cash transfers, schooling and child labor: Micro-simulating brazil's bolsa escola program. *The World Bank Economic Review 17(2).*

Bursztyn, L. and L. C. Coffman (2010). The schooling decision: Family preferences, intergenerational conflict, and moral hazard in the brazilian favelas. *Working Paper*.

Cameron, L. and M. Shah (2011). Can mistargeting destroy social capital and stimulate crime? evidence from a cash transfer program in indonesia. *mimeo*.

Crepon, B., M. Ferracci, G. Jolivet, and G. J. van den Berg (2012). Testing for treatment anticipation using data on private information shocks. *Working Paper*.

De Janvry, A., F. Finan, and E. Sadoulet (2009). Local electoral incentives and decentralized program performance.

De Janvry, A. and E. Sadoulet (2006). Making conditional cash transfer programs more efficient: Designing for maximum effect of the conditionality. *The World Bank Economic Review 20(1).*

DeBrauw, A. and J. Hoddinott (2010). Must conditional cash transfer programs be conditioned to be effective? the impact of conditioning transfers on school enrollment in mexico. *Journal of Development Economics*.

Ding, W. and S. Lehrer (2006). Do peers affect student achievement in china's secondary schools? *Working Paper*.

Duflo, E., P. Dupas, and M. Kremer (2010). Peer effects and the impacts of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*.

Duflo, E. and E. Saez (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics*.

Foster, A. D. and M. R. Rosenzweig (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy 103(6)*, 1176–1209.

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003). Does peer ability affect student achievement. *Journal of Applied Econonics 18*, 527–544.

Harbison, R. W. and E. A. Hanushek (1992). *Educational Performance of the Poor - Lessons from Rural Northeast Brazil.* Oxford University Press.

Hoxby, C. M. (2000). Peer effects in the classroom: Learning from gender and race variation. *NBER Working Paper 7867*.

Hoxby, C. M. and G. Weingarth (2005). Taking race out of the equation: School reassignment and the structure of peer effects. *NBER Working Paper*.

Jackson, M. O. and L. Yariv (2010). *Handbook of Social Economics.*, Chapter Diffusion, Strategic Interaction, and Social Structure.

Lindert, K., A. Linder, J. Hobbs, and B. de la Brire (2007). The nuts and bolts of brazils bolsa famlia program: Implementing conditional cash transfers in a decentralized context. *The World Bank, Social Protection Working Paper No. 0709*.

Lochner, L. (2007). Individual perceptions of the criminal justice system. *American Economic Review 97(1)*, 444–460.

Rincke, J. and C. Traxler (2011). Enforcement spillovers. *The Review of Economics and Statistics 93(4)*, 12241234.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth college roommates. *The Quarterly Journal of Economics 116(2)*, 681–704.

Schultz, P. T. (2004). School subsidies for the poor: Evaluating the mexican progresa poverty program. *Journal of Development Economics*.

Todd, P. E. and K. I. Wolpin (2006). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review 96(5).*

Van den Berg, G. J., A. Bergemann, and M. Caliendo (2009). The effect of active labor market programs on not-yet treated unemployed individuals. *Journal of the European Economic Association*.

Van den Berg, G. J. and B. van der Klaauw (2006). Counseling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment. *International Economic Review 47*, 895–936.

Van den Berg, G. J., B. van der Klaauw, and J. C. van Ours (2004). Punitive sanctions and the transition rate from welfare to work. *Journal of Labor Economics 22*, 211–241.

Appendix

Figure 1: Timing of Failure, Warnings and Justifications

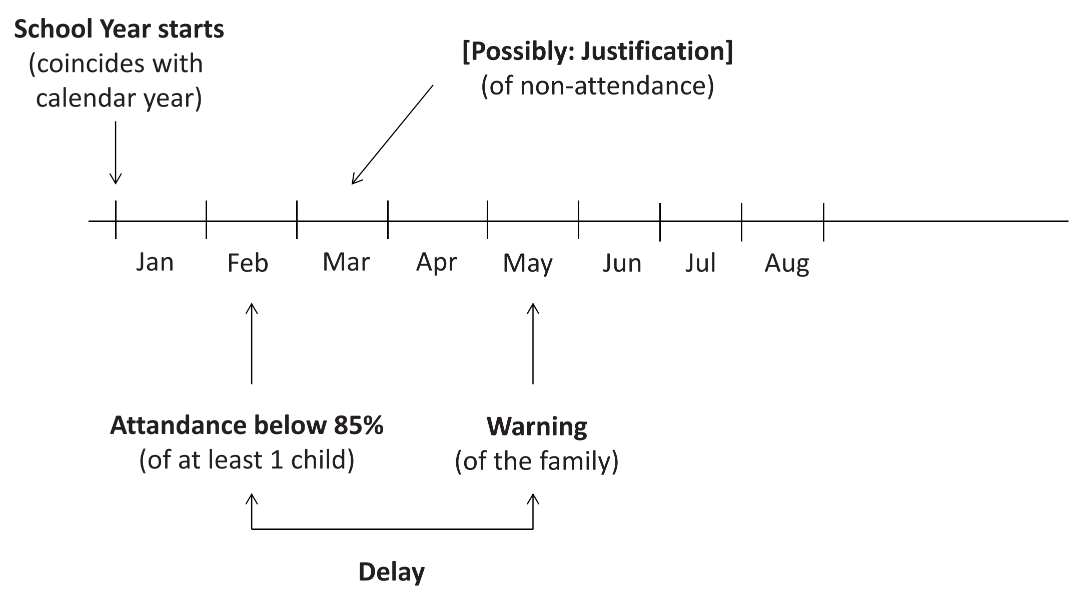

Figure 2: The Effect of Warnings: Periods Before and After Warning

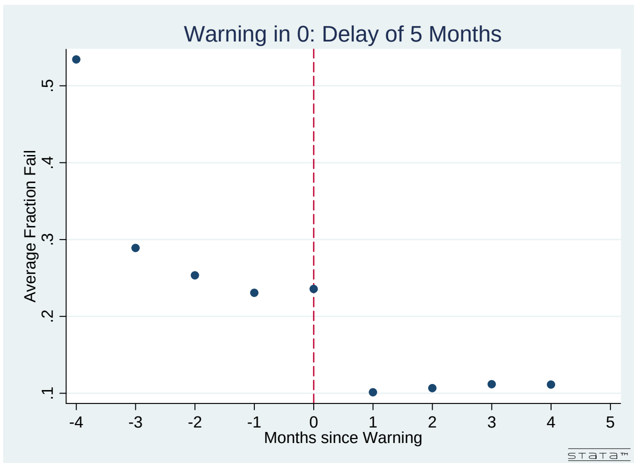

Table 1: Summary Statistics.

Variable	Obs	Mean	Std. Dev.
Male	15771234	0.527	0.499
Age	15771234	12.17	3170
Birth order	15771234	1937	1102
Nu of Sisters 6 to 10	15771234	0.275	0.521
Nu of Brothers 6 to 10	15771234	0.293	0.537
Nu of Sisters 11 to 14	15771234	0.337	0.565
Nu of Brothers 11 to 14	15771234	0.371	0.59
Nu of Sisters 15 to 17	15771234	0.193	0.43
Nu of Brothers 15 to 17	15771234	0.215	0.454
Failure to Attend	15771234	0.0399	0.196
Warning Stage 1	15771234	0.296	0.457
Warning Stage 2	15771234	0.0921	0.289
Warning Stage 3	15771234	0.0338	0.181
Warning Stage 4	15771234	0.0131	0.114
Warning Stage 5	15771234	0.00206	0.0454
Absence Justified	15771234	0.00918	0.0954
Nu of BFP Kids in Class	8515650	16.09	6836
Nu of BFP Kids in Grade	8515650	52.79	51.79
Nu of BFP Kids in School	8515650	256.3	221.5
Fraction of Peers Failed to Attend	8515650	0.0105	0.114
Fraction of Peers Warned	8515650	0.0193	0.0692
Fraction of Peers Warned WS 1	8515650	0.0114	0.0496
Fraction of Peers Warned WS 2	8515650	0.00486	0.0277
Fraction of Peers Warned WS 3	8515650	0.00203	0.0162
Fraction of Peers Warned WS 4	8515650	0.000859	0.0100
Fraction of Peers Warned WS 5	8515650	0.000241	0.00513
Max of Sibs' Peers Warned	8515650	0.0168	0.0655
Max of Sibs' Peers Warned (Other Schools)	8515650	0.0135	0.0597
Max of Sibs' Peers Warned (Oppos Sex)	8515650	0.00940	0.0498
Fraction of Peers Justified	8515650	0.00553	0.0373
Max of Sibs' Peers Justified	8515650	0.00435	0.0323
Max of Sibs' Peers Justified (Other Schools)	8515650	0.00513	0.0353
Max of Sibs' Peers Justified (Oppos Sex)	8515650	0.00283	0.0262

Table 2: Effect of Own Warnings on Children's Attendance Behavior: Months Before and After Warning.

Dependent Variable Dummy for Warning in Warning Stage	Decision of Noncompliance Pooled	 WS 1	 WS 2	 WS 3	 WS 4	 WS 5
Dummy Warn $t = -5$	0.1382***	0.1451***	0.1537***	0.1130***	0.0443**	-0.1244***
(0.004)	(0.005)	(0.010)	(0.015)	(0.022)	(0.039)	
Dummy Warn $t = -4$	0.1059***	0.1139***	0.1102***	0.0810***	0.0322	-0.0573
(0.004)	(0.005)	(0.009)	(0.014)	(0.023)	(0.043)	
Dummy Warn $t = -3$	0.2713***	0.2737***	0.2829***	0.2521***	0.2316***	0.1640***
(0.004)	(0.005)	(0.009)	(0.013)	(0.022)	(0.042)	
Dummy Warn $t = -2$	0.2026***	0.2065***	0.1971***	0.1915***	0.1956***	0.1837***
(0.003)	(0.004)	(0.007)	(0.011)	(0.019)	(0.037)	
Dummy Warn $t = -1$	0.0405***	0.0338***	0.0597***	0.0392***	0.0534***	-0.0239
(0.003)	(0.003)	(0.006)	(0.010)	(0.017)	(0.033)	
Dummy Warn $t = 0$	-0.0002	-0.0097***	0.0255***	-0.0032	0.0120	0.0152
(0.002)	(0.002)	(0.005)	(0.009)	(0.014)	(0.024)	
Dummy Warn $t = 1$	-0.0732***	-0.0751***	-0.0801***	-0.0848***	-0.0040	0.0390
(0.002)	(0.002)	(0.005)	(0.009)	(0.019)	(0.033)	
Dummy Warn $t = 2$	-0.0716***	-0.0734***	-0.0763***	-0.0701***	-0.0676***	-0.1070***
(0.002)	(0.002)	(0.005)	(0.009)	(0.020)	(0.039)	
Dummy Warn $t = 3$	-0.0722***	-0.0750***	-0.0757***	-0.0620***	-0.0654***	-0.0771*
(0.002)	(0.002)	(0.005)	(0.009)	(0.021)	(0.040)	
Dummy Warn $t = 4$	-0.0723***	-0.0714***	-0.0808***	-0.0739***	-0.0437**	-0.0437
(0.002)	(0.002)	(0.005)	(0.009)	(0.022)	(0.048)	
Dummy Warn $t = 5$	-0.0730***	-0.0749***	-0.0706***	-0.0653***	-0.1015***	-0.0826*
(0.002)	(0.002)	(0.005)	(0.010)	(0.018)	(0.050)	
Controls	Yes			Yes		
HH FE	Yes			Yes		
Time FE	Yes			Yes		
Observations	2,692,693			2,692,693		
R-squared	0.19			0.19		

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are month and year dummies and number of boys and girls in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18. All specifications include family fixed effects.

Table 3: Effect of Own Warnings on Children's Attendance Behavior: Months After Failure and Before and After Warning.

Dependent Variable Dummy for Warning in Warning Stage	Decision of Noncompliance Pooled	WS 1	WS 2	WS 3	WS 4	WS 5
Dummy Warn $t = -5$	0.1505***	0.1573***	0.1657***	0.1262***	0.0603***	-0.0994**
	(0.004)	(0.005)	(0.010)	(0.015)	(0.022)	(0.040)
Dummy Warn $t = -4$	0.1067***	0.1143***	0.1097***	0.0835***	0.0461**	-0.0395
	(0.004)	(0.005)	(0.009)	(0.014)	(0.022)	(0.042)
Dummy Warn $t = -3$	0.2677***	0.2682***	0.2806***	0.2547***	0.2388***	0.1736***
	(0.004)	(0.005)	(0.009)	(0.013)	(0.023)	(0.041)
Dummy Warn $t = -2$	0.1761***	0.1814***	0.1699***	0.1612***	0.1620***	0.1550***
	(0.003)	(0.004)	(0.007)	(0.011)	(0.018)	(0.037)
Dummy Warn $t = -1$	0.0206***	0.0122***	0.0468***	0.0202**	0.0306*	-0.0518
	(0.003)	(0.003)	(0.006)	(0.010)	(0.016)	(0.033)
Dummy Warn $t = 0$	0.0250***	0.0136***	0.0596***	0.0235***	0.0288**	0.0257
	(0.002)	(0.002)	(0.005)	(0.008)	(0.013)	(0.023)
Dummy Warn $t = 1$	-0.0296***	-0.0307***	-0.0380***	-0.0419***	0.0238	0.0540*
	(0.002)	(0.002)	(0.005)	(0.009)	(0.018)	(0.032)
Dummy Warn $t = 2$	-0.0330***	-0.0335***	-0.0445***	-0.0359***	-0.0310	-0.0686
	(0.002)	(0.002)	(0.005)	(0.009)	(0.019)	(0.044)
Dummy Warn $t = 3$	-0.0523***	-0.0562***	-0.0559***	-0.0406***	-0.0336	-0.0417
	(0.002)	(0.002)	(0.005)	(0.009)	(0.021)	(0.041)
Dummy Warn $t = 4$	-0.0686***	-0.0692***	-0.0754***	-0.0637***	-0.0311	-0.0308
	(0.002)	(0.002)	(0.005)	(0.009)	(0.022)	(0.046)
Dummy Warn $t = 5$	-0.0687***	-0.0721***	-0.0640***	-0.0567***	-0.0853***	-0.0881*
	(0.002)	(0.002)	(0.005)	(0.010)	(0.017)	(0.045)
Dummy Fail $t = 1$	0.1675***			0.1673***		
	(0.002)			(0.002)		
Dummy Fail $t = 2$	-0.0178***			-0.0181***		
	(0.002)			(0.002)		
Dummy Fail $t = 3$	-0.0487***			-0.0492***		
	(0.002)			(0.002)		
Dummy Fail $t = 4$	-0.0677***			-0.0685***		
	(0.002)			(0.002)		
Dummy Fail $t = 5$	-0.0957***			-0.0960***		
	(0.002)			(0.002)		
Controls	Yes			Yes		
HH FE	Yes			Yes		
Time FE	Yes			Yes		
Observations	2,692,693			2,692,693		
R-squared	0.20			0.20		

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are month and year dummies and number of boys and girls in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18. All specifications include family fixed effects.

Table 4: Effect of Own Warnings on Children's Attendance Behavior (Failure to Attend a Sufficient Number of Days in a Month).

Dependent Variable	**Decision of Noncompliance in a Given Month**	
	(1)	(2)
Warning Stage 1	0.0085***	0.0016***
	(0.000)	(0.000)
Warning Stage 2	0.0196***	-0.0338***
	(0.000)	(0.001)
Warning Stage 3	0.0284***	-0.0592***
	(0.001)	(0.001)
Warning Stage 4	0.0413***	-0.0821***
	(0.001)	(0.002)
Warning Stage 5	0.0647***	-0.1014***
	(0.003)	(0.005)
Dummies Month Fail	Yes	Yes
Controls	Yes	Yes
Individual FE	No	Yes
Time FE	No	Yes
Observations	6,190,311	6,190,311
R-squared	0.01	0.37

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are age dummies, birth order dummies, number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18, month and year dummies and month of failure dummies. The second specification includes individual fixed effects.

Table 5: Balance Test for Being Warned Early or Late

Dependent Variable	Balance Test					
Timing	Definition 1			Definition 2		
	Early Mean (Std Dev)	Late Mean (Std Dev)	Diff (P-Val)	Early Mean (Std Dev)	Late Mean (Std Dev)	Diff (P-Val)
Male	0.527 (0.419)	0.528 (0.419)	(0.6812)	0.528 (0.419)	0.528 (0.419)	(0.5705)
Age	11.94 (2.772)	11.94 (2.777)	(0.9721)	11.94 (2.774)	11.94 (2.777)	(0.7336)
Birthorder	1.702 (0.748)	1.703 (0.751)	(0.6751)	1.702 (0.752)	1.703 (0.752)	(0.8858)
Nu of Sisters 6 to 10	0.209 (0.399)	0.209 (0.399)	(0.7213)	0.209 (0.4)	0.208 (0.398)	(0.6954)
Nu of Sisters 11 to 14	0.25 (0.428)	0.251 (0.432)	(0.2920)	0.251 (0.429)	0.251 (0.433)	(0.8020)
Nu of Sisters 15 to 17	0.14 (0.31)	0.141 (0.31)	(0.3716)	0.14 (0.31)	0.141 (0.31)	(0.4712)
Nu of Brothers 6 to 10	0.224 (0.412)	0.223 (0.413)	(0.7107)	0.224 (0.412)	0.224 (0.412)	(0.9721)
Nu of Brothers 11 to 14	0.277 (0.452)	0.275 (0.45)	(0.1036)	0.277 (0.452)	0.274 (0.448)	(0.1248)
Nu of Brothers 15 to 17	0.156 (0.326)	0.157 (0.329)	(0.4638)	0.156 (0.327)	0.157 (0.329)	(0.5992)
Year 2009	0.511 (0.277)	0.51 (0.278)	(0.6313)	0.511 (0.278)	0.51 (0.278)	(0.4213)
Nu of BFP Kids in School	258 (201.9)	258.6 (203.7)	(0.4986)	258.7 (202.8)	257.7 (203.1)	(0.2203)
Nu of BFP Kids in Grade	51.74 (44.92)	52.08 (45.73)	(0.0737)	51.94 (45.22)	51.96 (45.67)	(0.8895)
Nu of BFP Kids in Class	15.81 (5.989)	15.81 (5.983)	(0.8535)	15.81 (5.994)	15.81 (5.972)	(0.7327)
Observations	119,289	140,539		178,290	118,677	

Table 6: Effect of Own Warnings on Children's Attendance Behavior: Causality.

Dependent Variable	Decision of Noncompliance in Given Month					
Timing	Definition 1			Definition 2		
	Early	Late	Diff	Early	Late	Diff
	Coeff/(SE)	Coeff/(SE)	(P-Val)	Coeff/(SE)	Coeff/(SE)	P-Val
Warning Stage 1 * Timing	0.0057***	0.0059***		0.0057***	0.0060***	
	(0.001)	(0.000)		(0.000)	(0.000)	
Warning Stage 2 * Timing	-0.0142***	-0.0093***		-0.0139***	-0.0075***	
	(0.001)	(0.001)		(0.001)	(0.001)	
Warning Stage 3 * Timing	-0.0247***	-0.0152***		-0.0231***	-0.0139***	
	(0.002)	(0.002)		(0.002)	(0.002)	
Warning Stage 4 * Timing	-0.0255***	-0.0189***		-0.0248***	-0.0165***	
	(0.004)	(0.003)		(0.003)	(0.003)	
Warning Stage 5 * Timing	-0.0148*	-0.0180***		-0.0221***	-0.0134*	
	(0.009)	(0.007)		(0.007)	(0.007)	
Lag Warning Stage 1 * Timing	-0.0006	-0.0007		-0.0008	-0.0006	
	(0.001)	(0.000)		(0.000)	(0.001)	
Lag Warning Stage 2 * Timing	-0.0248***	-0.0248***		-0.0251***	-0.0247***	
	(0.001)	(0.001)		(0.001)	(0.001)	
Lag Warning Stage 3 * Timing	-0.0438***	-0.0460***		-0.0431***	-0.0475***	
	(0.002)	(0.002)		(0.002)	(0.002)	
Lag Warning Stage 4 * Timing	-0.0683***	-0.0702***		-0.0694***	-0.0698***	
	(0.004)	(0.003)		(0.003)	(0.004)	
Lag Warning Stage 5 * Timing	-0.1022***	-0.0998***		-0.1011***	-0.0974***	
	(0.010)	(0.009)		(0.009)	(0.010)	
Dummies Month Fail	Yes			Yes		
Controls	Yes			Yes		
Individual FE	Yes			Yes		
Time FE	Yes			Yes		
Observations	4,430,501			4,430,501		
R-squared	0.39			0.39		

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are age dummies, birth order dummies, number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18, month and year dummies and month of failure dummies. All specifications include individual fixed effects.

Table 7: Effect of Own Warnings on Children's Attendance Behavior: Interpretation.

Dependent Variable	**Decision of Noncompliance**		
Parents informed in	Writing	Indiv Meeting	Home Visit
Warning Stage 1		0.0024	
		(0.002)	
Warning Stage 2		-0.0299***	
		(0.004)	
Warning Stage 3		-0.0564***	
		(0.008)	
Warning Stage 4		-0.0946***	
		(0.013)	
Warning Stage 5		-0.1339***	
		(0.034)	
Warning Stage 1 * Parents Informed	0.0019**	-0.0025	0.0012
	(0.001)	(0.002)	(0.001)
Warning Stage 2 * Parents Informed	0.0056***	-0.0118***	0.0039**
	(0.002)	(0.004)	(0.002)
Warning Stage 3 * Parents Informed	0.0026	-0.0050	-0.0023
	(0.003)	(0.008)	(0.003)
Warning Stage 4 * Parents Informed	-0.0025	0.0138	0.0023
	(0.006)	(0.013)	(0.006)
Warning Stage 5 * Parents Informed	-0.0333**	0.0527	0.0072
	(0.014)	(0.035)	(0.013)
Parents Informed	-0.0023	0.0005	-0.0040***
	(0.002)	(0.004)	(0.002)
Dummies Month Fail		Yes	
Controls		Yes	
Individual FE		Yes	
Time FE		Yes	
Observations		3,258,924	
R-squared		0.37	

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are age dummies, birth order dummies, number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18, month and year dummies and month of failure dummies. All specifications include individual fixed effects.

Table 8: Effect of the Quality of Enforcement on Children's Attendance Behavior: Delay in Receipt of Warning.

Dependent Variable	Decision of Noncompliance		
	(1)	(2)	(3)
Warning Stage 1	-0.0003	0.0011***	-0.0001
	(0.000)	(0.000)	(0.001)
Warning Stage 2	-0.0356***	-0.0427***	-0.0379***
	(0.001)	(0.001)	(0.001)
Warning Stage 3	-0.0601***	-0.0710***	-0.0639***
	(0.001)	(0.001)	(0.002)
Warning Stage 4	-0.0805***	-0.0930***	-0.0932***
	(0.002)	(0.002)	(0.003)
Warning Stage 5	-0.1031***	-0.1107***	-0.1108***
	(0.005)	(0.005)	(0.011)
Warning St 1 * Delay (in months)	0.0023***		
	(0.000)		
Warning St 2 * Delay (in months)	0.0154***		
	(0.000)		
Warning St 3 * Delay (in months)	0.0163***		
	(0.001)		
Warning St 4 * Delay (in months)	0.0128***		
	(0.001)		
Warning St 5 * Delay (in months)	0.0017		
	(0.003)		
Warning St 1 * Long Delay (top 20%)		0.0009*	
		(0.001)	
Warning St 2 * Long Delay (top 20%)		0.0380***	
		(0.001)	
Warning St 3 * Long Delay (top 20%)		0.0591***	
		(0.002)	
Warning St 4 * Long Delay (top 20%)		0.0722***	
		(0.006)	
Warning St 5 * Long Delay (top 20%)		0.0138	
		(0.027)	
Justification (lag 1)			-0.0020
			(0.002)
Justification (lag 2)			0.0157***
			(0.002)
Dummies Month Fail	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	6,190,311	6,190,311	2,697,988
R-squared	0.37	0.37	0.43

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are age dummies, birth order dummies, month and year dummies and number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18. All specifications include family fixed effects.

Table 9: Effect of Own Peers Getting Warned on a Child's Attendance Behavior (Failure to Attend a Sufficient Number of Days in a Month).

Dependent Variable	**Decision of Noncompliance in Given Month**			
	(1)	(2)	(3)	(4)
Frac of Own Peers Warned	-0.0388***	-0.0446***	-0.0266***	
	(0.002)	(0.002)	(0.003)	
Frac of Own Peers Warned (lag)	-0.0369***	-0.0382***	-0.0295***	
	(0.002)	(0.002)	(0.003)	
Frac of Own Peers Warned (lead)		-0.0044		
		(0.003)		
Any of Own Peers Warned				-0.0023***
				(0.000)
Any of Own Peers Warned (lag)				-0.0037***
				(0.000)
Frac of Own Peers Fail			0.5728***	0.5653***
			(0.026)	(0.025)
Frac of Own Peers Fail (lag)			0.1219***	0.1231***
			(0.016)	(0.016)
Warning Stage 1	0.0018***	0.0015***	0.0018***	0.0017***
	(0.000)	(0.001)	(0.001)	(0.001)
Warning Stage 2	-0.0338***	-0.0339***	-0.0325***	-0.0327***
	(0.001)	(0.001)	(0.001)	(0.001)
Warning Stage 3	-0.0563***	-0.0563***	-0.0537***	-0.0536***
	(0.002)	(0.002)	(0.002)	(0.002)
Warning Stage 4	-0.0751***	-0.0749***	-0.0746***	-0.0736***
	(0.003)	(0.003)	(0.003)	(0.003)
Warning Stage 5	-0.0922***	-0.0998***	-0.0878***	-0.0868***
	(0.008)	(0.010)	(0.009)	(0.009)
Dummies Month Fail	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Indiv FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	2,898,078	2,365,729	2,244,454	2,334,925
R-squared	0.40	0.38	0.41	0.42

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are age dummies, birth order dummies, number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18, month and year dummies and month of failure dummies. All specifications include individual fixed effects. The number of observations is reduced compared to the analysis on the effect of own warnings, since we can only define the peer group for a subset of observations.

Table 10: Effect of Siblings' Peers Getting Warned: All Siblings, Siblings at Other Schools and Siblings of Opposite Gender.

Dependent Variable	**Decision of Noncompliance in Given Month**		
	Siblings	Sibs in Other Schools	Sibs of Opposite Sex
	(1)	(2)	(3)
Frac of Own Peers Warned	-0.0342***	-0.0344***	-0.0355***
	(0.003)	(0.003)	(0.003)
Frac of Own Peers Warned (lag)	-0.0338***	-0.0358***	-0.0348***
	(0.003)	(0.003)	(0.003)
Max of Sibs Peers Warned	-0.0079***		
	(0.003)		
Max of Sibs Peers Warned (lag)	-0.0061**		
	(0.003)		
Max of Sibs Peers Warned		-0.0065**	
(Sib in Other Schools)		(0.003)	
Max of Sibs Peers Warned (lag)		-0.0020	
(Sib in Other Schools)		(0.003)	
Max of Sibs Peers Warned			-0.0073**
(Sib of Opposite Sex)			(0.003)
Max of Sibs Peers Warned (lag)			-0.0055*
(Sib of Opposite Sex)			(0.003)
Warning Stage 1	0.0020***	0.0020***	0.0020***
	(0.001)	(0.001)	(0.001)
Warning Stage 2	-0.0292***	-0.0292***	-0.0292***
	(0.001)	(0.001)	(0.001)
Warning Stage 3	-0.0492***	-0.0492***	-0.0492***
	(0.002)	(0.002)	(0.002)
Warning Stage 4	-0.0648***	-0.0648***	-0.0648***
	(0.003)	(0.003)	(0.003)
Warning Stage 5	-0.0822***	-0.0822***	-0.0822***
	(0.009)	(0.009)	(0.009)
Dummies Month Fail	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Indiv FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	2,185,469	2,185,469	2,185,469
R-squared	0.39	0.39	0.39

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are age dummies, birth order dummies, number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18, month and year dummies and month of failure dummies. All specifications include individual fixed effects. The number of observations is reduced as we only include families with at least two children.

Table 11: Effect of Peers Receiving Different Levels of Warnings on a Child's Attendance Behavior.

Dependent Variable	Decision of Noncompliance in Given Month				
Own Warning Stage	1	2	3	4	5
Frac Peers Warned (WS 1)	-0.0104***	0.0068	0.0152	-0.0069	0.0148
	(0.003)	(0.008)	(0.016)	(0.027)	(0.072)
Frac Peers Warned (WS 2)	-0.0013	-0.0206***	-0.0229	-0.0171	0.2040
	(0.007)	(0.007)	(0.021)	(0.042)	(0.148)
Frac Peers Warned (WS 3)	-0.0113	-0.0391**	-0.0418**	-0.0727	-0.0902
	(0.011)	(0.018)	(0.017)	(0.054)	(0.176)
Frac Peers Warned (WS 4)	0.0132	-0.0350	-0.0706	-0.0247	0.2688
	(0.019)	(0.032)	(0.043)	(0.029)	(0.250)
Frac Peers Warned (WS 5)	-0.0433	-0.0429	-0.0304	0.2761*	-0.0143
	(0.045)	(0.070)	(0.093)	(0.157)	(0.115)
Frac Peers Fail	0.3880***	0.5385***	0.6311***	0.7584***	0.3784*
	(0.028)	(0.064)	(0.047)	(0.069)	(0.207)
Dummies Month Fail	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Indiv FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	1,768,774	553,053	200,437	77,543	12,962
R-squared	0.49	0.51	0.51	0.54	0.62

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are age dummies, birth order dummies, number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18, month and year dummies and month of failure dummies. All specifications include individual fixed effects.

Table 12: Effect of Peers Receiving Either a Lower or Same Level of Warning Versus a Higher Level Warning on a Child's Attendance Behavior.

Dependent Variable	**Decision of Noncompliance in Given Month**			
Own Warning Stage	1	2	3	4
Frac Peers Warned (WS 1)	-0.0103***			
	(0.003)			
Frac Peers Warned (WS 2 to 5)	-0.0031			
	(0.005)			
Frac Peers Warned (WS 1 to 2)		-0.0084*		
		(0.004)		
Frac Peers Warned (WS 3 to 5)		-0.0358**		
		(0.015)		
Frac Peers Warned (WS 1 to 3)			-0.0143*	
			(0.008)	
Frac Peers Warned (WS 4 to 5)			-0.0651*	
			(0.037)	
Frac Peers Warned (WS 1 to 4)				-0.0242*
				(0.014)
Frac Peers Warned (WS 5)				0.2737*
				(0.158)
Frac Peers Fail	0.3880***	0.4970***	0.5998***	0.7585***
	(0.028)	(0.055)	(0.047)	(0.069)
Dummies Month Fail	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Indiv FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	1,768,774	602,21	219,035	77,543
R-squared	0.49	0.52	0.52	0.54

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are age dummies, birth order dummies, number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18, month and year dummies and month of failure dummies. All specifications include individual fixed effects.

Table 13: Effect of Own Peers Receiving a Justification for their Failure on a Child's Attendance Behavior (Failure to Attend a Sufficient Number of Days in a Month).

Dependent Variable	**Decision of Noncompliance in Given Month**			
	(1)	(2)	(3)	(4)
Frac of Own Peers Justified	0.1228***	0.1160***	0.1097***	
	(0.006)	(0.006)	(0.007)	
Frac of Own Peers Justified (lag)	0.0251***	0.0208***	0.0145**	
	(0.005)	(0.005)	(0.006)	
Frac of Own Peers Justified (lead)			0.0058	
			(0.005)	
Any Own Peer Justified				0.0182***
				(0.001)
Any Own Peer Justified (lag)				0.0035***
				(0.001)
Frac of Peers Fail		0.5782***	0.5994***	0.5669***
		(0.026)	(0.024)	(0.027)
Frac of Peers Fail (lag)		0.1228***	0.1119***	0.1215***
		(0.016)	(0.013)	(0.016)
Frac of Peers Fail (lead)			0.1230***	
			(0.013)	
Justified	-0.0795***	-0.0815***	-0.0797***	-0.0764***
	(0.002)	(0.002)	(0.002)	(0.002)
Justified (lag)	-0.0039*	-0.0042*	-0.0004	-0.0038*
	(0.002)	(0.002)	(0.003)	(0.002)
Warning Stage 1	0.0010*	0.0015***	0.0020***	0.0015***
	(0.001)	(0.001)	(0.001)	(0.001)
Warning Stage 2	-0.0359***	-0.0332***	-0.0290***	-0.0330***
	(0.001)	(0.001)	(0.001)	(0.001)
Warning Stage 3	-0.0584***	-0.0539***	-0.0516***	-0.0537***
	(0.002)	(0.002)	(0.002)	(0.002)
Warning Stage 4	-0.0802***	-0.0752***	-0.0715***	-0.0751***
	(0.003)	(0.004)	(0.004)	(0.004)
Warning Stage 5	-0.0954***	-0.0878***	-0.0943***	-0.0882***
	(0.009)	(0.010)	(0.013)	(0.009)
Dummies Month Fail	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Indiv FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	2,346,128	2,140,355	1,619,429	2,205,622
R-squared	0.41	0.41	0.39	0.42

Notes: Robust standard errors in parentheses (clustering at the household level). * $p<0.1$ ** $p<0.05$ *** $p<0.01$. Included controls are age dummies, birth order dummies, number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18, month and year dummies and month of failure dummies. All specifications include individual fixed effects.

Table 14: Effect of Siblings' Peers Receiving a Justification for their Failure on a Child's Attendance Behavior (Failure to Attend a Sufficient Number of Days in a Month).

Dependent Variable	**Decision of Noncompliance in Given Month**		
	Siblings	Siblings in Other Schools	Siblings of Opposite Sex
	(1)	(2)	(3)
Frac of Own Peers Justif	0.1156***	0.1098***	0.1156***
	(0.007)	(0.007)	(0.007)
Frac of Own Peers Justif (lag)	0.0276***	0.0205***	0.0270***
	(0.006)	(0.007)	(0.006)
Max Frac of Sib Peers Justified	0.0385***		
	(0.005)		
Max Frac of Sib Peers Justified (lag)	0.0012		
	(0.005)		
Max Frac of Sib Peers Justified		0.0293***	
(Sib in Other Schools)		(0.006)	
Max Frac of Sib Peers Justified (lag)		0.0146***	
(Sib in Other Schools)		(0.005)	
Max Frac of Sib Peers Justified			0.0417***
(Sib of Opposite Sex)			(0.007)
Max Frac of Sib Peers Justified (lag)			0.0014
(Sib of Opposite Sex)			(0.006)
Justified	-0.0783***	-0.0775***	-0.0777***
	(0.002)	(0.002)	(0.002)
Justified (lag)	-0.0055**	-0.0052*	-0.0053**
	(0.003)	(0.003)	(0.003)
Warning Stage 1	0.0009	0.0009	0.0009
	(0.001)	(0.001)	(0.001)
Warning Stage 2	-0.0315***	-0.0315***	-0.0315***
	(0.001)	(0.001)	(0.001)
Warning Stage 3	-0.0508***	-0.0507***	-0.0507***
	(0.002)	(0.002)	(0.002)
Warning Stage 4	-0.0690***	-0.0687***	-0.0687***
	(0.004)	(0.004)	(0.004)
Warning Stage 5	-0.0897***	-0.0901***	-0.0900***
	(0.010)	(0.010)	(0.010)
Dummies Month Fail	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Indiv FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	1,767,481	1,789,104	1,789,104
R-squared	0.41	0.41	0.41

Notes: Robust standard errors in parentheses (clustering at the household level). * p<0.1 ** p<0.05 *** p<0.01. Included controls are age dummies, birth order dummies, number of brothers and sisters in the age categories 0 to 5, 6 to 10, 11 to 15, and 16 to 18, month and year dummies and month of failure dummies. All specifications include individual fixed effects. The number of observations is reduced as we only include families with at least two children.