Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs*

Rajashri Chakrabarti†

Federal Reserve Bank of New York

Abstract

This paper compares two alternative voucher designs implemented in the U.S. The Milwaukee program was a "voucher shock" program that made low-income students eligible for vouchers. The Florida program was an accountability-tied voucher program that faced failing schools with "threat of vouchers" and stigma. In the context of a formal theoretical model, the study argues that the threatened schools will improve under the Florida-type program and this improvement will exceed that of the corresponding treated schools under the Milwaukee-type program. Using school-level scores from Florida and Wisconsin, and a difference-in-differences estimation strategy in trends, it then finds strong support in favor of these predictions.

Keywords: Vouchers, Incentives, Public School Performance, Mean Reversion

JEL Classifications: H4, I21, I28

*I thank Robin Boadway, Steve Coate, Julie Berry Cullen, Ron Ehrenberg, Dennis Epple, David Figlio, Caroline Hoxby, Brian Jacob, Miguel Urquiola, Mike Waldman and seminar participants at Cornell University, New York University, Northwestern University, University of Georgia, American Economic Association meetings, Econometric Society Meetings, Southern Economic Association Meetings, American Education Finance Association Conference and Society of Labor Economists Conference for helpful discussions, and the Program on Education Policy and Governance at Harvard University for its postdoctoral support. The views expressed here are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. All errors are my own.

†Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045. Email: Rajashri.Chakrabarti@ny.frb.org

1 Introduction

The 1983 report "A Nation at Risk"[1] and a series of similar reports have led to continued concern that American public schools may be lagging behind their counterparts elsewhere in the developed world. This has led to a wave of demands for public school reform. School choice and accountability in general, and vouchers in particular, are among the most hotly debated instruments of public school reform. This paper is motivated by the need to understand the effect of vouchers and, in particular, the designs of different kinds of vouchers on public school performance. It argues that all voucher programs are not created equal. There are often fundamental differences in voucher designs that affect public school incentives differently and in turn bring about different responses from them.

The last two decades have seen a spurt of publicly funded voucher programs in the United States. Interestingly, while they are all "voucher" programs, there are crucial differences in their designs. These programs can be classified into two generic designs:—the "voucher shock" (VS) design and the "threat of voucher" (TOV) design (see footnote 6). In this paper, I study and contrast (both theoretically and empirically) the impacts of these two voucher designs on public school incentives and performance. For this purpose, I focus on the two oldest publicly funded voucher programs implementing these two representative designs—the Milwaukee voucher program and the Florida voucher program respectively. Given the emphasis on school choice and vouchers as school reform initiatives, and especially the recent flurry of voucher programs in the country, it behooves us, both from a scholarly and policy perspective, to understand the impacts of these two alternative designs on public school incentives and performance.

The Milwaukee program can be looked upon as a "voucher shock" (VS) program with a sudden government announcement that the low-income public school population would be eligible for vouchers. Vouchers, under both programs were funded by public school revenue, and hence associated with monetary loss for the public schools. In particular, starting in the 1990-91 school year, the Milwaukee Parental Choice Program (MPCP) made all public school students with family income at or below 175% of the poverty line eligible for vouchers to attend private schools.

In contrast, the Florida program embedded vouchers in a school accountability system. Here the failing public schools were first threatened with vouchers and vouchers were implemented *only if* they failed to meet a government designated cutoff quality level. The key here is that meeting the pre-

[1] National Commission of Excellence in Education (1983),"A Nation at Risk: The Imperative for Educational Reform," Washington, DC: U.S. Government Printing Office.

designated cutoff allowed the school to completely escape vouchers, and the corresponding revenue loss. In particular, under the Florida opportunity scholarship program, all students of a public school became eligible for vouchers or "opportunity scholarships" if the school received two "F" grades in a period of four years. Therefore, a school getting an "F" for the first time *was exposed to the threat of vouchers but did not face vouchers unless and until it got a second "F" within the next three years.* In addition, "F" being the lowest performing grade exposed these schools to stigma.[2] This paper argues that differences in voucher designs between the Florida and Milwaukee programs will affect public school incentives differently and will induce very different responses from them. In particular, it argues that the Florida type voucher system will have a larger effect on response and performance of the threatened public schools than the Milwaukee-type "voucher shock" program (on the corresponding treated schools).

Apart from the above differences, the designs of the two programs were strikingly similar. In both programs, private schools were not permitted, by law, to discriminate between students who applied with vouchers—they had to accept all students unless oversubscribed and had to pick students randomly if oversubscribed. The voucher amount had to be taken as full payment of tuition in both programs, but parents were often responsible for transportation costs, uniforms and costs of extracurricular activities. Moreover, there was likely a relocation cost to switching schools. Note that, in both programs, private schools could choose whether or not to participate in the program, and the above rules applied only if they chose to participate. The system of funding was also similar. Under each program the average voucher amount equaled the state aid per pupil, and vouchers were financed by an equivalent reduction of state aid to the district. Thus state funding was directly tied to student enrollment and enrollment losses due to vouchers were reflected in a revenue loss for the public school. The average voucher amounts under the Florida (1999-2000 through 2001-2002) and Milwaukee (1990-1991 through 1996-1997) programs were respectively \$3,330 and \$3,346.[3] During the corresponding periods, vouchers as a percentage of total revenue per pupil were 41.55% in Florida and 45.23% in Milwaukee.

The paper develops its argument in the context of a formal theoretical model that captures the basic

[2] The Florida Department of Education classified schools according to five grades: -A, B, C, D, F (A-highest, F-lowest). In 1999, 78 schools got an "F" and students in 2 of those schools became eligible for vouchers. In 2000, 4 elementary schools got an F although none became eligible for vouchers. In 2001, no school got an F. In 2002, 64 schools got an F. Students in 10 of those schools became eligible for vouchers.

[3] I will focus on the Milwaukee program up to 1996-97. This is because following a 1998 Wisconsin Supreme Court ruling, there was a major shift in the program when religious private schools were allowed to participate in the program and the program entered into its second phase. Moreover, financing of the Milwaukee program saw some crucial changes, so that voucher amounts and revenue loss per student due to vouchers were not comparable between Florida and second phase Milwaukee.

features of the Florida and Milwaukee voucher programs. The objective of this model is to understand and compare the impacts of the two designs on public school incentives and responses. The model has three agents:–the public school, the households and the private schools. It derives the demand for public school endogenously from household behavior (modeling it as a function of peer group quality and public school quality), thus giving micro-foundations to the public school payoff function. In an equilibrium model of public school and household behavior the paper endogenously determines public school quality and its ingredients—public school effort and peer group quality. Both under complete information and under moral hazard (when public school effort is not observable), the model yields two empirically testable predictions that hold at the respective program equilibria—the threatened public schools will show an improvement in quality under the Florida-type voucher system and their improvement will exceed that of the corresponding treated schools under the Milwaukee-type program.

Using school-level test score data from Florida and Wisconsin, the paper next proceeds to test the two theoretical predictions. Implementing a difference-in-differences estimation strategy in trends, it estimates the treatment effects for each of the programs by comparing the post-program improvement of the treated schools with an appropriate set of control schools. Controlling for potentially confounding pre-program time trends and post-program common shocks, the paper finds considerable evidence in favor of both the theoretical predictions. These findings are quite robust in that they continue to hold after controlling for other confounding factors such as mean reversion, sorting, and withstand several sensitivity tests. The findings have important policy implications from the point of view of public school reform.

It might be worthwhile here to consider the intuition behind the differences in public school responses between the two voucher systems. Consider a school facing the Florida "Threat of Voucher" (TOV) program, a school that just received the first "F" grade. The school realizes that if it can meet the target cutoff (that is, avoid another "F" in the next three years), then it can obviate vouchers (and the corresponding loss of revenue and stigma). Therefore, it has a strong incentive to exert effort to escape the second "F". In contrast, the same school facing a Milwaukee-type VS program realizes that an increase in effort can serve to retain some marginal students, but it can in no way completely avoid vouchers, as they have already been introduced. Therefore, under the TOV program, a threatened school has more of an incentive to respond than the same school facing a VS program. The key design difference that leads to this difference in incentives and responses is that vouchers have not yet been introduced in the TOV program (unlike in the VS program). Rather, the TOV program issues a threat and provides a window to respond, and an adequate response can completely obviate vouchers.

A rich theoretical literature analyzes multiple issues relating to school vouchers. See, for example, Nechyba (1996, 1999, 2000) for distributional effects of alternative voucher policies in a general equilibrium framework that endogenizes residential choice; Epple and Romano (1998, 2002) for the effect of vouchers and alternative voucher designs on sorting by income and ability; and Nechyba (2003) for the effect of vouchers on income segregation and school quality. While Epple and Romano (2002) allows for public school technical inefficiencies (exogenously) and Nechyba (2003) allows for efficiency gain (exogenously) in public schools facing competition from vouchers, the above literature does not model public school behavior.

Manski (1992) examines the impact of vouchers on public school expenditure and social mobility, while allowing for rent-seeking public schools. Using a theoretical and computational model that includes information asymmetries, and modeling public schools as rent maximizers, Ferreyra and Liang (2012) simulate the effect of two policies: public monitoring of public schools and private school vouchers. They argue that no single tool is sufficient, rather they argue that a combination of the two will dominate private school vouchers.[4,5] Modeling public school behavior and assuming rent maximizing public schools, McMillan (2004) shows that under certain circumstances, public schools may find it optimal to reduce productivity when a voucher is introduced. The current paper also models public school incentives, behavior and quality, however, unlike McMillan, derives the demand for public school from equilibrium household behavior, and also models peer quality. However, the most important difference with McMillan (2004) as well as rest of the literature above is that unlike the above studies, this paper models and examines the effects of the alternative U.S. voucher designs on public school incentives and performance, and, consequently, compares and contrasts the effects of two alternative U.S. voucher programs (that characterize the two different voucher designs implemented in the U.S.) on public schools.[6] Especially

[4] Neal (2009) also argues that choice and accountability should not be seen as policy substitutes, but argue in favor of combining both policies.

[5] It should be noted here that the combination of private school vouchers and public school monitoring modeled in Ferreyra and Liang (2012) is different from the TOV scheme modeled in this paper. In the vouchers–public monitoring scheme that they model, vouchers are already imposed. In addition, they supplement this voucher system with public monitoring. In contrast, the key feature of the TOV system is that vouchers are not already imposed. Rather, the policymaker issues a threat, or establishes a target/cutoff, and vouchers are introduced *only if* the school fails to meet that cutoff. Also, while the policy intervention in the case of TOV is announcement of a cutoff, the public monitoring in Ferreyra and Liang (2012) takes the form of "detailed evaluations of public school performance, direct observation of classroom and administrative practices etc.", and is costly.

[6] The Cleveland voucher program (implemented in the 1995-96 school year), the District of Columbia voucher program

with the recent plethora of publicly funded voucher programs implemented in the U.S., it is absolutely essential to understand the impacts of these two alternative voucher designs on public school incentives and performance.

In the empirical literature, there has been a recent spurt of studies that look at the effect of vouchers on public school performance and behavior. Greene (2001) and Greene and Winters (2003), respectively studying the effects of the 1999 Florida program and the revised Florida program (after the 2002 grading rule changes), find positive effects on treated schools. Figlio and Rouse (2006) find some evidence of improvement of the treated schools in the high stakes state tests, but these effects diminish in the low stakes, nationally norm-referenced test. West and Peterson (2006) study the effects of the revised Florida program and find positive effects on student performance in treated schools. Rouse et al. (Forthcoming) and Chiang (2009) investigate the effect of the latter program on public school behavior and find evidence in favor of behavioral changes in threatened schools, such as more focus on instruction and teacher development. Analyzing the impact of the Milwaukee voucher program on public schools after the Wisconsin Supreme Court ruling of 1998, Hoxby (2003a,b) find a positive productivity response to vouchers. The current study has been greatly informed by the Hoxby studies, and it follows Hoxby in the treatment-control group classification in Milwaukee. Hoxby (2003a,b) look at the Milwaukee program after the Supreme Court ruling of 1998. In contrast, (while I consider this phase of the Milwaukee program towards the end of the paper), the focus of this paper is on the Milwaukee program before the court ruling. This is because, except for the key design differences identified above, the program characteristics in Florida were most similar to the characteristics of the Milwaukee program in its first phase. Chakrabarti (2008b) compares the impact of the second phase of the Milwaukee program with that in the first phase on public schools, and finds evidence in favor of a larger response in the second phase.

The fundamental difference of the present paper with the empirical studies above is its focus on the impact of alternative voucher designs on public school performance. In particular, there is no study thus far (either theoretical or empirical) that seeks to compare public school response to the different

(implemented 2004-05) and the Indiana voucher program (implemented 2011-12) are also of the VS type, while the Ohio voucher program (implemented 2006-07) and Louisiana voucher program (implemented 2012-13) are TOV-type. Among the VS-type programs, the Milwaukee program is the oldest; while among the TOV-programs, Florida program is the oldest. Another crucial aspect is that apart from the differences in designs highlighted in the paper (VS versus TOV), the other features of the programs were very similar between the Florida and Milwaukee programs (which is essential for credible comparisons of the VS and TOV designs). As a result, I focus on the Milwaukee and Florida programs.

U.S. voucher designs. This study fills this important gap. Second, unlike most of the above empirical studies, this study controls for mean reversion, differences in pre-existing trends between schools, and controls for the possibility that changes in student composition of schools may bias the program effects.[7,8] Finally, unlike any of the above papers, this paper combines a theoretical and an empirical part,—the theoretical part designed to model the basic features of the Florida and Milwaukee voucher programs, and to compare and contrast the impacts of the two designs on public school performance; and the empirical part aimed at testing the theoretical predictions.

2 The Model

I construct a model that captures the basic features of the Florida and Milwaukee programs. The objective of the model is to investigate, in a simple framework, the effect of vouchers on public schools, and specifically whether the differences in the Florida and Milwaukee designs would have different effects on public school incentives and response. There are three agents in the model: (i) the public school, (ii) the private schools, and (iii) the households. The public school is free and offers quality (q) to all households that choose to attend it. Quality q is a composite of two factors: public school effort (e) and public school peer-group quality (b). In keeping with the school choice literature (Manski (1992), Hoxby (2003a), McMillan (2004), Ferreyra and Liang (2012)), the objective of the public school is to maximize net revenue ("rent") which is simply defined as revenue minus costs.[9] Public school revenue is given by $p{\cdot}N$, where p is the exogenously given per pupil revenue and N is the number of students in public school. Public school cost (C_p) is given by $C_p(N, e) = c_1 + c(N) + C(e)$, where c_1 is a fixed cost. Both $c(.)$ and $C(.)$ functions are assumed to be increasing and strictly convex in their respective arguments. I assume $p - c_N > 0$, that is the "net marginal revenue" per student is positive.

There is a continuum of private schools providing a continuum of quality levels. Thus, in keeping with the feature of the U.S. voucher programs, private schools do not choose between students who

[7] Exceptions are Figlio and Rouse (2006) that control for all three, Hoxby (2003b) that controls for differences in pre-existing trends and Peterson and West (2006) that controls for changes in student composition of schools. Unlike the present study, the analysis in Figlio and Rouse (2006) focuses on the low stakes test. While Greene (2001) and Greene and Winters (2003) attempt to assess mean reversion effect, their analysis is based on post-program effects of schools. Hence any mean reversion effect is likely to be confounded with post-program response of schools. For a more detailed literature review as well as effect on low stakes test scores in Florida, see Chakrabarti (2008a).

[8] For the remainder of the paper, I will refer to school years by the calendar year of the spring semester.

[9] All results continue to hold if public schools are assumed to maximize enrollment subject to cost. An alternative formulation could be to model the public school as a quality maximizer. However, in that case there would be no argument for voucher programs as far as improving public school quality is concerned.

apply with vouchers.[10] Note, though, that in reality private schools can choose whether or not to enter (participate). I abstract from this here for simplicity. At the end of section 3.1, I discuss what might happen if private schools could choose whether to enter. Households pay a tuition $T = t \cdot Q$ $(t > 0)$ to attend a private school of quality Q.

Households are characterized by an income-ability tuple (y, α), where $y \in [0, 1]$ and $\alpha \in [0, 1]$; y and α are assumed to be independently and uniformly distributed. A household obtains utility (U) from the consumption of the numeraire good (x), school quality (θ) and its ability (α). The household utility function is assumed to be continuous and twice differentiable and is given by $U(x, \theta, \alpha) = h(x) + \alpha u(\theta)$. The functions h and u are increasing and strictly concave in x and θ respectively.

School qualities available to a household are public school quality and a continuum of (exogenously given) private school qualities. Public school quality $q = q(e, b)$ is a continuous, twice differentiable, increasing and concave function of public school effort $e \in [e_{min}, e_{max}]$ and public school peer quality b. Public school peer quality is defined as the mean ability of the public school student body.[11] If a public school household decides to switch to a private school with vouchers, it incurs a positive switching or relocation cost c.

The paper models three alternative scenarios: (i) a simple public-private system (PP) without vouchers (the baseline), which can be thought of as the pre-program scenario for both programs; (ii) the Milwaukee-type "voucher shock" (VS) program; and (iii) the Florida-type "threat of voucher" (TOV) program. As discussed above, the Florida voucher system faced "F" schools with both "threat of vouchers" and stigma. Most of this section deals with the effect of TOV. The effect of stigma is discussed at the end of this section. As discussed later, one would expect the effect of stigma to work in the same direction as TOV and hence stigma effect would serve to further reinforce the effects obtained in this section.

The simple public-private system consists of two stages. In the first stage, the public school chooses effort. In stage 2, households choose between schools after observing the last stage public school effort. Peer-group quality and public school quality are determined.

[10] Note that at equilibrium, private school quality will always exceed public school quality. Otherwise, no household would pay to attend a private school. Chakrabarti (2009) shows that for voucher students, random selection by private schools has indeed taken place.

[11] Public school quality can be thought of as being embodied in public school scores. The notion here is that public school scores reflect both public school effort and public school peer-group quality, which in turn depends on the abilities of the public school students. In other words, both public school and student characteristics contribute to school scores.

The VS program is analyzed in three stages. In the first stage, the government announces voucher v. In stage 2, facing v, the public school chooses effort. In stage 3, households choose between schools (after observing v and e). Peer-group quality and public school quality are obtained.

The TOV program is modeled in four stages. In the first stage, the Government announces the program and a corresponding cutoff quality $\bar{q}$ and voucher v. In stage 2, facing the program, public school chooses effort. Given the existing peer group quality, q is realized. In stage 3, the government imposes vouchers v if $q < \bar{q}$. No voucher is imposed if $q \geq \bar{q}$. In the last stage, households choose between schools (after observing effort and whether vouchers were imposed). Peer-group quality and public school quality are realized.

Each of the systems constitutes a game between two players: public school and households. Facing the relevant program and correctly anticipating household behavior, public school chooses effort to maximize rent. In the last stage, after observing the program, public school effort and whether vouchers have been introduced, households anticipate a certain peer quality and choose between schools. At equilibrium, anticipated peer quality equals actual peer quality. This yields an equilibrium peer quality and a corresponding allocation of households between public and private sectors. Equilibrium public school quality (composite of equilibrium public school effort and peer quality) is simultaneously obtained.

An equilibrium of the "threat of voucher" program is an effort-peer quality tuple (e_{TOV}, b_{TOV}), such that given the quality cutoff $\bar{q}$ and voucher v (i) e_{TOV} is a public school equilibrium, given b_{TOV} and (ii) b_{TOV} is a household equilibrium, given e_{TOV}. The "voucher shock" equilibrium is a peer-group quality b_{VS} and an effort e_{VS} such that given voucher v (i) e_{VS} characterizes the public school equilibrium, given b_{VS} and (ii) b_{VS} characterizes the household equilibrium, given e_{VS}. The public-private equilibrium is characterized by an effort-peer quality tuple (e_{PP}, b_{PP}), where (i) e_{PP} is an equilibrium of the stage 1 game, given b_{PP} and (ii) b_{PP} is an equilibrium of the stage 2 game, given e_{PP}.

3 Characterization of the program equilibria

This section solves for the household and public school equilibria and compares the public school qualities under the PP, VS, and TOV equilibria.

3.1 Household behavior

This subsection analyzes household behavior under the three systems in a common framework. A household (y, α) chooses private school iff $D = [h(y + v - t \cdot Q^*(y, \alpha, .) - c) + \alpha u(Q^*(y, \alpha, .))] - [h(y) + \alpha u(q(e, b))] > 0$ where $Q^*(y, \alpha, .)$ is the optimal private school quality choice of household (y, α), and v equals zero under the pre-program public-private system and under the Florida TOV system if the public school escapes vouchers. It can be seen that $\frac{\partial D}{\partial y} > 0$ and $\frac{\partial D}{\partial \alpha} > 0$ which imply stratification by income and ability respectively.

Suppose all households expect a peer group quality $b^e \in [0, 1]$. Then for each y and given t, v, e, c and expected peer group quality $b^e \in [0, 1]$, there exists a unique household $0 < \hat{\alpha} < 1$ such that all households with lower ability choose the public school and those with higher ability choose a private school. This $\hat{\alpha}$ is the unique solution to the equation:

$$[h(y + v - t.Q^*(.) - c) + \alpha u(Q^*(.))] - [h(y) + \alpha u(q(e, b^e))] = 0 \qquad (3.1.1)$$

Since the indirect utility and the q functions are continuously differentiable and $D_\alpha > 0$, by the implicit function theorem, $\hat{\alpha} = \hat{\alpha}(y; v, e, b^e, t, c)$ is a continuously differentiable function. Using the implicit function theorem it is straightforward to check that for each income level, the cutoff ability level $\hat{\alpha}$ is decreasing in v and increasing in e, b^e, t and c. Given all other parameters, the cutoff ability level varies inversely with y. Given b^e, peer group quality b is given by:

$$b = \frac{\int_0^1 \int_0^{\hat{\alpha}(y, b^e, .)} \alpha d\alpha dy}{\int_0^1 \int_0^{\hat{\alpha}(y, b^e, .)} d\alpha dy} \Rightarrow \frac{1}{2} . \frac{\int_0^1 \hat{\alpha}^2(y, b^e, .)dy}{\int_0^1 \hat{\alpha}(y, b^e, .)dy} = g(b^e, e, v, t, c) \qquad (3.1.2)$$

At equilibrium b corroborates the initial conjecture b^e, that is, $b = b^e$ (3.1.3)

In other words, if all households expect a peer-group quality, then at equilibrium this expectation has to be fulfilled. Mathematically, given parameters e, v, t, c, a fixed point in b is reached. A household equilibrium always exists.[12] From (3.1.1)-(3.1.3), the equilibrium peer quality satisfies the equation $b^* = g(b^*, e, v, t, c)$. The corresponding equilibrium allocation of households between public and private sectors is characterized by $\hat{\alpha}(y, b^*, .)$ for $y \in [0, 1]$. $N(b^*, e, v, t, c) = \int_0^1 \int_0^{\hat{\alpha}(y, b^*, .)} d\alpha dy = \int_0^1 \hat{\alpha}(y, b^*, .)dy$ gives the corresponding number of students in public school at the household equilibrium b^*.

Equilibrium number of public school students decreases with vouchers and increases with public school effort. The proof is in appendix A. The intuition is as follows. An increase in public school effort

[12] For proofs and discussion of existence and uniqueness of equilibrium, see working paper version of this paper, Chakrabarti(2008a). The equilibrium is unique if the marginal utility from peer quality is not too high.

leads to an increase in the equilibrium cutoff ability level, $\hat{\alpha}(y, b^*)$, at each income level. This occurs through two channels. Given b^*, an increase in e induces households just above the cutoff at each income level to switch to the public school. This increases peer quality, leading to a further influx of higher ability households just above the cutoff from the private to the public sector. The consequence is an increase in the equilibrium number of students with effort. Vouchers acting directly as well as indirectly through peer quality induce a flight of high ability public school households at each income level to the private sector at equilibrium.

The analysis here assumes that when vouchers are imposed, all households, irrespective of income, become eligible for them. Although this is the case in Florida, in Milwaukee vouchers are targeted only to the low-income population. I abstract from this here for simplicity. All results continue to hold under targeted vouchers and are available on request. The results and the corresponding proofs are also available in Chakrabarti (2008a). Note that given other parameters (e, v, t, c), the number of public school students is less in a household equilibrium where all households are eligible rather than where only the low-income are. The reason is that in the former case there is a flight of households at each income level, whereas in the latter case it is restricted only to a subset of income levels.

It might be worth reflecting a little bit at this point on the other features of the model. As in the actual programs, public school revenue is directly tied to the number of students in the model, vouchers lead to loss in public school revenue and private schools cannot choose between students who apply with vouchers. However, in the actual programs while private schools cannot discriminate between voucher students, they can choose whether or not to participate in the program. I abstract from this in the paper for simplicity, but it is worth discussing what would happen in this case. In this new scenario, not all private schools will enter (if there is an entry cost), so we no longer have a continuum of private schools. However, at least some private schools are likely to enter and participate in the program (and this has been the case in both Florida and Milwaukee). Assuming at least some private schools enter, for each income y, we will again have a unique $\hat{\alpha}(y, v, .)$, and $\hat{\alpha}(y, v, .) \leq \hat{\alpha}(y, 0, .)$, with strict inequality holding at least for some y. Signs of the partial derivatives of $\hat{\alpha}$ with respect to its various arguments remain the same as earlier. It follows that the equilibrium number of public school students still falls with vouchers, and increases with an increase in public school effort. It might be worth noting here that the voucher amount compared very favorably to tuition, so the voucher amount was not a constraining factor to private school entry. Interestingly, Chakrabarti (2009) finds that the characteristics of private schools that actually participated in the program (voucher schools) were very similar to those of the private schools that elected not to participate (non-voucher schools).

3.2 Public School Behavior

The public school correctly anticipates behavior in all the future stages of the corresponding game, and chooses effort to maximize rent. The rent function is given by $pN(e,v) - c_1 - c(N(e,v)) - C(e)$.

Proposition 1 *Equilibrium effort of a public school under the "voucher shock" program can be either greater or less than its effort in the pre-program public-private equilibrium.*

The proof is in Appendix A. In the pre-program simple public-private equilibrium, marginal revenue equals marginal cost of effort at e_{PP}. Vouchers affect both marginal revenue and marginal cost in multiple ways and these effects together determine whether or not the public school increases effort. More precisely, equilibrium effort increases iff the following expression is positive: $[(p - c_N)N_{ev} - c_{NN}N_vN_e]$ (3.2.1). Vouchers decrease the number of public school students. Since the cost function is convex in the number of students, vouchers decrease marginal cost on this account. This is captured by the second term in (3.2.1). The first term captures the change in net marginal revenue due to vouchers. This can either increase or decrease with vouchers, thus rendering the effect on public school effort ambiguous.[13] Public school effort increases if either net marginal revenue increases or the decrease in marginal revenue is less than the decrease in marginal cost.

Proposition 2 *For each voucher v, there exists a cutoff effort level $\bar{e}$[14] such that the equilibrium effort (e_{TOV}) of a public school facing the "threat of voucher" program exceeds both*
(i) its equilibrium effort under the "voucher shock" program, e_{VS} and
(ii) its equilibrium effort under the public-private system, e_{PP}.

The proof is in Appendix A. The Florida-type TOV program affects public school incentives in a way very different from the Milwaukee-type VS program. A Florida public school facing the threat has two options: it can choose to meet the cutoff or it can choose not to meet the cutoff. In the latter case, it is

[13] $N_{ev} = \int_0^1 \left[\frac{\delta^2 \hat{\alpha}(y,b^*,.)}{\delta e \delta v} + \frac{\delta^2 \hat{\alpha}(y,b^*,.)}{\delta e \delta b} \frac{\delta b^*}{\delta v} \right] dy$. There are two effects. Vouchers lead to an exodus of relatively high-ability households (at each income level) to private schools, so that the new marginal household has a relatively lower marginal valuation of quality. Consequently, the number of students gained due to a marginal increase in effort is lower under vouchers. This is captured by the negative first term. Second, since the marginal utility from school quality decreases with quality ($u_{qq} < 0$) the marginal number of students due to an increase in effort decreases with an increase in peer quality ($\frac{\delta^2 \hat{\alpha}(y,b^*,.)}{\delta e \delta b} < 0$). Since vouchers lead to a fall in peer quality, the marginal number of students increases due to this factor (which is captured by the positive second term).

[14] Note that since peer quality is known, announcing a cutoff in terms of effort is equivalent to announcing a corresponding cutoff in terms of quality.

in the same state as its counterpart under the VS program. It chooses the VS optimum effort e_{VS} and gets the VS rent, $R(e_{VS}, v)$. Since vouchers decrease rent, it follows that the school can be induced to satisfy a cutoff $\bar{e}$ strictly higher than e_{VS}, where the rent from $\bar{e}$ without vouchers exactly equals the rent from e_{VS} with vouchers. Thus, the fundamental feature of the TOV that induces a higher effort is that vouchers are not already imposed and a sufficient improvement can enable schools to escape vouchers. Note that any cutoff in the range $(e_{VS}, \bar{e}]$ induces an effort under the TOV program that is strictly higher than under the VS program. For the sake of simplicity and to avoid messy computations, the analysis here assumes that all households irrespective of income are eligible for vouchers under the VS program. However, this result also holds for vouchers targeted only to the low-income population under VS, as is the practice in Milwaukee. The formal proof is available on request and is available in Chakrabarti (2008a).[15]

Now consider the intuition behind the second part of proposition 2. The Florida TOV program introduces a discontinuity in the rent function at the cutoff effort level. If the cutoff is set at e_{VS}, then meeting it gives a higher rent than choosing to accept vouchers. Since e_{PP} is the rent maximizing effort under $v = 0$, setting the cutoff at e_{PP} gives an even higher rent to the public school. Given the strict concavity of the rent function, there exists a cutoff $\bar{e} > e_{PP}$ which satisfies the school's incentive constraint. Again, any cutoff in the range $(e_{PP}, \bar{e}]$ induces an effort under the TOV program that is strictly higher than under the PP program equilibrium.[16] As appendices B and C show, these results continue to hold when effort is not observable, but quality is, and there is no one-to-one relationship

[15] The intuition can be laid down in two steps. Call the VS program where all students are eligible the "universal voucher shock" (UVS) program and where only the low-income students are eligible the "targeted voucher shock" (TVS) program, and the corresponding equilibrium number of students and equilibrium effort N_{UVS}, N_{TVS} and e_{UVS}, e_{TVS} respectively. First, note that the equilibrium rent under the TVS is greater than that under the UVS. Under the TVS, the school can attract N_{UVS} students by giving a lower effort than under the UVS (follows from the discussion in page 10 and claim 1 in appendix A) and hence at a rent higher than under the UVS. Since the school chooses to attract N_{TVS} students, it must be the case that rent is higher under the TVS. Second, if vouchers when imposed in the Florida-type TOV program took a targeted form, then following the argument in proposition 2(i), the program could implement a cutoff $\bar{\bar{e}} > e_{TVS}$. But vouchers take the universal form in Florida, which implies that the rent would be smaller than the TVS rent if the school failed to meet the cutoff. This implies that there exists a cutoff $\bar{e} > \bar{\bar{e}} > e_{TVS}$ which satisfies the school's incentive constraint with equality and hence can be implemented by the TOV program. Any cutoff in the range $(e_{TVS}, \bar{e}]$ induces higher effort under the TOV program than under the TVS program. To summarize the above discussion, there are two features in the design of the Florida TOV that induce a higher effort than the TVS: (i) vouchers are not already imposed and (ii) the potential loss of students is greater. But, as it follows from the above discussion, the first factor is sufficient to induce a higher effort under the TOV.

[16] In the TOV program it may be reasonable to think that there is a stigma attached to being labeled as a 'voucher public school'. For example, Maureen Backentoss, assistant superintendent of curriculum and instruction of Lake County School District refers to it as a "glass of cold water in the face". In the presence of such a stigma, the public schools gain an additional utility if they are able to escape vouchers. This feature is absent in the VS program. Note that this will weigh results in favor of the TOV and will induce an even higher improvement under the TOV.

between the two. But the cutoff can no longer be set in terms of effort, which is now unobservable.

The basic intuition behind Proposition 2 can also be illustrated in a simple diagram. Consider Figure 1 Panel A. Let R_1 denote the net revenue function of the public school under the baseline public-private system. Then equilibrium effort under the public-private system is given by e_{PP}, and the equilibrium rent by R_{PP}. Milwaukee-type vouchers lead to a downward shift of the net revenue function, and let the new net revenue function be R_2. The equilibrium effort under VS is then given by e_{VS}, and the corresponding rent by R_{VS}. Panel A illustrates the scenario where $e_{PP} > e_{VS}$, while Panel B illustrates the scenario where $e_{PP} < e_{VS}$. As can be seen from the Figure, any target effort under the TOV program in the range $(e_{VS}, \bar{e}]$, leads to an equilibrium public school effort under TOV that is strictly higher than e_{VS}. On the other hand, any target effort in the range $(e_{PP}, \bar{e}]$ leads to an equilibrium effort under TOV that is strictly higher than e_{PP}. For example, consider an effort cutoff of e_{TOV}. If the public school matches this effort, it gets rent R_{TOV}. In contrast, if it fails to meet this cutoff, it gets a lower rent, R_{VS}. It follows that the public school matches e_{TOV} and gets rent R_{TOV}, and $e_{TOV} > e_{VS}$ and $e_{TOV} > e_{PP}$.

Using propositions 1 and 2 and the properties of the household equilibrium (see claim 1 in appendix A), the result below follows.

Corollary 1 *(i) Equilibrium quality of a public school facing "threat of vouchers" :*
(a) exceeds its equilibrium quality under the pre-program public-private system.
(b) exceeds its equilibrium quality under the "voucher shock" program.
(ii) Equilibrium quality of a public school under the "voucher shock" program can be either greater or less than its quality in the pre-program public-private equilibrium.[17]

Note that the statements in proposition 2 and corollary above are not statements relating to the public school system as a whole. Rather they relate to a school facing the threat of vouchers in Florida (a school that has got the first "F", a "threatened" school) and the same school facing a voucher shock program in Milwaukee. Thus the intended comparison is between a threatened school (alternatively referred to as "more treated" school in the empirical part) in Florida and a demographically and socioeconomically similar school ("more treated" school in the empirical part) in Milwaukee.

It is also worth noting here that while the above theoretical analysis on Florida focuses on the "threat of vouchers", the Florida voucher system has two components: "threat of vouchers" and stigma associated with getting the lowest performing grade "F". But, in the presence of such a stigma effect, the schools get higher utility if they are able to escape an "F". Thus, stigma affects incentives in the

[17] It follows from the above discussion that the corollary holds not just for the UVS system, but also the TVS system.

same direction as "threat of vouchers" and further induces schools to exert higher efforts. Hence, this implies that the equilibrium effort and quality of threatened public schools under the Florida voucher system would be even greater than that obtained above. Formal proof of this result including stigma effect is available on request.

4 Data

The data for this paper come from multiple sources. The Florida data consist of school-level data on test scores, grades, socio-economic characteristics of schools and school finances and are obtained from the Florida Department of Education (DOE). Data on socio-economic characteristics include data on sex-composition (1994-2002), percentage of students eligible for free or reduced-price lunches (1997-2002) and are obtained from the school indicators database of the Florida DOE. (As noted earlier, this paper refers to school years by the calendar year of the spring semester.) School finance data consist of several measures of school level and district level per pupil expenditures and are obtained from the school indicators database and the Office of Funding and Financial Reporting, Florida DOE.

School-level data on test scores are available on the Florida Comprehensive Assessment Test Sunshine State Standards (FCAT-SSS). (This test will be referred to as the FCAT in the remainder of the paper). The FCAT reading and math tests were first administered in the year 1998. Mean scale scores (on a scale of 100-500) on grade 4 reading and grade 5 math are available for 1998-2002. Mean scale scores (on a scale of 1-6) on the Florida grade 4 writing test, which was first administered in 1993, are available from 1994-2002.

The Wisconsin test scores analysis reported in this paper is obtained from Chakrabarti (2008b)—a more detailed data description is available there. The results reported in this paper (based on Chakrabarti (2008b)) use school-level data on socioeconomic characteristics, real per pupil expenditure, and test scores for (i) the Wisconsin Reading Comprehension Test (WRCT) (percentage of students above (% above) standard, 1989- 1996) (ii) the grade 5 Iowa Test of Basic Skills (ITBS) Reading (1987-1993) (iii) the grade 5 Iowa Test of Basic Skills (ITBS) Math (1987-1997) and (iv) Wisconsin Knowledge and Concepts (WKCE) Examination (1997-2002). The data are obtained from the Wisconsin Department of Public Instruction (DPI), the Milwaukee Public Schools (MPS), and the Common Core of Data (CCD) of the National Center for Education Statistics (NCES).

I also use data on addresses of public schools, private schools and participating choice schools in both Florida and Wisconsin. They are obtained respectively from the public and private school universe surveys of the NCES, the Florida DOE and Wisconsin DPI.

5 Empirical Strategy

5.1 Linking Theory with Empirics

The empirical part of the paper seeks to test the following two theoretical predictions: (i) A Florida-type voucher system will lead to an increase in quality of the threatened public schools. (ii) Quality improvement of threatened public schools in the Florida-type system will exceed that of (demographically/socioeconomically) similar treated public schools in the Milwaukee-type program. School quality is proxied by school scores.

Before moving on to the empirical part of the paper, I investigate the comparability of the settings in Florida and Milwaukee. First, an important factor is whether the more treated (or F) schools under the Florida TOV program were similar in terms of pre-program characteristics to the more treated schools under the Milwaukee program. Appendix Table D.1 reports the pre-program characteristics of more treated and control schools in Florida and Wisconsin. As can be seen, the more treated schools in Florida were indeed similar to the more treated schools in Wisconsin and except in one case, the differences between them were not statistically significant. Similarly, the control schools in Florida were similar to the control schools in Wisconsin and the differences between them were never statistically significant.

However, the treated schools were somewhat different from the control schools within each place. This is because Wisconsin schools outside Milwaukee were considerably more advantaged than Milwaukee schools. I ended up with this control group in spite of the strategy (following Hoxby (2003a,b)) of picking control schools as similar as possible to the Milwaukee more treated schools in terms of pre-program characteristics.

For purposes of comparison of program effects across the two programs, I use the C schools in Florida (described in next section) as the control group in Florida. This is because not only do the more treated schools need to be similar across the two programs in terms of pre-program characteristics, but it is important to have the control groups similar as well across the two places. As can be seen, the control group in Wisconsin was indeed very similar to the C schools in Florida and did not statistically differ from them in terms of any of the characteristics (Table D.1). Still another advantage of picking the C schools as the control group in Florida is that while the D schools were more similar to the more treated (F) schools in terms of grade and demographics, they were very close to getting an F grade and hence to some extent perceived an indirect threat and hence were to some extent treated by the program.

Because of the differences between the treated and control schools, one might argue that in the absence of the program the control group would have evolved differently from the more treated group. However, I have multiple years of pre-program data, and can check (and control) for any differences in pre-program trends of the treated and the control groups. This will get rid of any level differences between the treatment and control groups, and will also control for differences in pre-program trends, if any. It seems likely that once I control for differences in trends as well as in levels, any remaining difference between the treatment and the control groups will be minimal. In other words, my identifying assumption is that if the treated schools followed the same trends as the control schools in the immediate pre-program period, they would have evolved similarly in the immediate post-program period too in the absence of the program.

I also checked for differences in other dimensions across the two places. One important factor to consider is the extent of pre-existing competition (both private and public). Using public and private school addresses in the preprogram period, geocoding them and computing relevant distances, I find that more treated schools faced relatively more competition (both public and private) in both places, but the extent of relative competition was somewhat greater in Milwaukee than in Florida.[18] Note that I already control for differences in pre-existing levels and trends between schools in each of the two places, which, to some extent, takes care of these differences. However, the differences in relative pre-existing competition might affect treated schools differently in the two places in the post program period. Since relative competition faced by the Milwaukee more treated schools was greater than that in Florida, the estimates of the TOV effect in the paper as obtained from Florida will be underestimates. In this case, the difference in effects (on more treated schools) between TOV and TVS may be actually larger than that suggested by the empirical estimates in the paper. I also rerun the regressions (described in the following section) after including interactions of the program dummy respectively with pre-program numbers of private schools within one mile radius and pre-program numbers of public schools within one mile radius in each of the two places, so as to control for any differential effects that the differences in

[18] The measure of public (private) competition here is the number of public (private) schools within a certain radius. I use 1, 2, 3 and 5 mile radii.

pre-existing competition may have after the program. The results remain qualitatively similar and are available on request.

Another issue is the extent of post-program competition. While the Milwaukee program allowed voucher students to move to participating private schools, the Florida program allowed them to move to participating private schools as well as high performing (A, B, C) public schools. Therefore, to assess the extent of post-program competition, I investigate whether the total number of participating private schools and high performing public schools around an average more treated school in Florida was comparable to the number of participating private schools around an average more treated school in Milwaukee. I find that there were 0.67 participating private schools within 1 mile radius of an average more treated school in Milwaukee, while there were 0.60 participating private schools and high performing public schools within 1 mile radius of an average threatened school in Florida. These numbers are not very different,—if anything, competition in Florida was little less than in Milwaukee, so that the empirical estimates of the response of the more treated schools under TOV may be once again underestimates in that sense—that is, the differences in responses of the more treated schools under TOV versus TVS may be actually larger than what the paper suggests.

Fourth, since Milwaukee is an urban district, I repeated my analysis by comparing the improvement of more treated schools in Milwaukee with that of threatened schools in the largest urban district in Florida, Miami Dade County. The results are similar to those below and are available on request. The above discussion suggests that comparison of program effects in the two places can serve as a test of the relevant theoretical prediction. However, it is worth noting here that if there are factors or settings that affect more treated groups differently (relative to the control schools) *only* in the post-program period and *this phenomenon is asymmetric* across Florida and Milwaukee, then this might bias my empirical results. Any pre-existing differences in trends or levels are controlled for, but if there are differences that arise only in the post-program period, affect the more treated schools differently, and are unique to only one of the places, then these can act as potential confounding factors. While the preceding analysis in this section as well as a variety of sensitivity checks later give confidence to the results, because of this potential caveat, the results should be taken as strongly suggestive. If we indeed find in the empirical analysis that the treatment effects in Florida exceed the corresponding effects in Milwaukee, then that would indicate that the empirical findings are consistent with the theoretical predictions, and that the empirical analysis provide evidence and support in favor of the theoretical predictions.

5.2 Samples and Specifications

In Florida, the schools that received an "F" grade in 1999 were directly exposed to stigma and the threat of vouchers. These schools constitute the group of treated schools and will be referred to as "F schools".[19] The schools that received a D grade in 1999 were closest to the F schools in terms of grade but were not directly treated by the program. These schools will constitute my initial group of control schools and will be referred to as "D schools". Since the program was announced in June 1999 and the grades were based on the tests held in February 1999, the classification of schools into treatment and control groups is made here on the basis of their pre-program scores and grades. Using pre-program data, I first test whether the F and D schools exhibited similar trends before the program. If they had similar pre-program trends, I use the following set of specifications to investigate whether the F schools demonstrated a higher improvement in test scores in the post-program era. If the F schools demonstrated a differential pre-program trend, in addition to estimating these specifications, I also estimate their modified versions where I control for their pre-program differences in trends.

$$s_{it} = f_i + \alpha_0 t + \alpha_1 v + \alpha_2 (F * v) + \alpha_3 (v * t) + \alpha_4 (F * v * t) + \alpha_5 X_{it} + \epsilon_{it} \tag{1}$$

where f_i denotes school fixed effects, t is time trend, v is program dummy, $v = 1$ if year > 1999 and 0 otherwise. The variables v and $v * t$ respectively control for post-program common intercept and trend shifts. The coefficients on the interaction terms $F * v$ and $F * v * t$ estimate the program effects—α_2 captures the intercept shift and α_4 the trend shift of F schools. X_{it} denotes the set of school characteristics.

I also estimate a completely unrestricted model that includes year dummies to control for common year effects and interactions of post-program year dummies with the F school dummy to capture individual post-program year effects. This specification no longer constrains the post-program year-to-year gains of the F schools to be identical and allows the program effect to vary across the different years. The coefficients $\gamma_{1i}, i = \{2000, 2001, 2002\}$ capture the post-program year effects.

$$s_{it} = f_i + \sum_{i=1999}^{2002} \gamma_i D_i + \sum_{i=1999}^{2002} \gamma_{1i} (F * D_i) + \beta_2 X_{it} + \epsilon_{it} \tag{2}$$

The above specifications assume that D schools are not affected by the program. Although D schools did not face any direct threat from the program, they might have faced an indirect threat since they

[19] I restrict my analysis to elementary schools in both Florida and Milwaukee as there were too few treated middle and high schools in the respective places to justify analysis.

were close to getting "F".[20] Therefore, I next allow the F and D schools to be different treated groups with varying intensities of treatment ("more treated" and "less treated" groups respectively). I compare their post-program improvements with 1999 C schools (C schools from now on), which were the next higher up in the grade scale, using the above three specifications after adjusting for another treated group. It should be noted here that since both D and C schools might have faced the threat to some extent, my estimates may be underestimates (lower bounds), but not overestimates. All specifications I describe here are fixed effects regressions. I also estimate OLS counterparts of each of these specifications which also include dummies for the different treated groups. An important concern is that the treatment effects based on the above methodology may be biased due to the presence of potentially confounding factors such as mean reversion, sorting, presence of stigma effect associated with grade F. These issues are discussed in section 6.2.

The empirical analysis and the results for the Milwaukee part are obtained from Chakrabarti (2008b). The treatment–control classification strategy there follows Hoxby (2003a,b). It uses the basic Hoxby (2003a,b) intuition that the extent of treatment of the Milwaukee schools depended on the percentages of their students eligible for free or reduced price lunches, since the free or reduced price lunch eligible students of the MPS were the ones eligible for vouchers.[21] I classify the schools into three treatment groups based on their pre-program (1990) percentages of free or reduced price lunches—schools that had at least 66% of their students eligible as "more treated"; schools with eligibility between 66% and 47% as "somewhat treated"; and schools with less than 47% eligibility as "less treated". I refer to this sample as the 66-47 sample. For robustness of the results to alternative sample specifications, see Chakrabarti (2008b). I focus on the 66-47 sample in this paper because the baseline characteristics of the more treated (control) group in this sample are most similar to those of the more treated (control) schools in Florida.

The control group criteria are also based on Hoxby (2003a,b). For similarities and differences of this treatment–control strategy with Hoxby (2003a,b), see Chakrabarti (2008b). Since all schools in Milwaukee were potentially affected by the program, the control group consists of Wisconsin schools

[20] In fact, there is some anecdotal evidence that D schools might have responded to the program. The superintendent of Hillsborough county, which had no F schools in 1999, announced that he would take a 5% pay cut if any of his 37 D schools received an F on the next school report card. (For more evidence, see Innerst, 2000).

[21] While 175% was the cutoff poverty level for eligibility, this cutoff was not strictly enforced (Hoxby (2003b)) and households within this 10% margin (175% and free or reduced price lunch eligibility cutoff 185%) were often allowed to apply. Also there were very few students who fell in the 175%-185% range.

outside Milwaukee that: (i) had at least 25% of their population eligible for free or reduced-price lunch (ii) had black students compose at least 15% of the school population, and (iii) had locales (as defined by CCD) as similar as possible to the Milwaukee schools. Since Wisconsin schools outside Milwaukee were much more advantaged, I could find only 33 such elementary schools,—they were also geographically located close to Milwaukee.

The empirical methodology employed for Milwaukee is the same as above,—I estimate OLS and fixed-effects versions of the two specifications (1)-(2) after adjusting for the relevant years, number of treatment groups and controlling for (any) differences in pre-program trends. The baseline characteristics of the more treated and control groups in the two places are shown in Table D.1—the more treated and control groups in Wisconsin were demographically/socioeconomically very similar to the corresponding groups in Florida.[22]

6 Results

6.1 Obtaining the Treatment Effects in Florida and Milwaukee

In Florida, investigation of pre-program trends reveals that in reading and math, F schools exhibit no differential trends with respect to D schools, although they exhibit a negative differential trend relative to C schools in reading. In writing, F schools exhibit negative differential trends relative to both D and C schools. (These results are not reported here but are available on request.) Whenever there is a difference in pre-program trends, the regressions reported control for these differences by including interactions between trend and the respective treatment dummies.[23] Table 1 presents the effects of the Florida TOV program on F school reading, math and writing scores as compared to the D schools. For reading, the first two columns report results from the linear model 1, and the final two columns from the

[22] Note that the test scores under consideration in Florida are high stakes FCAT test scores. Since the threat in the Florida program was given in terms of grade, the *response* of the Florida threatened schools has to be assessed in terms of the high stakes test. Since this study is interested in examining the *response* of the threatened schools, it focuses on the high stakes tests in Florida. For example, even if there is no improvement in the low stakes test, it cannot be concluded that the public schools did not respond to the program. The test under consideration in Wisconsin are state tests (WRCT or WKCE) and district test (ITBS). These tests are low stakes. Note that the comparison of the effects on the more treated schools in the two places in terms of these tests is legitimate as the Florida FCAT tests are high stakes to the threatened schools precisely because of the Florida accountability-tied voucher system, that is, the Florida accountability-tied voucher program had the capacity to make the corresponding tests high stakes unlike the Milwaukee-type voucher program (that did not have an accountability component.)

[23] When data are available for only two years before program (for example, reading and math), the pre-program difference between treatment and control groups can be either a trend difference or a year effect. Specification 1 controls for this pre-program difference assuming it is a trend difference, and specification 3 controls for it assuming this difference is a year effect. Results from regressions without controlling for these pre-program differences are qualitatively similar.

non-linear model 2. Both OLS and fixed effects (FE) estimates in the first two columns show positive intercept and trend shifts for the F schools , although the latter is not significant in the fixed-effects estimate. These effects are disaggregated in columns (3) and (4) where the coefficients reflect the effects of the program after one, two and three years. Both the OLS and fixed-effects estimates show positive and significant year effects in each of the years after program.

For math and writing, the first column reports results from the linear model 1, and the second column from non-linear model 2.[24] In math, there is a positive, significant, large intercept shift after the program although there is no evidence of any trend shift. Column (6) shows evidence of positive significant F school year effects in math in each of the three years after the program. In writing, column (7) shows positive and statistically significant intercept and trend shifts for the F schools. The last column shows positive, significant year effects in writing in each of the three years after the program.

Next, considering D schools as an additional treated group, Table 2 looks at the effect of the program on F (more treated) and D (less treated) schools as compared to C schools. For each subject, the first two columns present results from model 1, the last two columns from model 2. In reading, F schools exhibit positive significant trend and intercept shifts that exceed the corresponding shifts of D schools. In both math and writing, F schools exhibit positive, significant and large intercept shifts that are statistically greater than that of D schools. Results from the unrestricted model show positive significant year effects in reading, math and writing for F schools in each of the years after program. Although many of the D school effects are also positive significant, the F school shifts are both economically and statistically larger in each of the years.[25] Since the C schools are most similar to the Wisconsin control schools, the estimates of treatment effects obtained from this sample are used for comparison with the corresponding Milwaukee effects. To put the above effects into perspective, the standard deviation of FCAT reading scores in this sample is 20, and the standard deviations of FCAT math and writing scores respectively are 20 and 0.53. To summarize the above results, using different samples, different subjects, different specifications, and both OLS and FE estimates, the results above show considerable improvement of F schools after the program. Although D schools show non-negligible improvement, their improvement is

[24] In many of the tables, only the fixed effects estimates are reported. The OLS results are very similar to the FE estimates and hence are omitted.

[25] It should be noted here that both F and D schools received additional funds from the state. However, all results above are obtained after controlling for real per pupil expenditure. The results are not sensitive to inclusion of real per pupil expenditure, nor do they change after including a polynomial in real per pupil expenditure. Moreover, even in the pre-1999 period, the critically low performing schools received extra assistance,—however this did not result in improved performance of this group in this period, as Table 4 shows in a different context.

smaller and also statistically different from those of F schools.

The results for Milwaukee are obtained from Chakrabarti (2008b), and the relevant ones are collated in Table D.2. Using the 66-47 sample, it presents the effects of the Milwaukee "voucher shock" program on WRCT (% above), ITBS reading, and ITBS math scores of different treatment groups. Except the positive and statistically significant effect in WRCT reading in its second year, there is no other statistically significant evidence of any effect of the program. Although many of the effects are positive, they are often not statistically significant and do not always have the right hierarchy (example, the somewhat treated effects often exceed the corresponding more treated effects). Thus the results in Milwaukee are mixed. To put the effects into perspective, the standard deviation of WRCT (% above) scores in this sample is 16, and the standard deviations of ITBS reading and math scores respectively are 18.45 and 16.71.

6.2 Investigating Potentially Confounding Issues

Before moving to the comparison of treatment effects, it is essential to investigate whether the effects above are biased due to the presence of potentially confounding factors, for if so, such a comparison may not serve as a valid test for the second prediction.

Mean Reversion

Mean-reversion is the statistical tendency whereby high or low scoring schools tend to score closer to the mean subsequently. Since the F schools were low scoring in 1999, a natural question to ask would be whether the improvement in Florida is driven by mean reversion rather than the program. Since I do a difference-in-differences analysis, my estimates will be contaminated by mean reversion only if F schools revert to the mean to a greater extent than the D schools and/or the C schools.

For a first pass at the mean-reversion issue, following Kane and Staiger (2002), I investigate whether the schools that were low scoring in 1998 were also low scoring in 1999. Interestingly, in each of reading, math and writing, more than 70% of schools that ranked in the bottom tenth percentile in 1998 also ranked in the bottom tenth percentile in 1999. This implies that although there may be mean reversion, it may not be a major problem.

A more direct way to approach mean-reversion is to check by how much the schools that received an "F" grade in 1998 improved during 1998-1999 compared to those that received a "D" (or "C") grade in 1998. Since this was the pre-program period, the gain can be taken to approximate the mean reversion effect and can be subtracted from the post-program gain of F schools compared to D schools (or C schools) to get at the mean-reversion corrected program effect.

The accountability system of assigning letter grades to schools started in the year 1999. The pre-1999 accountability system classified schools into four groups I-IV (I-low, IV-high). However, using the state grading criteria and data on percentage of students in different achievement levels in each of FCAT reading, math and writing, I was able to assign letter grades to schools in 1998.

The state assigned school grades based on FCAT reading, math and writing scores. In FCAT reading and math, it categorized students into five achievement levels (1-5) that correspond to specific ranges on the raw-score scale. Using current year data, it designated a school an "F" if it was below the minimum criteria in reading, math and writing, a "D" if it is below the minimum criteria in one or two of the three subject areas, and "C" if it is above the minimum criteria in all three subjects but below the higher performing criteria in all three. In reading and math at least 60% (50%) of the students had to score level 2 (3) and above while in writing at least 50% (67%) had to score 3 and above to meet the minimum (higher performing) criteria in that respective subject. The schools that were assigned grades "F", "D" or "C" in 1998 using this criteria will henceforth be called the 98F schools, 98D schools, and 98C schools, respectively. Using data for 1998 and 1999, Table 3 Panel A finds that in comparison to 98D schools, 98F schools show no evidence of mean reversion in reading or math, although there is mean reversion in writing. In comparison to 98C schools (Panel B), there is no evidence of mean reversion in reading; both 98D and 98F schools show comparable amounts of mean reversion in math; and only 98F schools show mean reversion in writing.

An alternative way to get around the problem of mean reversion is to do a regression discontinuity (RD) analysis. I briefly describe the strategy here, but the results are not reported here for lack of space. They are available in Chakrabarti (2008a) and are also available on request. The Florida program created a highly non-linear and discontinuous relationship between school achievement and the probability that the school's students would become eligible for vouchers in the near future. Exploiting this, the strategy is to compare F schools just below the cutoff between "F" and "D" with D schools just above the cutoff. Based on the state grading criteria, I construct a discontinuity sample where both F and D schools failed to meet the minimum criteria in reading and math in 1999, while in writing, only F schools failed the minimum criteria. In this sample, the probability of treatment varies discontinuously as a function of a continuous variable, the percentage of students scoring at or above 3 in 1999 FCAT writing. There is a sharp cutoff at 50%. Using this sample, there is a discontinuous relationship between assignment to treatment (i.e. facing the threat of vouchers) and schools' percentages of students scoring at or above 3 in writing (Chakrabarti 2008a).[26] The results from the regression discontinuity analysis are consistent and qualitatively similar to those from the above analysis (and are available on request).

Although the Milwaukee program is not conditional on low performance of schools, the more treated schools were also among the lowest scoring schools in each of the subject areas before the program. However, there is no evidence of any mean reversion in any of the subject areas in Milwaukee (see Chakrabarti (2008b)).

"Threat of Vouchers" Versus Stigma

The objective of this paper is to compare the effect of the whole Florida voucher system (that incorporates both "threat of vouchers" and stigma) with the effect of the Milwaukee voucher program on comparable treated schools. Still, it might be instructive to try to separate out the two effects in Florida. In this section, I investigate whether it is possible to say whether the above effects on the Florida threatened schools were caused by the "threat of vouchers" or stigma. I use the following strategies to investigate this issue.

First, although the system of assigning letter grades to schools started in 1999, Florida had an accountability system in the pre-1999 period which categorized schools into four groups 1-4 (1-low, 4-high) based on FCAT writing, and reading and math norm referenced test scores. Using FCAT writing data for two years (1997 and 1998), I investigate whether schools that were categorized in group 1 in 1997 improved relative to the 1997 group 2 and group 3 schools during the period 1997-98. The rationale is that if there was a stigma effect of getting the lowest performing grade, group 1 schools should improve in comparison to the group 2 and 3 schools even in the absence of the TOV program. Table 4 shows that there is no evidence that this has been the case.[27]

Second, all schools that received an F in 1999 received higher grades (A,B,C,D) in the years 2000, 2001. Therefore although stigma effect on F schools might have been operative in 2000, this was not

[26] I also consider two corresponding discontinuity samples where both groups fail the minimum criteria in reading and writing (math and writing). F schools fail the minimum criteria in math (reading) also, unlike D schools. In these samples, the probability of treatment changes discontinuously as a function of the percentage of students at or above level 2 in math (reading) and there is a sharp cutoff at 60%.

[27] I do not use the pre-1999 reading and math norm referenced test (NRT) scores because different districts used different NRTs during this period, which varied in content and norms. Also districts often chose different NRTs in different years. Thus these NRTs were not comparable across districts and across time. Moreover, since districts could choose the specific NRT to administer each year, the choice was likely related to time varying (and also time-invariant) district unobservable characteristics which also affected test scores.

likely to have been the case in 2001 or 2002 since none of the F schools got an F in the preceding year. Yet, as shown earlier F schools exhibited gains in both 2001 and 2002. Since F schools continued to face the threat of vouchers till 2002, this provides further evidence in favor of the TOV effect and against the stigma effect.

Third, I also use another strategy to investigate this issue. This strategy exploits the relationship between private school distribution around threatened schools and its relationship with threatened school response.[28] F schools that had more private schools in their near vicinity would likely lose more students if vouchers were implemented, and hence would face a greater threat of vouchers than those that had less. However, since stigma was a "bad" label associated with F, these schools would face the same stigma. Therefore if the response was caused by "threat of vouchers", then one would expect to see a greater response from F schools that had more private schools in their near vicinity. This, however would not be the case if the response was driven by stigma. To investigate this issue, I exploit the pre-program distribution of private schools, and investigate whether threatened schools that had more private schools in their immediate vicinity showed a greater response. Interestingly, this indeed was the case. The results are not reported here for lack of space, but are available on request.

To summarize, the above analysis suggests that the F-school effects obtained above were driven by "threat of vouchers" rather than stigma. Note though that even otherwise, the effects obtained above capture the effects of the whole program on F-schools, that is the effects of a combination of threat of vouchers and stigma generated by a voucher system that embeds vouchers in an accountability framework. As outlined above, *the objective of the paper is to identify the effect of the whole Florida-voucher system, that is the effect of an accountability tied voucher program on the threatened schools, and compare this effect with the corresponding effect of the Milwaukee program on similar schools.*

Sorting

Vouchers affect public school quality not only through direct public school response but also through changes in student composition and peer quality brought about by sorting. All these three factors get reflected in public school scores. This issue is important in Milwaukee since over the years students left the MPS with vouchers. In Florida, on the other hand, no school became eligible for vouchers in the years 2000 or 2001. Therefore the effects in Florida (for each of the years 2000, 2001 and 2002) were not likely to have been affected by this factor.

[28] I would like to thank David Figlio for suggesting this strategy.

First, each of the regressions control for demographic composition of schools. However any change in student composition in terms of unobservable factors may not be controlled for by these factors. It may be noted that inclusion of demographic controls do not change results by much, either in Florida or in Milwaukee. Also, to investigate this issue, using appropriately modified versions of models (1)-(2) (and with the dependent variable replaced by demographic variables), I examine whether demographic composition of the different Milwaukee treated groups changed after the program relative to the control schools. I do not find such evidence (Chakrabarti (2008b)).

While vouchers leading to sorting was not applicable in Florida, F, D, C grades could lead to differential sorting of students in these types of schools.[29] Note that the effects above would be driven by sorting only if F schools faced a relative flight of low performing students and/or a relative influx of high performing students in comparison to D or C schools. There is no a priori reason as to why this might happen. However, to investigate this issue further, I repeat the same exercise (as above) in Florida also,—there is no evidence of any relative shift of the demographic composition of F schools in comparison to D or C schools after the program. (The results are available on request.)

6.3 Comparing the Treatment Effects in Florida and Milwaukee

In this section, I compare the effects of the Florida and Milwaukee programs on the respective more treated schools in reading and math both before and after correcting for mean reversion. (No writing test data are available in Milwaukee during the relevant period.) The results are presented in Table 5—the figures are based on those in Tables 2, 3 and D.2, and all figures are expressed in terms of respective sample standard deviations. The comparison results presented here correspond to model 2, results from model 1 present a similar picture. Pre-correction results show positive and significant effect sizes in Florida in each of the years and subject areas which exceed the corresponding Milwaukee effect sizes economically in every case. The Florida effects are also statistically different from the Milwaukee effects in the first and third years after program, in both reading and math.

Mean reversion corrected effect sizes are obtained by subtracting the effect size attributed to mean reversion (obtained from expressing the relevant coefficients in Table 3, panel B in terms of respective standard deviations) from the F school effect sizes in each of the three years after program. The estimates in reading are the same as earlier. In math, although the effect sizes fall in Florida, they are still positive

[29] Figlio and Lucas (2004) find that following the first assignment of school grades in Florida, the better students differentially selected into schools receiving grades of "A", though this differential sorting tapered off over time.

and considerably larger than those in Milwaukee.[30] Once again, the Florida effects are statistically different from the corresponding Milwaukee effects in the first and third years after program in both subject areas. These results provide strong evidence in favor of both the theoretical predictions. It should be noted that since none of the F schools got an "F" in either 2000 or 2001, the mean reversion corrected effect sizes attributed to the Florida program in the second and third years may be underestimates.[31]

The Milwaukee program saw a major shift and entered into its second phase when following a 1998 Wisconsin Supreme Court ruling, the religious schools were allowed to accept choice students for the first time in the 1998-99 school year. (I will refer to the post-shift period as second phase Milwaukee or Milwaukee phase II.) This led to a massive increase in the number of MPCP schools and students and the MPS membership fell for the first time. It is tempting to compare the treatment effect in Florida with that in Milwaukee phase II also. However, it is not clear whether this comparison is legitimate. Except for the key differences outlined above, the other features of the two programs were very similar and comparable between Florida and Milwaukee phase I (as described in the introduction). However this was not so in phase II. Due to some funding changes, the voucher amount ($5,220 on average) as well as the revenue loss per student per year was much higher in Milwaukee Phase II than in either Florida or Milwaukee Phase I. Moreover, in Florida we observe the effect of the program only in its first three years while in Milwaukee Phase II we observe the program 9-12 years after it was first implemented. It is reasonable to expect that adjustments and/or effects of adjustments take time to get reflected in test

[30] I also do a pair-wise non parametric test (sign test), where I ignore the significance of coefficients and consider only their signs. Under the null of equal effects the probability that any one effect size in Florida exceeds the corresponding one in Milwaukee is $\frac{1}{2}$. Under the null, $D =$Florida effect-Milwaukee effect follows a binomial distribution. D is positive in all cases. The probability of getting all positive D under the null is very small and hence the null of equal effects can be comfortably rejected.

[31] It might be worth comparing these effect sizes with those obtained in the literature. Greene (2001), and Figlio and Rouse (2006) study the effect of the 1999 Florida program on F schools in the first year after program. Greene (2001) finds F-school effects (relative to C schools) to be 0.42 standard deviations in reading and 0.67 standard deviations in math, while Figlio and Rouse (2006) finds 0.09 of a standard deviation improvement in reading and 0.23 of a standard deviation improvement in math for the F schools. These effects are not directly comparable to those in the current study due to the differences in methodology, specifications and samples. For example, the estimates in Greene (2001) do not control for mean reversion or pre-program trends or changes in demographic compositions of schools. The effects in Figlio and Rouse (2006) are relative to all higher-graded schools (so their comparison group was different), do not control for pre-existing trends or mean reversion, and also pertain to data on the group of Florida districts for which they had micro data. Note that the main focus of Figlio and Rouse (2006) was looking at low stakes tests (not high stakes tests, so the estimates above are their initial estimates). Chiang (2009) and Rouse et al. (forthcoming) look at the second phase of the Florida program after the 2002 changes. Thus they look at a different time period and the program parameters had also changed, and their methodology and samples were also different, so their effect sizes are not directly comparable to those in this paper. Using a regression discontinuity estimation strategy with a 28-point bandwidth, Chiang finds effect sizes of 0.11 in reading and 0.12 in math one year after the program shift. Rouse et al. (forthcoming) finds 0.6-0.14 standard deviation effect sizes in math and 0.6-0.10 standard deviations effect in reading for F schools (relative to the sample of all higher-graded schools) also one year after the program shift.

scores. Since each of these would indicate a higher response in Milwaukee Phase II, it is not clear that the effect of the Florida program will still be higher than that in Milwaukee Phase II. In spite of these problems, I compare the treatment effect in Florida with that in Milwaukee Phase II, but the results should be interpreted with the above caveats in mind.

The first four columns of Table 6 present estimates before correcting for mean reversion, while the last four columns present mean-reversion corrected estimates. All figures are in terms of sample standard deviations. The Florida effects are the same as earlier. The Milwaukee estimates correspond to non-linear regressions corresponding to model (2) (for the actual regression results, see Chakrabarti (2008b)) run on the WKCE reading and math test scores (1997-2002) using the 66-47 sample. While interpreting these results, it should be remembered that the caveats mentioned earlier are likely to bias the Milwaukee phase II effects upwards. Inspite of that, the Florida effects for each of the years and each of the subject areas, and both before and after mean reversion correction are economically larger (and also statistically so in some cases) than the corresponding Milwaukee estimates except second year reading (which are the same).

Consistent with the above findings, there is considerable anecdotal evidence that suggests that F schools have responded to the program. Escambia county implemented a 210-day extended school year in its F schools (typical duration was 180 days), implemented an extended school day at least twice a week and added small group tutoring in afternoons and Saturdays. Palm Beach County targeted its fourth grade teachers for coaching and began more frequent and closer observations of teachers in its F schools. (For more evidence, see Innerst, 2000.) In the words of Carmen Varela-Russo, associate superintendent of technology, strategic planning and accountability, Broward County Public Schools, "People get lulled into complacency". . . "the possibility of losing children to private schools or other districts was a strong message to the whole community."

7 Conclusion

This paper examines the role of vouchers as instruments of public school reform. It argues that voucher design matters,—differences in voucher designs affect public school incentives differently and hence induce different responses from them. Therefore, understanding the effects of different voucher designs is essential to the formulation of effective voucher policies. This study contributes in this direction by comparing the effects of two U.S. voucher programs—that characterize the two different voucher designs implemented in the U.S. so far—on public school incentives and performance. The Florida program was an accountability tied voucher system that incorporated both "threat of vouchers" and stigma. Schools receiving the first "F" grade faced stigma as well as were directly threatened by vouchers. But vouchers were introduced *only* if they failed to meet a certain government designated quality cutoff. The Milwaukee program, on the other hand, was a "voucher shock" program with a sudden government announcement that all low income public school students would be eligible for vouchers. In the context of an equilibrium theory of public school and household behavior, this paper argues that the Florida-type program should bring about an improvement in performance of the threatened public schools and this improvement should exceed the improvement (if any) of the corresponding more treated schools in the Milwaukee-type program. Using data from Florida and Milwaukee, and a difference-in-differences estimation strategy in trends, it then finds robust evidence in favor of the theoretical predictions. These findings are robust to alternative specifications, samples and potentially confounding explanations, continue to hold after adjusting for mean-reversion, and survive a regression discontinuity analysis.

The main contribution of this paper is that it is the first to show that differences in voucher designs—specifically, the two alternative voucher designs implemented in the United States—can have very different incentive and performance effects on affected public schools,—and it does so theoretically as well as empirically. The findings have important policy implications which are all the more relevant in the context of the present concern over public school performance. They suggest that if the objective of the policy maker is to improve the quality of public schools, then the Florida-type "threat of voucher" program rather than the Milwaukee-type "voucher shock" program promises to be more effective.

Appendix A: Proofs of results

Claim 1: *Equilibrium number of public school students falls with vouchers and increases with effort.*

Proof. Step 1: *Equilibrium peer-group quality falls with vouchers and increases with effort.*

Effect of an increase in e:

$$\frac{\delta b^*}{\delta e} = \frac{\frac{\delta g(b^*,.)}{\delta e}}{1-\frac{\delta g(b^*,.)}{\delta b}} \text{ where } \frac{\delta g(b^*,.)}{\delta e} = \frac{1}{N(b^*,.)}.\int_0^1 (\hat{\alpha}(y,b^*,.)-b^*).\frac{\delta\hat{\alpha}}{\delta e}dy$$

The denominator is positive from uniqueness. Consider

$\int_0^1 (\hat{\alpha}(.)-b)dy = \int_0^1 [\hat{\alpha}(.) - \frac{1}{N(.)}.\int_0^{\hat{\alpha}(.)} \alpha d\alpha]dy$ (A.1).

For any y, $[\hat{\alpha}(.).N(.) - \int_0^{\hat{\alpha}(.)} \alpha d\alpha] = \hat{\alpha}(.)[\int_0^1 \hat{\alpha}(.)dy - \frac{\hat{\alpha}(.)}{2}]$, which is positive. Therefore, A.1>0.

It can be checked that $\frac{\delta^2 \hat{\alpha}}{\delta e \delta y} < 0$. $\hat{\alpha}(y)$ is inversely related to y. If for some large y, $[\hat{\alpha}(y, b^*, .) - b^*] < 0$, then there must exist some $y = y_1$, $y_1 \in (0, 1)$ such that $\hat{\alpha}(y_1, b^*, .) = b^*$. Then,

$$\int_0^{y_1} (\hat{\alpha}(y, b^*, .) - b^*) dy > | \int_{y_1}^1 (\hat{\alpha}(y, b^*, .) - b^*) dy|$$

$$\Rightarrow \int_0^{y_1} (\hat{\alpha}(y, b^*, .) - b^*). \frac{\delta \hat{\alpha}(y_1, .)}{\delta e} dy > | \int_{y_1}^1 (\hat{\alpha}(y, b^*, .) - b^*). \frac{\delta \hat{\alpha}(y_1, .)}{\delta e} dy|$$

$$\Rightarrow \int_0^{y_1} (\hat{\alpha}(y, b^*, .) - b^*). \frac{\delta \hat{\alpha}(y, .)}{\delta e} dy > | \int_{y_1}^1 (\hat{\alpha}(y, b^*, .) - b^*). \frac{\delta \hat{\alpha}(y, .)}{\delta e} dy| \Rightarrow \frac{\delta b^*}{\delta e} > 0$$

Effect of an increase in v:

$$\frac{\delta b^*}{\delta v} = \frac{\frac{\delta g(b^*, .)}{\delta v}}{1 - \frac{\delta g(b^*, .)}{\delta v}} \text{ where} \frac{\delta g(b^*, .)}{\delta v} = \frac{1}{N(b^*, .)}. \int_0^1 (\hat{\alpha}(y, b^*, .) - b^*). \frac{\delta \hat{\alpha}}{\delta v} dy$$

The denominator is positive from uniqueness. Since $A.1 > 0$, $\frac{\delta \hat{\alpha}}{\delta v} < 0$, $\frac{\delta^2 \hat{\alpha}}{\delta v \delta y} > 0$ and $\hat{\alpha}(y)$ is inversely related to y, the numerator is negative. Therefore, $\frac{\delta b^*}{\delta v} < 0$.

Step 2: *Equilibrium cutoff ability at each income level falls with vouchers and increases with effort.* Follows from $\frac{\delta \hat{\alpha}(y; b^*, .)}{\delta e} = \frac{\delta \hat{\alpha}(y; b^*, .)}{\delta e} | b^* + \frac{\delta \hat{\alpha}(y; b^*, .)}{\delta b}. \frac{\delta b^*}{\delta e}$ and $\frac{\delta \hat{\alpha}(y; b^*, .)}{\delta v} = \frac{\delta \hat{\alpha}(y; b^*, .)}{\delta v} | b^* + \frac{\delta \hat{\alpha}(y; b^*, .)}{\delta b}. \frac{\delta b^*}{\delta v}$ and step 1. From step 2 and definition of $N(b^*, .)$, the proof follows.

Proof of Proposition 1. Under the VS program, e_{VS} solves the first order condition:
$\frac{\delta R(e,v)}{\delta e} = (p - c_N) N_e(e, v) - C_e(e) = 0$.

Comparative statics with respect to v yields:

$$\frac{\delta e}{\delta v} = \frac{-[(p - c_N) N_{ev} - c_{NN} N_v N_e]}{(p - c_N) N_{ee} - c_{NN} N_e^2 - C_{ee}} \qquad A.2$$

The denominator is negative from the strict concavity of the rent function. Also $p - c_N > 0$ and $c_{NN} N_v N_e < 0$. $N_{ev} = \int_0^1 \left[\frac{\delta^2 \hat{\alpha}(y, b^*, .)}{\delta e \delta v} + \frac{\delta^2 \hat{\alpha}(y, b^*, .)}{\delta e \delta b} \frac{\delta b^*}{\delta v} \right] dy$. $\frac{\delta b^*}{\delta v} < 0$ from claim 1. It can be seen that $\frac{\delta^2 \hat{\alpha}(y, b^*, .)}{\delta e \delta b} < 0$ and $\frac{\delta^2 \hat{\alpha}(y, b^*, .)}{\delta e \delta v} < 0$. Therefore $N_{ev} \gtrless 0$ which implies that A.2 $\gtrless 0$. ■

Proof of Proposition 2.

Proof of part (i): $pN(e_{VS}, 0) - c(N(e_{VS}, 0))) - C(e_{VS}) > pN(e_{VS}, v) - c(N(e_{VS}, v)) - C(e_{VS})$, since vouchers decrease rent. By the strict concavity of the rent function, $\exists\ \bar{e} > e_{VS}$ that satisfies the public school's incentive constraint under TOV with equality
$pN(\bar{e}, 0) - c(N(\bar{e}, 0)) - C(\bar{e}) = pN(e_{VS}, v) - c(N(e_{VS}, v)) - C(e_{VS})$.

Proof of part(ii): $pN(e_{PP}, 0) - c(N(e_{PP}, 0))) - C(e_{PP}) > pN(e_{VS}, 0) - c(N(e_{VS}, 0))) - C(e_{VS}) > pN(e_{VS}, v) - c(N(e_{VS}, v)) - C(e_{VS})$. The first inequality follows because e_{PP} is the rent maximizing effort under $v = 0$. Given strict concavity of the rent function, $\exists\ \bar{e} > e^{PP}$ that satisfies the public school's incentive constraint under TOV with equality. ■

Appendix B: Moral hazard problem – unobservable public school effort

This appendix relaxes the assumption of complete observability of public school effort and examines whether under unobservable public school effort, the equilibrium effort under the TOV program still exceeds those under the PP and the VS programs. Given public school effort $e \in [e_{min}, e_{max}]$, "effective effort" e' is realized according to the distribution $F(e'/e)$, where $e' \in [e'_{min}, e'_{max}]$. Although e is not publicly observable, all agents have complete knowledge of the set $[e_{min}, e_{max}]$ and the family of conditional distributions $F(e'/e)$ for $e \in [e_{min}, e_{max}]$. The corresponding density $f(e'/e)$ satisfies the strict monotone likelihood ratio property (MLRP). $F(e'/e)$ satisfies the convexity of the distribution function condition (CDFC) i.e $F_{ee}(e'/e) > 0$ for all $e' \in [e'_{min}, e'_{max}]$ and $e \in [e_{min}, e_{max}]$. Public school quality ($q = q(e', b)$) is a composite of two factors: (i) "effective effort" e' and (ii) peer group quality (b) and can be thought of as being embodied in school scores. All agents observe quality q but not the actual public school effort e that generated it. The uncertainty signifies the absence of any direct one-to-one relationship between the effort of teachers and administrators, and school scores.

The simple public-private system, the Milwaukee-type VS program and the Florida-type TOV program are modeled as follows. The public-private system has two stages. In the first stage, the public school chooses effort which is not observable by the other agents of the economy. Quality q is realized and observed by all agents in the economy. In stage 2, households choose between schools. The VS program consists of three stages: In the first stage, the Government announces the voucher v. In stage 2, facing v, the public school chooses effort which is not observable. Quality is realized and observed by all agents in the economy. In stage 3, households choose between schools. The Florida program is modeled in four stages: In the first stage, the Government announces the program and a cutoff quality level $\bar{q}$ and voucher v. In stage 2, the public school chooses effort. Quality is realized and observed by all agents in the economy. In stage 3, government imposes vouchers if $q < \bar{q}$. No voucher is imposed if $q \geq \bar{q}$. In the final stage, households choose between schools.

Household behavior is basically the same as earlier, the only difference is that instead of using effort itself, they use a noisy representation of effort, effective effort e' to make their school choices.

The public school anticipates household behavior and chooses e to maximize expected rent: $ER(v,.) = \int_{e'_{min}}^{e'_{max}} [pN(e',v,b^*,.) - c(N(e',v,b^*,.))] f(e'/e)de' - c_1 - C(e)$ where $v = 0$ under the public-private system. The expected rent function is strictly concave under CDFC.

Equilibrium public school effort under the VS program can be either greater or less than the PP system. (Proof available on request.) The intuition behind this is as follows. With imposition of vouchers, rent falls at each realization of e'. An increase in e increases the probability of higher e' realizations. However, the above fall in rent can either increase or decrease in e'. This implies that vouchers may induce public schools to correspondingly decrease or increase effort in response to vouchers.

Under the Florida TOV program the public school faces a quality cutoff $\bar{q}$ or equivalently an "effective effort" cutoff $\bar{e'}$ and chooses e to maximize its expected rent. The school's expected rent under the TOV program is given by the following. Under CDFC, H is strictly concave in e.
$H = \int_{e'_{min}}^{\bar{e'}} [pN(e',v,.) - c(N(e',v,.))] f(e'/e)de' + \int_{\bar{e'}}^{e'_{max}} [pN(e',0,.) - c(N(e',0,.)] f(e'/e)de' - c_1 - C(e)$.

Proposition 3 *(i)There exists e'_1, $e'_{min} \leq e'_1 < E_1$ such that if the cutoff $\bar{e'} \in [e'_1, e'_{max}]$ the effort under the "threat of voucher" program unambiguously exceeds that under the "voucher shock" program i.e., $e_{TOV} > e_{VS}$. (ii) There exists e'_2, $E_2 < e'_2 \leq e'_{max}$ such that if the cutoff $\bar{e'} \in [e'_{min}, e'_2]$ the effort under the "threat of voucher" program unambiguously exceeds that under the pre-program public-private equilibrium i.e., $e_{TOV} > e_{PP}$.*

The intuitive argument behind this proposition is as follows. First consider the TOV and the VS programs. Facing the TOV program, if the school chooses e_{VS} (the equilibrium effort under the VS program), then at each realization of $e' < \bar{e'}$, its rent is the same as in the VS program. On the other hand, for each realization of $e' \geq \bar{e'}$, its rent is higher. Therefore the school chooses an effort strictly higher than e_{VS} to increase its probability of falling above $\bar{e'}$ since it follows from the MLRP that an increase in effort increases the probability of higher e's. The intuition behind $\bar{e} > e_{PP}$ is similar. Choosing e_{PP} under the TOV gives it the same rent as the PP program at each realization above $\bar{e'}$ but lower rent at each realization below $\bar{e'}$. The school chooses an effort strictly above e_{PP} to increase (decrease) the probability of realizations above (below) $\bar{e'}$.[32] Thus, the results here parallel those in the

[32] However, although rent falls at each realization of e' with vouchers, this fall (or alternatively, the gain in rent from avoiding vouchers) may either increase or decrease with e'. Depending on this, under certain circumstances as proposition 3 indicates, at very low levels of cutoff, the public school effort under TOV may be less than VS and at very high levels of cutoff, effort under TOV may be less than PP. The intuition is as follows. First consider TOV versus VS. If $\bar{e'}$ is low, schools escape vouchers for low values of e' also. If it is the case that the gain in rent from avoiding vouchers is largest for lower values of e' then since an increase in effort decreases the probability of occurrence of lower values of e', public school

complete information model (Proposition 2).

Appendix C: Proof of Result in Appendix B

Proof of Proposition 3.

Proof of part (i): Evaluating the first order condition under the TOV program at e_{VS}:

$$\frac{\delta H}{\delta e}|_{e_{VS}} = \frac{\delta H}{\delta e}|_{e_{VS}} - \frac{\delta ER(v,.)}{\delta e}|_{e_{VS}} = \int_{\bar{e'}}^{e'_{max}} [r(e',0) - r(e',v)] f_e(e'/e_{VS}) de' = \int_{\bar{e'}}^{e'_{max}} \beta(e',v) f_e(e'/e_{VS}) de'$$

where $\beta(e',v) = [r(e',0) - r(e',v)]$ and $r(e',V) = pN(e',V) - c(N(e',V))$, $V = \{0, v\}$. MLRP implies that there exists E_1, $f_e(e'/e_{VS}) \lesseqqgtr 0$ according as $e' \lesseqqgtr E_1$. Now if the cutoff $\bar{e'} \geq E_1$ then $\frac{\delta H}{\delta e}|_{e_{VS}} > 0$ since $\beta(e',v) > 0$ so that $\bar{e} > e_{VS}$. There are two cases if $\bar{e'} < E_1$. Let $\int_{e'_{min}}^{e'_{max}} \beta(e',v) f_e(e'/e_{VS}) de' = A_1$. Although $f_e(e'/e_{VS}) \lesseqqgtr 0$ according as $e' \lesseqqgtr E_1$, $\beta(e',v)$ may be increasing or decreasing in e'. Therefore $A_1 \gtrless 0$. (Note that $A_1 \lessgtr 0$ implies $e_{PP} \lessgtr e_{VS}$).

Case 1: If $A_1 > 0$ then for any $\bar{e'} \in (e'_{min}, E_1)$, $\frac{\delta H}{\delta e}|_{e_{VS}} > 0$ and $e_{TOV} > e_{VS}$.

Case 2: If $A_1 < 0$ then $\exists e'^1 \in (e'_{min}, E_1)$ such that $|\int_{e'^1}^{E_1} \beta(e',v) f_e(e'/e_{VS}) de'| = \int_{E_1}^{e'_{max}} \beta(e',v) f_e(e'/e_{VS}) de'$ then for any $\bar{e'} \in (e'^1, E_1)$, $\frac{\delta H}{\delta e}|_{e_{VS}} > 0$ and $e_{TOV} > e_{VS}$.

Using cases (1) and (2) define $e'_1 = [min\{e' \in [e'_{min}, E_1] : \int_{e'}^{e'_{max}} \beta(e',v) f_e(e'/e_{VS}) de' > 0]$. Then for any $\bar{e'} \in [e'_1, e'_{max},]$, $e_{TOV} > e_{VS}$. Note that $e'_1 \geq e'_{min}$ according as $A_1 \lessgtr 0$.

Proof of part (ii): Evaluating the first order condition under the TOV program at e_{PP}:

$$\frac{\delta H}{\delta e}|e_{PP} = \frac{\delta H}{\delta e}|e_{PP} - \frac{\delta ER(0,.)}{\delta e}|e_{PP} = \int_{e'_{min}}^{\bar{e'}} [r(e',v) - r(e',0)] f_e(e'/e_{PP}) de' = -\int_{e'_{min}}^{\bar{e'}} \beta(e',v) f_e(e'/e_{PP}) de'$$

MLRP implies that there exists E_2, $f_e(e'/e_{PP}) \lesseqqgtr 0$ according as $e' \lesseqqgtr E_2$. Now if the cutoff $\bar{e'} \leq E_2$ then $\frac{\delta H}{\delta e}|e_{PP} > 0$ so that $e_{TOV} > e_{PP}$. There are two cases if $\bar{e'} > E_2$. Let $A_2 = \int_{e'_{min}}^{e'_{max}} \beta(e',v) f_e(e'/e_{PP}) de'$. Again similarly as above $A_2 \gtrless 0$. (Note that $A_2 \gtrless 0$ implies $e_{PP} \gtrless e_{VS}$).

Case 1: If $A_2 < 0$ then for any $\bar{e'} \in (E_2, e'_{max})$ $\frac{\delta H}{\delta e}|e_{PP} > 0$ and $e_{TOV} > e_{PP}$.

Case 2: If $A_2 > 0$ then $\exists e'^2 \in (E_2, e'_{max})$ such that $|\int_{e'_{min}}^{E_2} \beta(e',v) f_e(e'/e_{PP}) de'| = \int_{E_2}^{e'^2} \beta(e',v) f_e(e'/e_{PP}) de'$ then for any $\bar{e'} \in (E_2, e'^2)$, $\frac{\delta H}{\delta e}|e_{PP} > 0$ and $e_{TOV} > e_{PP}$.

may find it profitable not to increase effort. Now consider TOV versus PP. If $\bar{e'}$ is high, vouchers will be incurred at high values of e' also. If it is the case that the fall in rent due to vouchers is highest for high values of e', then the school may not have an incentive to increase effort since an increase in effort increases the probability of occurrence of higher e'.

Using cases (1) and (2) define $e'_2 = [max\{e' \in [E_2, e'_{max}] : \int_{e'_{min}}^{e'} -\beta(e', v) f_e(e'/e_{PP}) de' > 0]$. Then for any $\bar{e'} \in [e'_{min}, e'_2]$, $e_{TOV} > e_{PP}$. Note that $e'_2 \leq e'_{max}$ according as $A_2 \gtreqless 0$. ■

References

Chakrabarti, Rajashri (2008a),"Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs," Federal Reserve Bank of New York Staff Paper Number 315.

Chakrabarti, Rajashri (2008b),"Can Increasing Private School Participation and Monetary Loss in a Voucher Program Affect Public School Performance? Evidence from Milwaukee," Journal of Public Economics volume 92, Number 5-6, 1371-1393.

Chakrabarti, Rajashri (2009),"Do Vouchers Lead to Sorting Even Under Random Private School Selection? Evidence from Milwaukee ," Federal Reserve Bank of New York Staff Paper Number 379.

Chakrabarti, Rajashri (Forthcoming),"Vouchers, Public School Response and the Role of Incentives: Evidence from Florida," Economic Inquiry.

Chiang, Hanley (2009),"How accountability pressure on failing schools affects student achievement," volume 93, Number 9-10, 1045-1057.

Epple, Dennis and Richard Romano (1998), "Competition between Private and Public Schools, Vouchers and Peer Group Effects," American Economic Review 62(1), 33-62.

Epple, Dennis and Richard Romano (2008), "Educational Vouchers and Cream Skimming" International Economic Review 49 (4), 1395-1435.

Ferreyra, Marta M. and Pierre J. Liang (2012), "Information Asymmetry and Equilibrium Monitoring in Education," Journal of Public Economics 96, 237-254.

Figlio, David and Maurice Lucas (2004), "What's in a Grade? School Report Cards and the Housing Market", American Economic Review, 94 (3), 591-604.

Figlio, David and Cecilia Rouse (2006), "Do Accountability and Voucher Threats Improve Low-Performing Schools?" Journal of Public Economics, 90(1-2), 239-255.

Greene, Jay (2001), "An Evaluation of the Florida A-Plus Accountability and School Choice Program," New York: Manhattan Institute for Policy Research.

Greene, Jay P. and Marcus A. Winters (2003), "When Schools Compete: The Effects of Vouchers on Florida Public School Achievement," Education Working Paper 2.

Hoxby, Caroline (2003a), "School Choice and School Productivity (Or, Could School Choice be the tide that lifts all boats?)", in C. Hoxby (ed.), The Economics of School Choice, University of Chicago Press, Chicago.

Hoxby, Caroline (2003b), "School Choice and School Competition: Evidence from the United States", Swedish Economic Policy Review, 10, 11-67.

Innerst, Carol (2000), "Competing to win: How Florida's A+ Plan has Triggered Public School Reform," The Center for education Reform.

Kane, T. J., and D. O. Staiger(2002b). Volatility in test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.), Brookings Papers on Education Policy (pp. 235-283). Washington,DC: Brookings Institution Press.

McMillan, Robert (2004),"Competition, Incentives, and Public School Productivity," Journal of Public Economics, 88, 1871-1892.

Manski, Charles F. (1992), "Educational Choice (Vouchers) and Social Mobility," Economics of Education Review, 11(4), 351-369.

Neal, Derek (2009), "Designing Incentive Systems for Schools," in Springer, M. (ed.) Performance Incentives: Their Growing Impact on American K-12 Education. Brookings Institution Press. Washington D.C.

Nechyba, Thomas J. (1996), "Public School Finance in a General Equilibrium Tiebout World: Equalization Programs, Peer Effects and Private School Vouchers," NBER working paper # 5642.

Nechyba, Thomas J. (1999), "School Finance Induced Migration Patterns: The case of Private School Vouchers," Journal of Public Economic Theory 1 (1), 5-50.

Nechyba, Thomas J. (2000), "Mobility, Targeting, and Private-School Vouchers", American Economic Review 90(1), 130-146.

Nechyba, Thomas J. (2003), "Introducing School Choice into Multi-District Public School Systems", in C. Hoxby (ed.), The Economics of School Choice, University of Chicago Press, Chicago.

Rouse, Cecilia, Jane Hannaway, Daniel Goldhaber, and David Figlio (Forthcoming), "Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure," American Economic Journal: Economic Policy.

West, Martin and Paul Peterson (2006), "The Efficacy of Choice Threats Within Accountability Systems: Results from Legislatively Induced Experiments." The Economic Journal, Vol. 116, C46-C62.

Table 1: Effect of "Threatened Status" on FCAT Reading (1998-2002), Math (1998-2002) and Writing (1994-2002) Scores

(Sample of treated F and control D schools in Florida)

	Grade 4 Reading				Grade 5 Math		Grade 4 Writing	
	OLS	FE	OLS	FE	FE	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Trend	-0.59	-0.20			12.80***		0.21***	
	(0.57)	(0.67)			(0.76)		(0.004)	
Program dummy	-5.25***	-5.30***			0.31		0.10***	
	(1.46)	(0.84)			(0.91)		(0.02)	
Program dummy * trend	5.35***	5.57***			-9.11***		-0.13***	
	(0.57)	(0.76)			(0.85)		(0.01)	
Treated * Program dummy	2.71*	2.97*			7.90**		0.30***	
	(1.60)	(1.78)			(2.29)		(0.05)	
Treated * Program dummy * trend	1.57**	1.10			-0.71		0.04**	
	(0.74)	(1.00)			(1.04)		(0.02)	
Treated * 1 year after program			4.85***	4.85***		6.78***		0.35***
			(1.49)	(1.68)		(1.63)		(0.04)
Treated * 2 years after program			4.71***	3.30*		7.25***		0.37***
			(1.78)	(1.71)		(1.82)		(0.04)
Treated * 3 years after program			8.01***	7.08***		5.35***		0.43***
			(1.49)	(1.78)		(2.00)		(0.05)
Year dummies	N	N	Y	Y	N	Y	N	Y
Controls	N	Y	N	Y	Y	Y	Y	Y
Observations	2567	2550	2567	2550	2524	2524	4476	4476
R-squared	0.11	0.77	0.11	0.77	0.76	0.76	0.84	0.85
p-value[1]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

*, **, ***: significant at the 10, 5, and 1 percent level respectively. [1]p-value of F-test of the program effect on treated schools. Robust standard errors adjusted for clustering by school district are in parentheses. All regressions are weighted by the number of students tested. Columns (1)-(2), (5), (8) report results from model 1, columns (6), (9) from model 2 and columns (3)-(4), (7), (10) from model 3. The OLS columns include an F dummy. Columns (8)-(10) include an interaction term of treated dummy with trend. Controls include race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Table 2: Effect of "Threatened Status" on FCAT Reading (1998-2002), Math (1998-2002) and Writing (1994-2002) Scores

(Sample of treated F, D and control C schools in Florida)

	Reading				Math				Writing			
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Less treated * program	1.70^{**}	1.20			-0.23	-0.15			0.08^{***}	0.08^{***}		
	(0.84)	(1.00)			(0.92)	(1.09)			(0.02)	(0.02)		
More treated * program	4.88^{**}	4.99^{**}			$9.05^{***\dagger\dagger}$	$9.02^{***\dagger\dagger}$			$0.38^{***\dagger\dagger}$	$0.37^{***\dagger\dagger}$		
	(2.04)	(2.43)			(2.13)	(2.23)			(0.09)	(0.05)		
Less treated * program	5.22^{***}	4.85^{***}			0.29	0.30			-0.03	-0.03		
*trend	(0.81)	(0.93)			(0.60)	(1.04)			(0.02)	(0.02)		
More treated * program	7.73^{***}	8.02^{***}			0.91	0.54			$0.01^{\dagger\dagger}$	0.00		
*trend	(2.47)	(2.08)			(1.10)	(1.01)			(0.02)	(0.02)		
Less treated * 1 year after			4.13^{***}	3.53^{***}			1.06^{***}	0.97			0.05^{**}	0.05^{**}
			(0.75)	(0.76)			(0.73)	(0.85)			(0.02)	(0.02)
Less treated * 2 years after			5.84^{***}	5.52^{***}			2.83^{**}	2.54^{***}			0.00	0.00
			(0.96)	(0.80)			(1.44)	(0.94)			(0.03)	(0.02)
Less treated * 3 years after			8.60^{***}	7.94^{***}			3.89^{***}	3.47^{***}			-0.03	-0.03
			(0.93)	(0.87)			(1.20)	(0.92)			(0.04)	(0.02)
More treated * 1 year after			$9.45^{***\dagger\dagger}$	$9.32^{***\dagger\dagger}$			$9.56^{***\dagger\dagger}$	$8.96^{***\dagger\dagger}$			$0.40^{***\dagger\dagger}$	$0.39^{***\dagger\dagger}$
			(1.92)	(1.87)			(2.02)	(1.59)			(0.07)	(0.04)
More treated * 2 years after			$11.01^{***\dagger}$	$10.75^{***\dagger}$			$11.61^{***\dagger\dagger\dagger}$	$11.00^{***\dagger\dagger}$			$0.39^{***\dagger}$	$0.37^{***\dagger}$
			(2.39)	(1.87)			(2.99)	(1.77)			(0.07)	(0.04)
More treated * 3 years after			$17.08^{***\dagger\dagger}$	$16.03^{***\dagger\dagger}$			$11.39^{***\dagger\dagger}$	$11.94^{***\dagger\dagger}$			$0.42^{***\dagger}$	$0.39^{***\dagger}$
			(2.64)	(1.91)			(3.46)	(1.95)			(0.05)	(0.05)
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Observations	6034	5933	6034	5933	6003	5909	6003	5909	10646	10587	10646	10587
R-squared	0.44	0.86	0.44	0.86	0.44	0.83	0.44	0.83	0.72	0.85	0.74	0.86
p-value[1]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

*, **, ***: significant at the 10, 5, and 1 percent level respectively. †,††: more treated significantly different from less treated at 5 and 1 percent level respectively. [1]p-value of the F-test of program effect on more treated schools. Robust standard errors adjusted for clustering by school district are in parentheses. The OLS columns include more treated and less treated dummies. All regressions are weighted by the number of students tested. Columns (1)-(2), (5)-(6), (9)-(10) include program dummy, trend and an interaction of trend with program dummy while columns (3)-(4), (7)-(8), (11)-(12) contain year dummies. Columns (1)-(2), (5)-(6), (9)-(12) include interactions of trend with less treated and more treated dummies respectively and (3)-(4), (7)-(8) include interaction of D_1 dummy ($D_1 = 1$ if year $>$ 1998) with less treated and more treated dummies respectively. Controls include race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Table 3: Mean Reversion of the 98F Schools Compared to 98D and 98C Schools, 1998-1999.

Panel A: 98F and 98D Schools Dependent Variable: FCAT Score, 1998-99.

	Reading		Math		Writing	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	FE	OLS	FE	OLS	FE
trend	2.27***	2.01***	14.25***	14.02***	0.04***	0.04***
	(0.67)	(0.43)	(0.65)	(0.49)	(0.01)	(0.01)
98F*trend	-0.45	-0.65	1.03	1.17	0.14***	0.14***
	(1.77)	(1.14)	(1.81)	(1.19)	(0.03)	(0.02)
Observations	1353	1353	1354	1354	1355	1355
R^2	0.64	0.93	0.63	0.91	0.33	0.85

Panel B: 98F, 98D and 98C Schools Dependent Variable: FCAT Score, 1998-99.

	Reading		Math		Writing	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	FE	OLS	FE	OLS	FE
trend	1.76**	1.76***	9.57***	9.71***	0.03***	0.03***
	(0.56)	(0.35)	(0.50)	(0.36)	(0.01)	(0.01)
98F*trend	0.18	-0.55	4.67***	4.63***	0.14***	0.14***
	(1.78)	(1.12)	(1.80)	(1.16)	(0.03)	(0.02)
98D*trend	0.41	0.16	4.61***	4.22***	0.01	0.01
	(0.88)	(0.54)	(0.82)	(0.58)	(0.02)	(0.01)
Observations	2605	2605	2608	2608	2608	2608
R^2	0.76	0.96	0.76	0.94	0.38	0.87

*, **, ***: significant at the 10, 5, and 1 percent level. All regressions include race, sex, % of students eligible for free or reduced-price lunches and real per pupil expenditure as controls. OLS regressions in Panel A include 98F dummy; OLS regressions in Panel B include 98F and 98D dummies. Sample of 98F and 98D schools: s.d of FCAT reading, math and writing respectively are 18.9, 18.05, 0.30. Sample of 98F, 98D, 98C schools: s.d of FCAT reading, math and writing respectively are 21.16, 21.56 and 0.31.

Table 4: Is there a Stigma Effect of getting the Lowest Performing Grade? Effect of being Categorized in Group 1 on FCAT Writing Scores

	Using FCAT Writing Scores, 1997-1998					
	Sample: Group 1, 2 Schools			Sample: Group 1, 2, 3 Schools		
	OLS	FE	FE	OLS	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Trend	0.52^{***}	0.52^{***}	0.48^{***}	0.48^{***}	0.48^{***}	0.46^{***}
	(0.04)	(0.03)	(0.04)	(0.02)	(0.01)	(0.02)
Group 1 * trend	-0.01	-0.02	-0.02	0.03	0.01	0.02
	(0.08)	(0.06)	(0.06)	(0.07)	(0.05)	(0.05)
Group 2 * trend				0.03	0.04	0.04
				(0.04)	(0.03)	(0.03)
Controls	N	N	Y	N	N	Y
Observations	314	314	314	1361	1361	1358
R-squared	0.49	0.84	0.85	0.52	0.87	0.87

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. Robust standard errors adjusted for clustering by school district are in parentheses. All regressions are weighted by the number of students tested and include race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure as controls. OLS regression in column (1) includes group 1 dummy, OLS regression in column (4) includes group1 and group 2 dummies.

Table 5: Comparing the Impact of Florida and Milwaukee Programs

Using performance in reading test [WRCT (% above) 1989-97 and FCAT Reading 1998-2002] and math test [ITBS Math 1986-1997 and FCAT Math 1998-2002]

						Corrected for Mean Reversion				
	Reading			Math		Reading			Math	
	Wisconsin WRCT (1)	Wisconsin ITBS (2)	Florida FCAT (3)	Wisconsin ITBS (4)	Florida FCAT (5)	Wisconsin WRCT (6)	Wisconsin ITBS (7)	Florida FCAT (8)	Wisconsin ITBS (9)	Florida FCAT (10)
More Treated * 1 year after prog	-0.06††	0.06††	0.47***	-0.24†††	0.45***	-0.06††	0.06††	0.47***	-0.24††	0.24***
More treated * 2 years after prog	0.38*	0.36	0.50***	0.26	0.55***	0.38*	0.36	0.50***	0.26	0.34***
More treated * 3 years after prog	0.35††	0.15††	0.80***	-0.13†††	0.60***	0.35††	0.15††	0.80***	-0.13††	0.39***

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. †, ††, †††: Milwaukee effect significantly different from corresponding Florida effect at 10, 5, and 1 percent level, respectively. All figures are in terms of respective sample standard deviations. All figures are obtained from regressions that contain school fixed effects, year dummies, interactions of year dummies with the respective treatment dummies, race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure. Standard deviation of FCAT reading scores = 20, Standard deviation of FCAT math scores = 20, Standard deviation of WRCT (% above) reading scores = 16, Standard deviation of ITBS reading scores = 18.45, Standard deviation of ITBS math scores = 16.71. For standard deviations corresponding to the mean reversion sample, see footnote for table 4.

Table 6: Comparing the impact of Florida and Milwaukee phase II programs

					Corrected for Mean Reversion			
	Reading		Math		Reading		Math	
	Wisconsin WKCE (1)	Florida FCAT (2)	Wisconsin WKCE (3)	Florida FCAT (4)	Wisconsin WKCE (5)	Florida FCAT (6)	Wisconsin WKCE (7)	Florida FCAT (8)
More Treated * 1 year	$0.20^{\dagger\dagger}$	0.47^{***}	0.27	0.45^{***}	$0.20^{\dagger\dagger}$	0.47^{***}	0.03	0.24^{***}
More treated * 2 years	0.50^{***}	0.50^{***}	0.38^{***}	0.55^{***}	0.50^{***}	0.50^{***}	0.14^{***}	0.34^{***}
More treated * 3 years	$0.53^{***\dagger}$	0.80^{***}	0.57^{***}	0.60^{***}	$0.53^{***\dagger}$	0.80^{***}	0.33^{***}	0.39^{***}

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. †, ††, †††: Milwaukee effect significantly different from corresponding Florida effect at 10, 5, and 1 percent level, respectively. All figures are in terms of respective sample standard deviations. All figures are obtained from regressions that contain school fixed effects, year dummies, interactions of year dummies with the respective treatment dummies, race, sex, free-reduced lunch percentage and real per pupil expenditure. Standard deviation of FCAT reading scores = 20, Standard deviation of FCAT Math Scores = 20, Standard deviation of WKCE reading scores = 13.07, Standard deviation of WKCE math scores = 15.01, Standard Deviation of WKCE Math for the mean reversion sample=14.4, Standard Deviation of FCAT Math for the mean reversion sample= 20.04

Table D.1: Pre-program Characteristics of More Treated and Control Schools, Florida and Wisconsin

Panel A	Florida	Wisconsin	Florida–Wisconsin
More Treated Schools	(std.dev.)	(std. dev.)	[p-value]
% black	62.79	66.55	-3.76
	(28.23)	(32.22)	[0.56]
% hispanic	18.95	18.07	0.88
	(23.40)	(24.54)	[0.87]
% white	17.18	10.21	6.97
	(19.54)	(10.68)	[0.07]
%male	51.38	52.25	-0.87
	(4.84)	(2.60)	[0.34]
% free-reduced lunch	85.80	84.5	1.3
	(9.95)	(6.48)	[0.50]

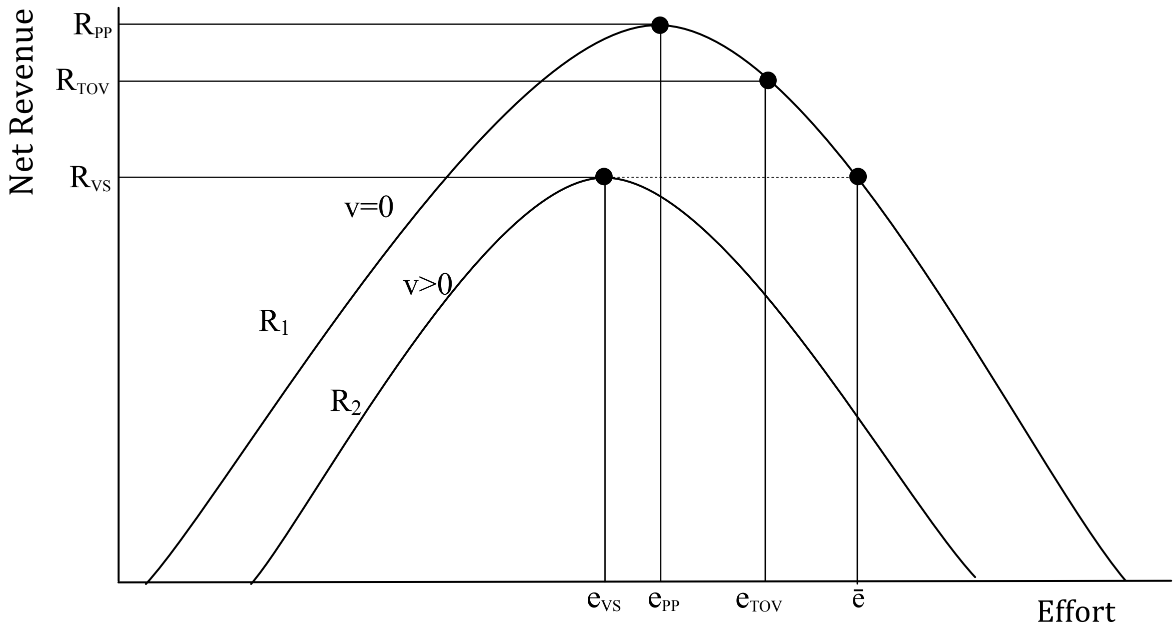
Panel B	Florida	Wisconsin	Florida–Wisconsin
Control Schools	(std.dev.)	(std. dev.)	[p-value]
% black	18.12	22.37	-4.25
	(14.17)	(12.93)	[0.10]
% hispanic	15.49	14.84	0.17
	(21.23)	(6.02)	[0.86]
% white	63.59	60.85	2.73
	(22.33)	(12.80)	[0.49]
% male	51.38	50.63	0.76
	(4.84)	(2.29)	[0.43]
% free-reduced lunch	50.14	44.95	5.19
	(17.51)	(11.66)	[0.10]

The more treated schools in Florida are the F schools and the control schools are the C schools. This table uses the 66-47 sample for the Milwaukee program; control schools constitute Wisconsin schools outside Milwaukee (see section 5.2 for details).

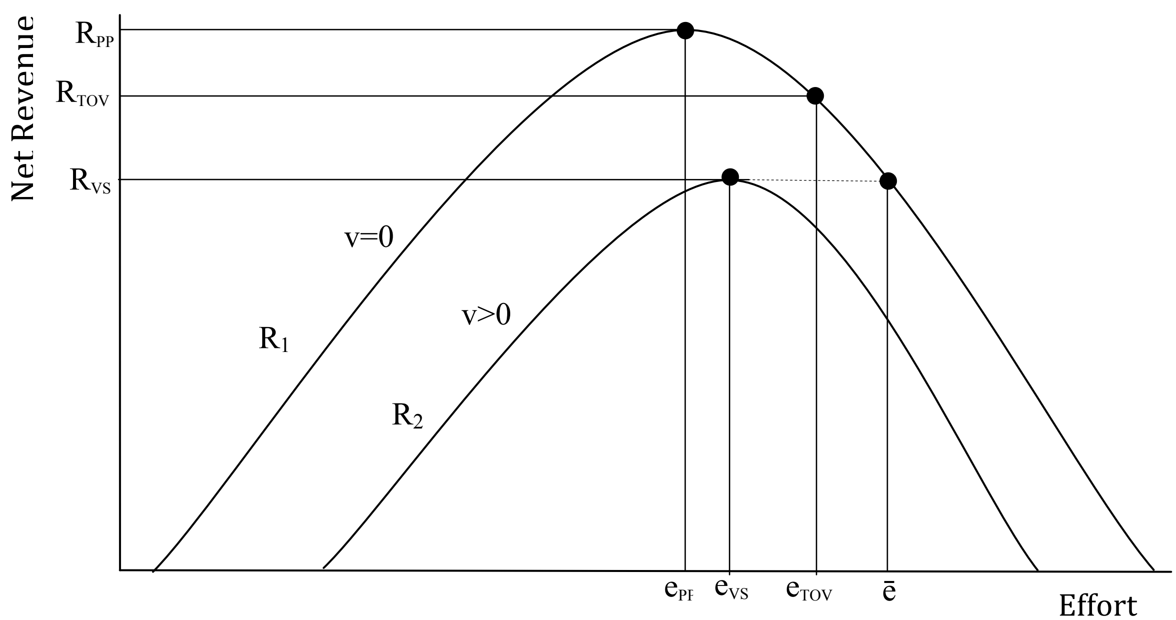
Table D.2: Effect of the Milwaukee Voucher Program on Test Scores by Treatment Status

	Using treatment groups					
	WRCT (% above)		ITBS Reading		ITBS Math	
	(1)	(2)	(3)	(4)	(5)	(6)
Somewhat treated * program dummy	3.50		3.21		0.39	
	(2.59)		(5.45)		(2.81)	
More treated * program dummy	2.85		3.40		-2.97	
	(3.32)		(5.79)		(3.13)	
Somewhat treated * program dummy *trend	0.64		1.22		0.61	
	(0.47)		(2.02)		(0.54)	
More treated * program dummy * trend	0.67		3.40		0.75	
	(0.62)		(5.79)		(0.63)	
Somewhat treated * 1 year after program		2.03		4.15		-1.35
		(2.81)		(4.49)		(2.94)
Somewhat treated * 2 years after program		5.38**		7.83		6.14*
		(2.43)		(5.17)		(3.38)
Somewhat treated * 3 years after program		5.01		6.78		2.47
		(3.03)		(5.31)		(3.31)
More treated * 1 year after program		-0.92		1.12		-4.02
		(3.33)		(3.86)		(3.26)
More treated * 2 years after program		6.06*		6.59		4.36
		(3.14)		(5.15)		(3.83)
More treated * 3 years after program		5.69		2.85		-2.22
		(3.16)		(5.18)		(3.54)
Year dummies	N	Y	N	Y	N	Y
Observations	1195	1195	717	717	1127	1127
R-squared	0.50	0.58	0.55	0.55	0.58	0.60
p-value[1]	0.06	0.02	0.68	0.62	0.49	0.28

*, **, ***: significant at the 10, 5, and 1 percent level respectively. [1]p-value of the F-test of joint significance of more treated shift coefficients. Robust standard errors adjusted for clustering by school district are in parentheses. This table uses the 66-47 sample. For each subject, the first column presents results corresponding to model 1, the second column results corresponding to model 3. All regressions include school fixed effects and control for race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure. Odd-numbered columns include program dummy, a time trend and an interaction of trend with program dummy, while even-numbered columns include year dummies.



Panel A



Panel B

Figure 1. Analyzing the Relationship between "Public-Private", "Voucher Shock" and "Threat of Voucher" Equilibria