What Good is a Good Fit? Religious Matching and Educational Outcomes

CATHERINE RUTH PAKALUK

The Catholic University of America

Abstract: This paper introduces the notion of "fit" between students and schools. Fit is a student-school pairing with respect to intangible characteristics like culture, religion, or educational focus. Using NLSY97 data, I attempt to identify effects of fit along one observable vector: religious affiliation. I test whether students matched to schools of their own religion have outcomes different from their un-matched peers. OLS and IV estimates of religious match-effects range between +5 and +8 percentile points on standardized tests. The IV analysis relies upon a relative distance instrument previously unused in school studies. I discuss advantages and limitations of the instrument.

Keywords: Educational economics, human capital, productivity, school choice

Choice-based school reforms have been promoted on two grounds: first, on the supposition that school choice introduces a market-like competitive mechanism which may raise educational efficiency and productivity. Several studies, notably Hoxby (1994, 2000, and 2002) and Epple and Romano (1998), have examined whether various types of choice (intra- and inter-district, public or private) appear to provide these benefits.

The second motivation for choice-based school reform involves a common-sense notion of sorting according to student and school characteristics. It is supposed that in a given educational market there may be pairings of students and schools which are more or less productive. Under this assumption, efficiency gains can be made under constant school quality simply by re-assigning students to more productive pairings. The Utah legislature seemed to have had this in mind when it argued that parents are "best informed to make decisions for their children, including the educational setting that will *best serve* their children's interests and educational needs" (emphasis added).[1] Likewise, parents who exercise school choice often use the language of goodness of fit, or "good fit" when selecting a school.[2] This second argument for school-choice, namely, that a good fit has value for educational outcomes, has not been studied explicitly.

There are two primary empirical questions: first, whether sorting according to "interests and needs" is productive *in itself*, independently of school quality; second, whether the magnitude of such an effect is sizeable enough to be of importance for school-reform policy. In this paper I address these questions with respect to the most common type of sorting in the American school system—sorting by religious affiliation.

I. WHAT IS A GOOD FIT?

What exactly does it mean for a school to be a good fit? Let "fit" be a multi-dimensional interaction effect between students and schools. A good fit may occur along several vectors: educational focus, pedagogical method, religious tradition, cultural values, racial and ethnic identity, behavioral identity, and individualized personal characteristics. The following are some considerations regarding dimensions of fit:

1. *Multiplicity.*—A particular student-school pairing may exhibit a high quality of fit along some vectors, and a lower quality along others.
2. *Heterogeneity.*—Some vectors of fit may be highly individualistic. Consider the case of ethnic identity. A high degree of ethnic fit may be very productive for some groups of students (such as minorities) but not for others. On the other hand, some vectors may act upon all students with similar effect.

Overall, it is important to identify some of the most common vectors along which productive pairings are found (if any), and to see which ones might operate *in common* for many students. This will have implications for what kind of aggregation would be appropriate in measuring effects of fit.

Is a good fit productive per se, or is it orthogonal to outcomes of policy importance? To address this question empirically, a few challenges in particular loom large. First, there is the difficulty posed by limited types of schools. The available stock of schools clearly does not represent all the types of schools which could be conceived of as making good pairings with students along various vectors. For example, we observe schools which use the Montessori method, and those which don't, but we may not observe schools which use the Montessori method, are in the Catholic tradition, and are single-gender.

A second difficulty is generated by the fact that some of the most productive vectors of fit may be unobservable, such as personality or self-identity. For example, consider the possibility that one aspect of a good fit may be pedagogical method. Some parents may discover, experimentally, as it were, that their child learns better in a Montessori program, but this knowledge is likely to be unobservable to policy makers and researchers. Further, due to the nature and magnitude of the default public school system, many parents will never observe that their children are better suited to a Montessori-method of instruction. Ultimately, the nature of this problem is two-sided: among those children in Montessori schools, researchers cannot distinguish between those who are there because their parents deemed it a better fit, and those who are there for other reasons. Similarly, among those children in public or other private schools, researchers cannot distinguish between those who are there because Montessori-method instruction was not a good fit for them, and those who are there because they have never been able to try Montessori-method instruction.

A third difficulty is the familiar identification problem resulting from the process of selection into good-fit schools. Correcting for this will involve many of the same techniques employed in the private and Catholic schools literature. Additionally, fit effects would need to be distinguished from school effects. This is highly doable given a rich data set and good instruments.

II. WHY RELIGIOUS MATCHING?

In this paper I pick just one vector of fit to test: religious affiliation. I ask whether students matched to schools by religion exhibit better outcomes than their unmatched peers.[3] There are several reasons why this is a good place to start. To begin, values-matching is correlated with positive outcomes in many situations where relationships matter, such as counseling and prison programs (Hurst et al. 2008; LaVigne et al. 2007). In psychotherapy, values-matching helps establish a so-called "therapeutic alliance" between clinicians and patients (Bordin 1979; Horvath and Symonds 1991; Martin et al. 2000). Religious matching between students and schools seems to be a valid analog to that sort of values "fit".

In addition to this, religious matching is a good vector of fit to test for many empirical reasons. First, religion is *observable* in both students and schools; we can readily measure whether there is a match. Second, religious affiliation is the *most common form of sorting* in the private school system—this allows for a reasonably large number of matches even within random samples drawn from the general population. As regards other types of matching (e.g. educational focus or pedagogy) we do not observe a meaningful supply of types of schools, so it would be hard to observe actual sorting with respect to these vectors.

Third, religion itself is *plausibly exogenous to academic outcomes*. Parents choose (or inherit) religious affiliation and typically affiliate their children with them; when children are young, affiliation is as if randomly assigned. As such, it will be easier to identify match-effects without a very serious concern over confounding with academic outcomes. In contrast, note the empirical challenge to estimating the effect on math scores of being matched to a math-science magnet school.

A fourth reason to study the effects of good fit through religious matching is that religious school attendance is often likely to be a *revealed-preference measure of a "good fit"* that includes more than one vector. To see this, consider that religious schools are chosen over competing (and cheaper) alternatives (public schools)—but more importantly, they are also chosen from among other possible religious or private schools in a given region. So unless parents pick a school randomly *solely* on the basis of religious preference, it is likely that religious matching may capture other types of fit that are less observable.

Fifth, for a large enough sample of students in religious schools we could *isolate school-effects from match-effects*. This possibility arises due to the fact that virtually all religious schools admit students of different creeds. So there is a small but potentially significant portion of students in religious schools that are not matched by religious affiliation (McDonald and Schultz 2008; Trivitt and Wolf 2011). The non-matched subgroup of students in religious schools will turn out to be very helpful (indeed critical) for identifying true match-effects and distinguishing them from school-quality.

Finally, and importantly, we know something about the *supply of religious schools*—at least enough to think about some ways of correcting for the selection problem in identifying the effects of a match.

III. RELATION TO OTHER WORK

This paper is related to a number of studies in the economic literature, most especially to the long line of work on the effects of Catholic schools beginning with Coleman, Hoffer, and Kilgore (1982), Coleman and Hoffer (1987), and Hoffer, Greeley, and Coleman (1985). These studies compared public and Catholic schools, often finding that Catholic schools outperformed public schools, even with the inclusion of extensive controls. Related early studies in sociology include Noell (1982), Lee, Bryk and Holland (1993) and Witte (1996).

Subsequent research in economics has aimed to clarify the character and magnitude of the effects of Catholic schools. One line of research (Evans and Schwab 1995; Eide et al. 2004; Dee 2004) has tended to relocate the effect away from quantitative outcomes (such as scores) and onto more qualitative outcomes, such as high school graduation, college matriculation, and civic participation. Another line has found quantitative effects, but only in terms of interactions with certain sub-populations (Neal 1997; Figlio and Stone 1999; Grogger and Neal 2000). Notably, Neal (1997) found real effects for urban minority students, with little to no effect for white suburban students. Given these idiosyncratic findings, it remains a puzzle whether and how Catholic schools affect students differently from other schools, public and private.

This paper diverges from previous research in that I attempt to measure an interaction effect between Catholic students and Catholic schools, but also between non-Catholic students and schools of their same religion. Effectively, I am asking a distinct question from the Catholic schools literature: whether religious schools have any special advantage for students of the same religion. If there is any sense in which this is true, religious matching itself has been an important omitted variable in studies of Catholic schools.

The literature beginning with Neal (1997) which identifies quantitative effects of Catholic schools only for certain sub-populations supports a generalized theory of "fit effects". While it is clear that religious-fit is not driving Neal's findings (since urban blacks are not typically Catholic), it is clear that Catholic schools appear to affect different groups of students differently. It would not be hard to suppose that another kind of fit may be identified as complementing, or competing with, interaction effects previously observed.

Regarding religious matching, the present investigation further complicates the question of how religious affiliation matters to studies of private schooling. First, there are those studies that take religious affiliation as a kind of instrument to predict Catholic school attendance—presuming (the identifying assumption) that religious affiliation matters more to attendance than to outcomes (Evans and Schwab 1995; Kim 2011). Second, there are those which take the opposite view, namely that religious affiliation is essentially endogenous to student outcomes and therefore useless as a reliable instrument (Altonji et al. 2005). But whereas Altonji et al. find the fault to lie in the possibility that the instrument is correlated to other variables which themselves effect outcomes, I raise the question here of whether religious affiliation as an instrument is unreliable more directly, that is, if matching itself is productive.

The question underlying my study, and the corresponding critique of religious affiliation as an instrument, share a basic intuition with Akerlof and Kranton (2002) in which a model of identity in schooling is presented which is closely related to the idea of match or fit. They postulate that student utility increases with proximity to a self-chosen social category. School administrators can try to influence social categories within a school in order to change the prescriptions that follow from students' chosen identities and consequent behavior. In contrast, my notion of matching supposes that the administrator's own categorization (and the consequent culture of the school, or "ideal student") is relevant to student outcomes. Despite this difference, the concepts of identity and match share the insight that the student's perception of familiarity and fit with the school environment (either at the level of the social group or at the level of the school) may have very strong influence over student outcomes—indeed possibly stronger than other traditionally studied variables. The critical intuition is to consider the student herself as a decision maker—whose choices and responses depend upon her school environment.

The remainder of the paper is structured as follows. I first describe my methodology, including measurement of the match variable, estimating equations, and presentation of the instrumental variables. In section three, I describe the NLSY97 data used in this study. In section four I present my findings which include estimates of the effect of religious matching which have never before been estimated. Finally, I include a discussion of the implications of the findings and directions for future research.

IV. METHODOLOGY

a. Religious Matching as a 0-1 Proxy for Fit

Although "fit" as described above is a conceptually rich variable, I do not attempt to measure a continuous level of religious matching, even though there may be "better" and "worse" matches with respect to this vector. Rather, I allow religious matching to be a simple binary variable, and measure it in the most intuitive way possible. I define an 'alpha' (or 'primary') match when a student of any particular religious affiliation attends a school of the same religious denomination in the academic year prior to the outcome measure (standardized test). Students who fit this description are coded '1'—and all students, including public school students, for whom the criteria fails are coded '0'.

I also define a second binary match variable, 'beta' (or 'broad') match, which includes all the alpha-match students, plus additional students who attend a school with a religious outlook *broadly similar* to their own religious affiliation. For example, Catholics in some Protestant schools and some Protestants in Catholic schools are coded as matched under this variable. I determined similarity according to the denominational family trees assembled by The Association of Religion Data Archives (www.thearda.com). More students are matched by the beta classification than the alpha, though the strength of religious identification under the beta-match is likely to vary more widely.

b. Empirical Strategy

I begin with the aim of estimating a standard linear outcomes production function (Equation 1) where y_i is test scores (age-adjusted), v_i is the *0-1* match variable, X_i is a vector of student characteristics, including demographics, income, family structure, religious affiliation, parent's education, and regional controls. θ_1 is the coefficient of interest.

$$y_i = \theta_0 + \theta_1 v_i + \Theta_2 X_i + e_i \quad (1)$$

Interaction effects between students and schools, or between particular sub-groups of students and types of schools, are suppressed here as I have only student-level data. In consequence, I am estimating a single interaction effect: the interaction between students and schools of the same religious affiliation. The coefficient of interest, θ_1 or the "match effect", has never been estimated. For that reason, even the OLS estimates presented in this paper have real significance. A biased estimate is preferable to no estimate.

c. Non-random Selection into Matched Schools and a New IV Strategy

Clearly, the match variable v_i is somewhat troublesome as an explanatory variable. Students whose families select a matched school may differ in unobserved ways from those whose parents did not select a matched school, and these differences may correlate with outcomes (Altonji et al. 2005). For example, if families with an especially high level of commitment to education are more likely to choose a matched school, we could observe upward bias in the estimate.

Alternatively, estimates of matching could be biased downward for three important reasons: first, if parents choose matched schools especially in the case where children are struggling in a local public school, match effects could be biased downward. Second, families who choose a matched school may make other (unobservable) sacrifices to do so, such as working a second job, or giving up other amenities in favor of paying tuition. These sacrifices may have a negative effect on outcomes, offsetting the overall advantage of the match. Finally, most matched schools are small church-based schools with low per-pupil budgets, and inferior structures.[4] If school-quality is not well identified in a given data set, then estimated religious match effects could be confounded with effects of school-quality.

In this study I use a new type of distance-based instrument—relative distance—which has been used in studies of hospital quality under the name "differential distance" (McClellan et al. 1994; Sloan et al. 2001) but never for private school studies. This instrument exploits variation in the *relative distance* between the nearest matched school and the nearest unmatched (public) school. The relative distance instrument compares an implicit cost of the good (travel distance) with the implicit cost of its substitute, and relies on the insight that students are not equally positioned with respect to local public schools. It is thus designed to be related to variation in *whether* students are matched, but plausibly unrelated to variation in student outcomes. Smaller relative distances should correlate with an increased likelihood of attending a matched school.

Earlier distance-based instruments used in private school studies were absolute distances (Tyler 1994; Neal 1997)—where the instrument measured the absolute distance between the home and the private school. Others have examined absolute distance interacted with some other factor (Altonji et al. 2005). All of these are, however, subject to a standard critique: that the choice of residence may be determined in part by proximity to a desired school. Therefore, living close to a private school can't be interpreted as a *random treatment* which affects the likelihood of attending private school. Rather proximity may capture exactly what one wishes to remove from the equation—the fact that a family chose the particular school in the first place (for unobserved reasons).

But the relative distance instrument should survive this critique somewhat more easily than an absolute distance instrument. Consider that if a family moves closer to a private school of choice, this may result in a *smaller or larger relative distance*, depending on the location and supply of public schools in a given region. The logical question can be broken into two parts. For some families, choice of residence is unrelated to choice of schooling. For these families, both the relative and the absolute distance instruments should be valid. For the remaining families who choose a home based on proximity to a public or private school, the relative distance instrument should continue to be valid so long as we assume that public schools are randomly distributed *with respect to the residence choices of those families*. (While regional differences in the supply or density of public schools may moderate this assumption somewhat, district- or county-level controls should clear the remaining doubt.) Importantly, for this latter group of families, the instrument will have little to no meaning: for them, we cannot interpret large or small relative distances as influencing likelihood of matching. But this matters little—what matters is that these families will not generate systematic *spurious associations* between relative distance and matching.

Having considered the two major criteria for a sound instrument—relevance and validity—there is additionally the question of whether the instrument is plausibly monotonic (see Angrist and Krueger 2001; Murray 2006; and Bound et al. 1995). Does the relative distance instrument predict matching equally well for all subgroups? Given the strong existing differences between whites, blacks and Hispanics with respect to religious affiliation, and the disproportionate supply of Catholic schools among religious schools, it is worth questioning whether the relative distance instrument can perform equally well for these racial and ethnic subgroups. This will have consequences for the interpretation of the IV estimates. In the results section below I present some evidence on this question. Theoretically, however, the challenge here is unrelated to the *relative* distance instrument per se. This critique applies to an absolute distance instrument as well.

Methodologically the relative distance instrument was created as follows: for a student of religion X in zip code Y, I obtain the distance to the nearest matched school (of religion X) and the nearest unmatched (public) school in zip code Y. I take the difference between these two distances and call this the "typical relative distance" (TRD): so-called "typical" since it will be the same differential for every student of religion X in zip code Y. I would like to be able to use an "actual relative distance" (ARD) for each student—but with only zip-code level specificity, many of the obtained differentials are just zero. For example, if a student lives in the same zip-code as a local public school and a matched school, the TRD for this student-zip-code-religion combination is just zero. I show the incidence of the zero-TRDs below.

This lack of specificity, the zero-TRDs, is the chief limitation of the relative distance instrument employed in this paper. It generates plausible but imprecise IV-estimates, as I will show below. However, this is a limitation entirely driven by the data in this sample. A better data set with more precise geo data could overcome this problem quite easily. There are reasons to remain optimistic about the value and future use of relative distance instruments in the private schools literature.

In this paper, I employ two instruments—the plain TRD described above, and the square of the TRD which is meant to account for the exponential costs of very large relative distances.

Although the two stages are estimated simultaneously (to obtain the correct standard errors) Equations 2 and 3 below show the implied first- and second-stage regression equations, respectively.

First-stage equation

$$v_i = \pi_0 + \Pi_1 X_i + \Pi_2 Z_i + \varepsilon_i \tag{2}$$

The variable v_i is the match variable, X_i is a vector of student demographic and geographic controls, and Z_i is the vector of excluded instruments.

Second-stage equation

$$y_i = \theta_0 + \theta_i \hat{v}_i + \Theta_2 \mathrm{X}_i + e_i \tag{3}$$

Here, y_i is student scores, $\hat{v}_i$ is the predicted probability of matching from the first stage, and X_i is again a vector of student demographic and geographic controls.

III. DATA

I employ data from the 1997 National Longitudinal Survey of Youth (first round), together with some limited data from the associated School Survey Files, the restricted access zip code files, 1997 Geocode Data, and the NCES Private School Universe Survey Data with the Common Core of Data. All the data is indexed at the student level.

Table 1 presents summary statistics for my sample. The original NLSY97 sample contains data on 8984 students who entered the survey in 1997. I dropped those for whom religion or school religion was unknown, for a total full sample of 8817 students. The 8817 students in my sample were 48.9% male, 52.0% white, 25.9% black, 21.2% Hispanic.[5] 49.0% of the students in my sample lived in an MSA outside of the central city area. 32.2% lived in an MSA in a center city, and 17.7% lived outside an MSA. Of the students in my sample, 3.4% were matched to their school religion by the alpha-match indicator, and an additional 1.5% were matched by the beta-match indicator, for a total of 4.9% matched at all. A full 95.1% were not matched.

The mean typical relative distance was 15.09 miles (indicating the nearest matched school was on average 15 miles further than the nearest public school). The standard deviation in relative distance was 19.77 miles. The relative distance was capped at 50 miles in each direction. 22.4% of the students in my sample have a zero 'typical' relative distance—which means that they share a zip code with both the nearest matched and unmatched schools.

I have data from two standardized tests: the PIAT Math Percentile (updated) score (available for 5925 students), and an ASVAB Math-Verbal Composite Percentile score (available for 6963 students). Both sets of scores are age-adjusted to 3-month intervals to control for the age at which the test was taken.

Table 1

Summary Statistics

			Matching			Typical Relative Distance		Mean Outcomes *(Percentile)*	
		Full Sample	Alpha-match	Beta-match	None	Mean *(in miles)*	Proportion Zero	PIAT-Math	ASVAB
	N=	8817	302	430	8387	8817	1972	5925	6963
Full Sample									
	8817	1.00	0.034	0.049	0.951	15.09 (19.77)	0.224	37.55 (27.37)	45.43 (29.18)
Sex									
Male	4312	0.489	0.033	0.048	0.952	15.48 (20.03)	0.213	37.04 (26.90)	46.39 (28.57)
Female	4505	0.511	0.036	0.050	0.950	14.71 (19.51)	0.234	38.02 (27.79)	44.5 (29.74)
Race*									
Black	2284	0.259	0.009	0.027	0.973	13.52 (17.25)	0.175	26.37 (23.85)	28.64 (23.77)
Hispanic	1866	0.212	0.025	0.030	0.970	8.47 (16.97)	0.271	29.68 (23.89)	35.60 (25.79)
White	4587	0.520	0.051	0.067	0.933	18.61 (21.15)	0.228	46.33 (27.31)	56.52 (27.59)
Geography*									
Rural *Not in MSA*	1564	0.177	0.018	0.026	0.974	27.01 (20.12)	0.155	37.77 (26.92)	43.39 (28.37)
Suburban *MSA, not city*	4322	0.490	0.039	0.054	0.946	13.34 (19.01)	0.220	40.72 (27.88)	50.17 (29.25)
Urban *MSA, city*	2838	0.322	0.038	0.055	0.945	11.29 (18.29)	0.265	32.77 (26.28)	39.45 (28.38)

Note. All statistics are sample or sub-sample means. Standard deviations are included in parentheses under the statistic for quantitative variables. Starred sections have category Ns that do not sum to 8817 because other categories are omitted.

Table 2 presents summary statistics on the type of school attended and religious affiliation of students in my sample. Of the 8817 students in my sample, 92.2% attended public schools, 3.5% attended Catholic school, 1.9% attended non-Catholic religious schools, and 0.8% attended other private schools with no religious affiliation.

Throughout the analyses I rely on student-reported religious preference, which allows me to have the fullest possible sample (since only about half of the parents report on religion). Of the 8817 students in my sample, 55.4% reported a Christian religion other than Catholic, 28.5% reported a Catholic religious affiliation, 3.2% reported another religious affiliation, including Jewish, Mormon, and various Eastern religions (Hindu, Buddhist, etc.), and 12.9% reported no affiliation.

Table 2
Summary Statistics (Student Schools and Religious Affiliations)

		Full Sample	Sex		Race*			Geography*		
			Male	Female	Black	Hispanic	White	Rural *(Not in MSA)*	Suburban *(MSA, not city)*	Urban *(MSA, city)*
	N=	8817	4312	4505	2284	1866	4587	1564	4322	2838
School Type*										
Public	8131	0.922	0.922	0.923	0.941	0.948	0.903	0.947	0.921	0.911
Catholic	308	0.035	0.033	0.037	0.024	0.025	0.045	0.015	0.039	0.042
Other Religious	164	0.019	0.019	0.019	0.009	0.008	0.028	0.016	0.021	0.017
Non-Religious Private	75	0.008	0.010	0.010	0.004	0.004	0.013	0.012	0.007	0.009
Student Religion										
Catholic	2512	0.285	0.272	0.297	0.049	0.650	0.256	0.178	0.331	0.279
Other Christian	4880	0.554	0.580	0.528	0.812	0.257	0.545	0.685	0.511	0.541
Other	285	0.032	0.032	0.032	0.011	0.013	0.050	0.013	0.039	0.034
None	1140	0.129	0.115	0.143	0.128	0.080	0.149	0.125	0.119	0.146

Note. All statistics are sample or sub-sample means. Starred sections have category *Ns* that do not sum to 8817 because other categories are omitted.

V. RESULTS

a. Matching

Who is matched by religion and how do they compare to their unmatched peers? My first set of results aims to characterize the matched population and highlight the differences between matched students and their unmatched peers. I present the raw findings instead of moving straight to the regression analysis because this is the first paper to examine religious matching directly. It is important to establish and characterize the size of the raw "match" effect before seeking to fully identify the effect in an econometric sense.

How do matched students compare to their unmatched peers? Table 3 provides descriptive statistics comparing matched and unmatched students.

Table 3

Matched Students Compared to their Unmatched Peers
(Both Match Classifications)

	Alpha-match N=302	Unmatched N=8515	Beta-match N=430	Unmatched N=8387
Demographics				
Male	0.464	0.490	0.481	0.489
Black	**0.070**	**0.266**	**0.142**	**0.265**
Hispanic	**0.152**	**0.214**	**0.133**	**0.216**
White	**0.768**	**0.511**	**0.714**	**0.510**
Family Background				
Mother's Education *(in years)*	**13.2**	**11.4**	**13.0**	**11.4**
Gross Family Income	**67,251**	**44,536**	**67,932**	**44,154**
Both Biological Parents	**0.712**	**0.482**	**0.698**	**0.479**
Single Biological Mom	**0.156**	**0.287**	**0.153**	**0.289**
Two parents, but Only Bio. Mom	**0.060**	**0.112**	**0.060**	**0.113**
Religiosity *(0-600)*	**403**	**377**	**412**	**376**
Geography				
Rural *Not in MSA*	**0.093**	**0.180**	**0.095**	**0.182**
Suburban *MSA, not city*	**0.553**	**0.488**	**0.540**	**0.488**
Urban *MSA, city*	0.354	0.321	**0.365**	**0.320**
Typical Relative Distance *in miles*	**4.22**	**15.48**	**8.02**	**15.45**
Outcomes				
PIAT Math Percentile	**51.82**	**37.03**	**51.31**	**36.85**
ASVAB Math-Reading	**64.26**	**44.71**	**63.05**	**44.48**
HS Graduation	**0.992**	**0.862**	**0.989**	**0.861**
Delinquency *(0-10)*	**0.94**	**1.35**	**0.93**	**1.35**

Note. Bold denotes statistically significant differences at the 95% level or higher.

There are several things to note in Table 3. First, matched students are more likely to be white and less likely to be either black or Hispanic. However, Hispanics are about twice as likely as blacks to be alpha-match students, while blacks and Hispanics are equally likely to be beta-match students. This is clearly because Hispanics are more likely to be Catholic, and most religious schools are Catholic.

Second, matched students of both classifications come from families with about 50% higher income than unmatched peers, and are significantly more likely to reside in an intact family. Interestingly, matched students are from families that are only moderately more religious than their unmatched peers.

Third, typical relative distance appears to correlate well with matching, which supports the relevance of the instrument. Alpha-match students average about 4.22 miles relative distance between matched and unmatched schools, while beta-match students average 8.02 miles relative distance. This means that for alpha-match students, the nearest matched school is about 4.22 miles further than the nearest local public school. For beta-match students the nearest matched school is 8.02 miles further than the nearest local public school. Unmatched students, by comparison, have a matched school that is on average about 15.5 miles further than the nearest local public school.

Finally, there is a significant raw "match effect". For the PIAT, alpha-match students score about 14.8 percentile points higher than their unmatched peers, and beta-match students score about 14.5 points higher than their unmatched peers. For the ASVAB, alpha-match students score about 19.6 percentile points higher than their unmatched peers, and beta-match students score about 18.6 points higher. These are about one-half (PIAT) to two-thirds (ASVAB) of a standard deviation in scores respectively. This is the effect I am interested in exploring—does it remain after controlling for all the other significant differences between matched and unmatched students? In the tables that follow I try to put this effect in context, to see how it compares with effects that are more familiar, such as the Catholic-schools effect, and the private school effect.

Is the raw match effect the same as the Catholic schools effect? Table 4 first shows the raw "Catholic schools effect", and then breaks this down to show the "match effect" within the Catholic schools effect.

Table 4
Catholic School Effects Compared to Match Effects Within Catholic Schools

	Catholic School Effects		Catholic School: Matched vs. Unmatched	
	Catholic N=318	Non-Catholic N=8499	Alpha-match N=251	Unmatched N=67
Demographics				
Male	0.462	0.490	0.454	0.493
Black	**0.173**	**0.262**	**0.080**	**0.522**
Hispanic	**0.160**	**0.214**	0.175	0.104
White	**0.660**	**0.515**	**0.737**	**0.373**
Family Background				
Mother's Education *(in years)*	**12.97**	**11.43**	13.1	12.6
Gross Family Income	**65,637**	**44,554**	66,546	62,234
Both Biological Parents	**0.714**	**0.481**	0.737	0.627
Single Biological Mom	**0.151**	**0.287**	0.143	0.179
Two parents, but Only Bio. Mom	**0.066**	**0.112**	0.056	0.104
Religiosity *(0-600)*	384.4	377.9	380.9	401.3
Geography				
Rural *Not in MSA*	**0.072**	**0.181**	0.080	0.045
Suburban *MSA, not city*	**0.550**	**0.488**	0.578	0.448
Urban *MSA, city*	**0.377**	**0.320**	**0.343**	**0.507**
Typical Relative Distance *in miles*	**4.47**	**15.48**	**2.12**	**13.25**
Outcomes				
PIAT Math Percentile	**50.27**	**37.09**	**52.00**	**41.69**
ASVAB Math-Reading	**62.29**	**44.78**	**64.43**	**53.08**
HS Graduation	**0.996**	**0.862**	0.995	1.000
Delinquency (0-10)	**1.06**	**1.34**	1.03	1.16

Note. Bold denotes statistically significant differences at the 95% level or higher.

The raw Catholic schools effect is about 13 percentile points in the PIAT and 18 points in the ASVAB. These are very similar to the raw match effect. This is no surprise since the set of Catholic school students and the set of alpha-match students overlap significantly. Of the 318 Catholic school students, 79% are also alpha-match students, and of 302 alpha-match students, 83% are also in Catholic schools. Much more interesting are the outcome differences between alpha-match and unmatched students *conditional upon being in a Catholic school*. If the match effect from Table 3 were simply a Catholic schools effect, then we should not expect to see a large difference here. But in fact, conditional upon being in a Catholic school, there is a remaining 10 point percentile advantage in the PIAT and 11 point percentile advantage in the ASVAB for alpha-match students and these differences are statistically significant. Note also that there are almost no other statistical differences between these groups, with the exception of the percent black and white students. Income, parental education, religiosity and family structure variables are all highly similar. So, it appears that there is some effect which is different from a so-called "Catholic schools effect"—and this effect is experienced by matched students in non-Catholic schools to the same degree as Catholic students in Catholic schools.

Is the raw match effect some other kind of Catholic effect? Table 5 compares Catholic students by matching status, and then matched students by Catholic status.

The results here are quite striking. In the first two columns, we can observe the effect of matching *conditional upon being Catholic*: 15 points (PIAT) and 19 points (ASVAB) respectively. Note that this is the same as the "match effect for Catholic students". Now consider the second pair of columns. The effect of being a Catholic student, conditional upon matching, is negligible. The difference is not even statistically significant. Therefore, there is no "Catholic effect for matched students".

Table 5

Match Effects for Catholic Students and Catholic Effects for Matched Students

	Catholic Students: Matched vs. Unmatched		Matched Students: Catholic vs. Non-Catholic	
	Alpha-match N=251	Unmatched N=2261	Catholic Alpha-match N=251	Non-Catholic Alpha-match N=51
Demographics				
Male	0.454	0.468	0.454	0.510
Black	**0.080**	**0.041**	0.080	0.020
Hispanic	**0.175**	**0.517**	**0.175**	**0.039**
White	**0.737**	**0.437**	**0.737**	**0.922**
Family Background				
Mother's Education *(in years)*	**13.1**	**10.7**	13.1	13.6
Gross Family Income	**66,546**	**43,439**	66,546	70,723
Both Biological Parents	**0.737**	**0.585**	**0.737**	**0.588**
Single Biological Mom	**0.143**	**0.243**	0.143	0.216
Two parents, but Only Bio. Mom	0.056	0.081	0.056	0.078
Religiosity *(0-600)*	**380.9**	**346.7**	**380.9**	**501.2**
Geography				
Rural *Not in MSA*	**0.080**	**0.114**	**0.080**	**0.157**
Suburban *MSA, not city*	0.578	0.569	0.578	0.431
Urban *MSA, city*	0.343	0.312	0.343	0.412
Typical Relative Distance *in miles*	**2.12**	**7.29**	**2.12**	**14.52**
Outcomes				
PIAT Math Percentile	**52.00**	**36.76**	52.00	50.94
ASVAB Math-Reading	**64.43**	**45.57**	64.43	63.45
HS Graduation	**0.995**	**0.853**	0.995	0.976
Delinquency (0-10)	**1.03**	**1.29**	**1.03**	**0.490**

Note. Bold notes statistically significant differences at the 95% level or higher.

Table 6
PRIVATE SCHOOL EFFECTS COMPARED TO MATCH EFFECTS WITHIN PRIVATE SCHOOLS

	Private School Effects		Private School: Matched vs. Unmatched		
	Private N=686	Public N=8131	Alpha-match* N=302	Beta-match* N=430	Unmatched** N=256
Demographics					
Male	0.490	0.489	0.464	0.481	0.504
Black	**0.197**	**0.264**	**0.070**	**0.142**	**0. 289**
Hispanic	**0.141**	**0.218**	0.152	0. 133	0. 156
White	**0.647**	**0.510**	**0.768**	**0.714**	**0.535**
Family Background					
Mother's Education *(in years)*	**12.6**	**11.4**	**13.2**	**13.0**	**11.9**
Gross Family Income	**63,234**	**43,802**	**67,251**	**67,932**	**55,344**
Lives with Both Biological Parents	**0.622**	**0.479**	**0.712**	**0.698**	**0.496**
Lives Only with Biological Mom	**0.201**	**0.289**	**0.156**	**0.153**	**0.281**
Has two parents, but Only Bio. Mom	**0.074**	**0.113**	0.060	**0.060**	**0.098**
Religiosity *(ranked 0-600)*	**399.7**	**376.2**	402.7	**411.7**	**377.6**
Geography					
Rural *(Not in MSA)*	**0.121**	**0.182**	**0.093**	**0.095**	**0.164**
Suburban *(MSA, not city)*	0.499	0.489	**0.553**	**0.540**	**0.430**
Urban *(MSA, city)*	**0.369**	**0.318**	0.354	0.365	0.375
Typical Relative Distance *(in miles)*	**9.88**	**15.53**	**4.22**	**8.02**	**13.01**
Outcomes					
PIAT Math Percentile	**46.09**	**36.86**	**51.82**	**51.31**	**36.31**
ASVAB Math-Reading Composite	**57.81**	**44.36**	**64.26**	**63.05**	**48.33**
HS Graduation	**0.942**	**0.861**	**0.992**	**0.989**	**0.860**
Delinquency *(ranked 0-10)*	1.22	1.34	**0.94**	**0.93**	**1.72**

Note. Bold notes statistically significant differences at the 95% level or higher. *Significance is as compared to private school unmatched **Significance is as compared to private school beta-match

Is the raw match effect the same as the private schools effect? Finally, Table 6 presents descriptive statistics for private school effects, and the match effect *conditional upon private school attendance*. This is analogous to table 4.

There are two especially interesting things to note here. First, the "raw" private schools effect is not as big as the raw match effect. There is a 9 point advantage for private school students on the PIAT, and a 13.5 point advantage on the ASVAB. These are about 40 percent and 30 percent smaller than the match effects of 15 and 19 respectively. The second interesting thing is the match effect conditional upon being in a private school. For alpha-match students, the effect is 16 percentile points on both outcomes, and for beta-match students, the effect is 15 points on both outcomes. So, again, it does not seem that a "match" effect is the same as a private schools effect.

b. Match Effects: OLS Regressions

Full sample OLS findings Table 7 reports the OLS results for the alpha-match indicator (Panel A) and the beta-match indicator (Panel B), respectively, in two specifications for each outcome variable for a total of eight regressions. The categorical variables for race (white is baseline) are included to provide a reference for the magnitude of the effects.

Table 7
Religious Matching and Educational Outcomes
Ordinary Least Squares

Explanatory Variable	Dependent Variable: *PIAT-Math*		Dependent Variable: *ASVAB*	
Panel A	(1)	(2)	(3)	(4)
Alpha-match	5.662***	5.730**	8.766***	5.983**
	(1.987)	(2.665)	(1.721)	(2.309)
Private school (dummy)	---	-0.071	---	2.937*
		(1.854)		(0.593)
Black	-12.994***	-12.995***	-17.740***	-17.734***
	(1.087)	(1.088)	(1.014)	(1.013)
Hispanic	-8.900***	-8.901***	-11.821***	-11.817***
	(1.141)	(1.141)	(1.089)	(1.089)
R^2	0.2394	0.2394	0.3299	0.3303
Root mean square error	24.177	24.179	24.148	24.143
Observations	5920	5920	6958	6958
Panel B	(5)	(6)	(7)	(8)
Beta-match	6.457***	10.285***	8.632***	8.077***
	(1.629)	(2.758)	(1.428)	(2.320)
Private School (dummy)	---	-3.973*	---	0.577
		(2.304)		(1.941)
Black	-12.972***	-12.998***	-17.763***	-17.760***
	(1.086)	(1.086)	(1.012)	(1.012)
Hispanic	-8.840***	-8.877***	11.790***	-11.792***
	(1.139)	(1.140)	(1.087)	(1.087)
R^2	0.2404	0.2410	0.3308	0.3308
Root mean square error	24.161	24.157	24.131	24.133
Observations	5920	5920	6958	6958

Note. Huber robust standard errors are in parentheses. Regression errors are clustered at the household level. Student-level controls include gender, religion, MSA, household income, family structure, mother's education. Other controls include Census Bureau region, county, and rural/suburban/urban designation. *, **, ***= significant at the 90, 95%, and 99% level or higher, respectively.

The first two results of interest are the top coefficients in columns (1) and (3). These present the match effect for the alpha-match indicator for a specification that includes a full set of controls except a private school dummy. The match effect for alpha-match is +5.7 for the PIAT and +8.8 for the ASVAB. These are statistically significant at the 99% level, and large in magnitude—1/5 and 1/3 of a standard deviation in scores respectively. To put these in context, they are on the order of half the size of the black-white test score gap for this data. For the beta-match, which includes both the alpha-match students and 128 additional students who are matched to schools with very similar religious affiliations, columns (5) and (7) show that the match effects are +6.5 percentile points and +8.6 for the two outcomes respectively. These are, again, highly statistically significant.

In specifications (2) and (4) respectively, I include a private school control. For the PIAT test, the private school variable has no statistical significance, and is small in magnitude. For the ASVAB, the private school variable is positive and significant, on the order of +3 points, diminishing the match effect by a similar amount. I do not attempt a control for Catholic schools since it would be so highly correlated with the alpha-match variable. In the corresponding regressions for the beta-match, (6) and (8), the private school dummy has the opposite effect—it strengthens the match effect for the PIAT outcomes, and has no significant effect on the ASVAB scores. The conflicting findings for the private school dummy result, most likely, from the high degree of overlap between private and matched schools in my sample. These estimates are probably not trustworthy. This underscores the need to replicate this study in a much larger sample of private and religious schools where school-type effects can be identified more easily. However—with or without the private school control—both alpha- and beta-matches appear beneficial to scores to have a large and significant effect, by either match specification.

Table 8
Religious Matching and Educational Outcomes by Race
Ordinary Least Squares

Explanatory Variable	Dependent Variable: PIAT-Math					Dependent Variable: ASVAB				
	Full Sample	Without Blacks	Black Only	Hispanic Only	White Only	Full Sample	Without Blacks	Black Only	Hispanic Only	White Only
Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Alpha-match	5.662***	5.400***	10.682	4.117	5.114**	8.766***	8.592***	9.310*	11.804**	7.370***
	(1.987)	(2.078)	(7.609)	(4.454)	(2.382)	(1.721)	(1.822)	(5.072)	(4.798)	(2.028)
R^2	0.2394	0.2033	0.2478	0.1725	0.1662	0.3299	0.2571	0.2806	0.2732	0.1800
Root mean square error	24.177	24.914	21.468	22.705	25.522	24.148	25.043	20.850	22.931	25.446
Observations	5920	4360	1560	1241	3064	6958	5193	1765	1331	3796
Panel B	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Beta-match	6.457***	6.234***	8.977**	4.464	6.327***	8.632***	8.325***	11.792***	10.734**	7.172***
	(1.629)	(1.781)	(4.122)	(3.880)	(2.057)	(1.428)	(1.598)	(3.194)	(4.225)	(1.793)
R^2	0.2404	0.2043	0.2492	0.1728	0.1676	0.3308	0.2578	0.2848	0.2733	0.1806
Root mean square error	24.161	24.898	21.449	22.701	25.501	24.131	25.032	20.788	22.929	25.436
Observations	5920	4360	1560	1241	3064	6958	5193	1765	1331	3796

Note. Huber robust standard errors are in parentheses. Regression errors are clustered at the household level. Student-level controls include gender, religion, MSA, household income, family structure, mother's education. Other controls include Census Bureau region, county, and rural/suburban/urban designation.
*, **, ***= significant at the 90, 95%, and 99% level or higher, respectively.

OLS Results by Racial Subgroups A good question, examining these results, is whether the effect observed is really driven by the interaction effects identified earlier by Neal (1997), namely by large effects for urban minorities in Catholic schools. While my data are not rich enough to provide a full econometric answer to this question, Table 8 should provide some evidence to the contrary. Here I present OLS regressions identical to those in Table 7, but with the full sample omitting black students, and then for black, Hispanic, and white students separately. The coefficients on the separate race regressions, columns (3)-(5), can be interpreted as a proxy for the interaction effect between race and matching.

What we observe here is that the large and significant match effects are not driven by minority effects. For the PIAT the match effects are large and significant for the full sample without blacks (+5.4 for the alpha-match, +6.2 for the beta-match) and also for whites without black or Hispanic students (+5.1 for the alpha-match and +6.3 for the beta-match). A similar pattern obtains for the ASVAB. We do see that the point estimates for blacks are generally larger than for whites. That they lack significance under the alpha-match is entirely a function of the very small number of alpha-matched black students (only 21). There are three times as many black students who are beta-matched (61), and thus the beta-match coefficients are estimated more precisely. While these coefficients may not completely settle the question—they provide very strong evidence that the match effects identified in Table 7 are not minority interaction effects masquerading as a match.

c. Match Effects: IV Results

First stage results Table 9 presents evidence on the relation between the relative distance instrument and the likelihood of matching. The first-stage regresses the 0-1 match variable upon the instrument(s) in order to establish relevance. The first three regressions, columns (1), (2) and (3), show the results for the alpha-match indicator. The first column shows that the instruments jointly have significance for the probability of a match in the absence of controls. Regressions (2) and (3) include the full set of controls: regression (2) shows that the TRD instrument alone performs worse than it does with the second instrument, TRD^2, which is presented in the third regression. A similar finding obtains for the beta-match indicator. We learn here that the instruments are consistently significant at the highest possible level, but the coefficients are small in magnitude. This suggests a weak instrument problem (Bound et al. 1995; Murray 2006)—no surprise given the major limitation of the instrument mentioned above, which is a high percentage of zero-value instruments. The usual diagnostic for the first-stage (Stock, Wright, and Yogo 2002) is an implied F-statistic for the first stage. These are presented along with the IV regressions in table 11. (All are above 20.) Based on the R^2 and the coefficients the first-stage results suggest using both instruments.

Table 9
Relative Distance And Religious Matching (First-Stage)

Explanatory Variable	Ordinary Least Squares Dependent Variable: *Alpha-match*			Ordinary Least Squares Dependent Variable: *Beta-match*		
	(1)	(2)	(3)	(4)	(5)	(6)
Typical relative distance	-0.003***	-0.001***	-0.003***	-0.003***	-0.001***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Typical relative distance^2	-0.000***	---	0.000***	0.000***	---	0.000***
	(0.000)		(0.000)	(0.000)		(0.000)
Control variables	No	Yes	Yes	No	Yes	Yes
R^2	0.0139	0.1001	0.1028	0.0079	0.0892	0.0912
Root mean square error	0.181	0.173	0.173	0.215	0.206	0.206
Observations	8809	8809	8809	8809	8809	8809

Note. Huber robust standard errors are in parentheses. Regression errors are clustered at the household level. Student-level controls include gender, religion, MSA, household income, family structure, mother's education. Other controls include Census Bureau region and rural/suburban/urban designation. *, **, ***= significant at the 90, 95%, and 99% level or higher, respectively.

Reduced form results Table 10 presents the reduced-form results: OLS regression of the scores on the relative distance instruments. This provides statistical support for the exclusion restriction. Regressions (1) and (2) show that the TRD and TRD^2 have no explanatory power for PIAT scores after the standard controls are included. The same pattern is observed for the ASVAB scores—the instruments have no predictive power in the presence of full controls.

Table 10
Relative Distance And Educational Outcomes (Reduced-Form)

Explanatory Variable	Ordinary Least Squares Dependent Variable: *PIAT-Math*		Ordinary Least Squares Dependent Variable: *ASVAB*	
	(1)	(2)	(3)	(4)
Typical relative distance	-0.203*	-0.025	-0.263***	-0.010
	(0.113)	(0.110)	(0.094)	(0.084)
Typical relative distance^2	0.007***	0.000	0.008***	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Control variables	No	Yes	No	Yes
R^2	0.0066	0.2043	0.0082	0.3029
Root mean square error	34.284	30.777	29.069	24.432
Observations	5922	5922	6958	6958

Note. Huber robust standard errors are in parentheses. Regression errors are clustered at the household level. Student-level controls include gender, religion, MSA, household income, family structure, mother's education. Other controls include Census Bureau region and rural/suburban/urban designation. *, **, ***= significant at the 90, 95%, and 99% level or higher, respectively.

Second-stage IV results Table 11 presents the main IV results for the match effects. For the two separate outcomes, PIAT and ASVAB, I present four regressions for each match indicator, for a total of 16 regressions. The first regression in each quadrant is simply the OLS regression from the relevant columns of table 7. These are provided for ease of comparison with the IV results. The second regression provides the results for the preferred IV-2SLS specification which includes both instruments. The third regression replicates the second but with a LIML estimation (Angrist and Pischke 2008). The fourth presents the results for a just-identified specification using only the TRD instrument (Angrist and Pischke 2008; Murray 2006).

The upper panel presents the findings for the alpha-match indicator and the lower panel presents the findings for the beta-match indicator. As for the OLS regressions, black and Hispanic categorical indicators (white is baseline) are included to provide a point of comparison for the magnitude of the effects.

Columns (2) and (3) show a pattern of IV coefficients that is repeated in the four quadrants of the table. The magnitude and direction of the IV alpha-match effects are fairly similar to the magnitude and direction of the OLS coefficients: +7.1 (IV-2SLS) compared to +5.7 (OLS). However, the standard errors on the IV estimates are large, and so the coefficients lack statistical significance. In general, the IV-LIML coefficients are very similar to the IV-2SLS which is comforting since LIML estimation is less biased, and provides a cross-check on the IV-2SLS estimate. Finally, the coefficient in column (4) is what results from the just-identified regression using only the TRD instrument, and not the TRD^2. The fact that this coefficient (+15) is so much larger than that in column (2) is somewhat troubling. We expect that the just-identified regression is least biased if there is at least one strong IV—however, the first-stage regressions show that the IVs perform much better together. Since they are relatively weak, the results from this just-identified regression are not preferable to the ones in columns (2) and (3).

The implied first-stage F-statistic is 21.105 for these regressions, comfortably above 10 [Stock, Wright and Yogo, 2002]. Finally, the Hansen statistic is small, which means the hypothesis that the instruments are jointly valid can't be rejected. This depends upon the assumption that at least one of the instruments is valid [Murray, 2006].

The IV findings for the alpha-match effect on the ASVAB are similar in spirit: +14.5 (IV-2SLS) compared to +8.8 (OLS). For the beta-match indicator: PIAT is +7.0 (IV-2SLS) compared to +6.5 (OLS), and ASVAB is +14.8 (IV-2SLS) compared to +8.6 (OLS).

The F-statistics are in general better for the ASVAB regressions, as high as 44.8 for the alpha-match, which means that the instruments are not hopelessly weak in spite of being somewhat meaningless for a large portion of my sample.

Table 11
RELIGIOUS MATCHING AND EDUCATIONAL OUTCOMES (Second-Stage)

EXPLANATORY VARIABLE	Dependent Variable: PIAT-Math				Dependent Variable: ASVAB			
	OLS	IV-2SLS	IV-LIML	IV-2SLS	OLS	IV-2SLS	IV-LIML	IV-2SLS
Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Alpha-match	5.662***	7.106	7.114	15.001	8.766***	14.487	14.564	25.403
	(1.987)	(23.336)	(23.491)	(28.131)	(1.721)	(17.556)	(17.776)	(21.532)
Black	-12.994***	-12.889***	-12.889***	-12.797***	-17.740***	-17.999***	-17.998***	-17.818***
	(1.087)	(1.090)	(1.091)	(1.104)	(1.014)	(1.006)	(1.007)	(1.024)
Hispanic	-8.900***	-8.571***	-8.571***	-8.143***	-11.821***	-11.399***	-11.395***	-10.790***
	(1.141)	(1.695)	(1.701)	(1.884)	(1.089)	(1.409)	(1.417)	(1.571)
R^2	0.2394	0.1946	0.1946	0.1913	0.3299	0.3041	0.3041	0.2947
Root mean square error	24.177	24.570	24.570	24.620	24.148	24.340	24.340	24.510
F-statistic	---	21.105	21.105	25.370	---	34.586	34.586	44.796
Hansen's statistic	---	0.258	0.258	---	---	0.780	0.780	---
Panel B	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Beta-match	6.457***	6.998	7.002	11.156	8.632***	14.792	14.886	23.670
	(1.629)	(18.820)	(18.912)	(20.799)	(1.428)	(16.954)	(17.202)	(20.050)
Black	-12.972***	-12.898***	-12.898***	-12.854***	-17.763***	-18.049***	-18.048***	-17.934***
	(1.086)	(1.068)	(1.069)	(1.071)	(1.012)	(0.985)	(0.986)	(0.995)
Hispanic	-8.840***	-8.594***	-8.594	-8.379***	-11.790***	-11.391	-11.386	-10.900***
	(1.139)	(1.477)	(1.480)	(1.534)	(1.087)	(1.378)	(1.387)	(1.499)
R^2	0.2404	0.1955	0.1955	0.1942	0.3308	0.3043	0.3042	0.2943
Root mean square error	24.161	24.550	24.550	24.570	24.131	24.340	24.340	24.510
F-statistic	---	22.319	22.319	32.405	---	26.863	26.863	35.098
Hansen's statistic	---	0.209	0.209	---	---	0.697	0.697	---
Instruments								
TRD	---	Yes	Yes	Yes	---	Yes	Yes	Yes
TRD2	---	Yes	Yes	No	---	Yes	Yes	No
Observations	5920	5920	5920	5920	6958	6958	6958	6958

Note. Huber robust standard errors are in parentheses. Regression errors are clustered at the household level. Student-level controls include gender, religion, MSA, household income, family structure, mother's education. Other controls also include Census Bureau region, county, and rural/suburban/urban designation. *, **, *** denote significance at the 90, 95%, and 99% levels or higher.

What can we conclude from Table 11? It would be difficult to argue that the point estimates for the IV match-effects have much meaning given the size of the standard errors. However, if we think about the IV coefficients as a check on the underlying bias of the OLS findings resulting from endogeneity or selection effects, then a plausible story here is that the IV results support the direction and magnitude of the OLS findings.

The IV-diagnostics are generally good—the instrument performs well in the first- and second-stages, but since the instrument has a zero-value so often, as does the match variable (0-1 with only a small portion of students having a 1), there is an inherent difficulty in predicting 0-1 matches with a variable which lacks specificity. In fact, given these limitations, it is somewhat surprising that the TRD performs as well as it does. This means that a second lesson from the IV results is that the relative distance instrument should be considered as having some promise for future school studies, and should be tried with more precise geo-data.

Monotonicity of the Instrument Table 12 presents evidence on whether the instrument performs equally well for the various racial subgroups. This is motivated by the fact that religious matching itself is a function of religious affiliation and the supply of religious schools—both of which are experienced by racial groups quite differently. The regressions in Table 12 essentially reproduce the first-stage regressions from Table 9 but for blacks, Hispanics and whites separately. The results confirm what might be suspected—that the relative distance instrument predicts matching for Hispanics and whites fairly well, but not well at all for the black population in my sample. Considering, however, that so few black students are matched at all in this sample (21 and 61 respectively) it is not clear that the lack of monotonicity arises from the failure of the instrument per se, or from the limitations of the data (little variation in matching for blacks, and many zero-TRDs).

It should be reassuring, on the other hand, that the instrument performs fairly consistently for both whites and Hispanics. While it is a better predictor for whites, the differences between them are first-order and not second-order. Overall, Table 12 provides some evidence that the IV is plausibly monotonic. We already knew that the IV estimates for this sample lack precision, so this finding of non-monotonicity cannot add much to our caution about the IV coefficients. However, there is enough in Table 12 to suppose that the relative distance instrument could perform better with more specific data, and should not be abandoned too easily.

Table 12

Relative Distance and Religious Matching:
First-stage Results by Race

Explanatory Variable	Ordinary Least Squares Dependent Variable: *Alpha-match*						Ordinary Least Squares Dependent Variable: *Beta-match*					
	Black	Black	Hispanic	Hispanic	White	White	Black	Black	Hispanic	Hispanic	White	White
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
TRD	-0.001**	0.000	-0.003***	-0.003**	-0.005***	-0.006***	-0.001	-0.001	-0.003***	-0.003**	-0.004***	-0.006***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
TRD^2	0.000*	0.000	0.000**	0.000**	0.000***	0.000***	0.000	0.000	0.000***	0.000**	0.000**	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Control variables	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.0050	0.1821	0.0068	0.0474	0.0257	0.1223	0.0028	0.1017	0.0063	0.0485	0.0167	0.1109
Observations	2284	2284	1866	1866	4587	4587	2284	2284	1866	1866	4587	4587

Note. Huber robust standard errors are in parentheses. Regression errors are clustered at the household level. Student-level controls include gender, religion, MSA, household income, family structure, mother's education. Other controls include Census Bureau region, county, and rural/suburban/urban designation. <1% of Black students are alpha-matched and <3% are beta-matched. The very small sub-sample size may account for the lack of significance in this group. *, **, ***= significant at the 90, 95%, and 99% level or higher, respectively.

VI. DISCUSSION

a. Contributions and Limitations of the Present Study

This paper finds large and significant effects of religious matching on student outcomes using NLSY97 data. These effects are on the order of +5 to +8 percentile points in standardized tests. These estimates—even from the OLS—are important because they have never been estimated before. Issues of bias are likely to plague the question of religious match effects, as they do the entire class of related private and Catholic schools literature. But the new instrument used here, relative distance, can be improved upon with more accurate geo data. The data used here were significantly limited due to the lack of finer address data. Many students lived in the same zip code as both the nearest matched and unmatched school—rendering a relative distance of "zero". Better data is attainable in theory, and represents a realistic future research plan. In spite of this, the IV estimates are similar in direction and magnitude to the OLS estimates, and provide some formal support for the direction and magnitude of the OLS effects.

b. Believability of the Estimates

The match effects identified here are large—ranging between 1/5 and 1/3 of a standard deviation in test scores—or about ½ the size of the black-white test score gap in this data. How much should we trust these estimates? Are they believable?

In the first place, to what extent do we expect the OLS estimates to suffer from self-selection bias? This question cannot be answered directly. But it is worth pointing out that the match-effects in OLS should suffer less from self-selection bias than in the estimate of a general Catholic schools effect, or a general private schools effect. In these cases, one readily sees that families may choose a Catholic or private school for brand reasons, or a general sense that these schools are "better" than public alternatives (Trivitt and Wolf 2011). But the sort of selection effect that would plague the findings in this paper would depend upon a story about parents believing that a "religiously-matched" school would be better for scores. This is dubious and seems to run contrary to common sense. Therefore, the OLS findings should stand a little firmer against concerns about selection bias than OLS findings for general Catholic or private schools. (This does not rule out other kinds of bias, though, such as those listed in section 2.3.)

A second question related to believability is whether we know that the match effect here is not really some other effect previously identified as being important in Catholic schools—especially, the interaction effect for minorities in urban areas (Neal 1997). Table 8 presents my best possible answer to this question. When the OLS findings have been broken down by minority group, match effects, though largest for black students, remain large and statistically significant for white and Hispanic students as well. This augurs against the hypothesis that the match-effects are merely driven by minority students in urban Catholic schools.

A final question of importance related to believability is whether the match effects here are really Catholic school effects, or private school effects. Although I cannot isolate these effects from the match effect in this paper since there is too little variation in matches apart from Catholic schools, I *have shown* that thinking about students in religious schools as treated with a "religious-match" (or not) may be as fruitful as thinking about them as treated with a "catholic school". Put differently, my results really flip the question around: although I can't reject that hypothesis that the match effect is really a Catholic school effect, there is enough evidence here to suggest that previous studies identifying a Catholic school effect may have really identified a match effect. My results suggest that matching should be taken seriously as a new way of thinking about *who is being treated with what* in the analysis of private and religious schools.

What about the objection that previously identified Catholic schools effects are especially observable for black and Hispanic urban students, only some of which (Hispanics) are typically Catholic? This would seem problematic if we think in terms of the alpha-match indicator which requires strict matching in religious creed. But if we think about the beta-match indicator things get more interesting. American blacks have long been characterized by greater religiosity than whites (Taylor et al. 1996; Taylor et al. 1999). It may well be that urban black families, though Protestant, are more religiously "identified" with Catholic schools than previously thought. This may be especially true in an era (post-Vatican II) when Catholic schools exhibit fewer of the traditional accoutrements of Catholic schools, such as the presence of religious sisters, and have come to reflect a more mainstream American "Christianity". There is more to uncover in the literature on the sociology of religion which is highly relevant to the question of fit and match in the sense of religious tradition. See especially Chaves (1994) on religiosity and authority, and Woodberry et al. (2012) on the measurement of American religious traditions.

Ultimately, the OLS estimates presented here are a starting place for future research, though I make no claim to have exhausted all the possible ways to make this effect disappear.

c. Intriguing Questions Raised by the Present Study

This paper and its findings raise the intriguing question of what it is that schools actually do. What is the mechanism by which schools generate or promote academic growth in students? Consider two models, or analogies, for schooling. The first, a factory model, or black box, supposes that schools *act upon* students in the manner of producing something from raw materials. The second, a garden model, supposes that schools provide *an environment* for the mind to develop naturally, according to its own pace and progress, as the soil for a plant. If the factory model gets something right, then a good school can be identified in part by the objective quality of various inputs (teachers, curriculum, physical plant, and even the quality of the peers). However, if the garden model gets something right, the effectiveness of a school may vary tremendously by student, and this variation may depend upon how comfortable various students feel—and how nurturing the environment might be for specific students. While the factory model helps to illuminate the importance of certain oft-studied objective measures (class size, teacher quality, etc…), the garden model helps to explain others, such as the importance of minority teachers for minority students (Hess and Leal 1997; Dee 2004). The garden model is consistent with potentially large effects of fit when properly identified and measured; importantly, it can also provide a framework for understanding why some students make large academic gains in non-traditional school settings, such as home-education.

d. Relation to policy and future research

With respect to policy, it is important to know what benefit, if any, accrues to parents and students when they aim to get a "good fit". Research has implicitly treated the question of fit as a matter of taste or individual utility. But if a good fit between student and schools *by itself* raises outcomes—independently of many other school or student characteristics, then the state too would reap the benefit of a good fit.

Studies of choice to date have not been able to distinguish between whether choice improves outcomes through competitive markets or through sorting. For example, Hoxby (2000) found that increased school choice improves student outcomes and lowers per pupil costs, thereby increasing overall efficiency. But her data do not allow her to back-out estimates of student-school interaction effects that would be needed to make a judgment about fit effects. Further, since her sample is limited to public schools, it's hard to think about meaningful student-school pairings in this setting. Without doubt there is some kind of sorting according to student-school interaction that occurs within and across residential districts; and if Tiebout (1956) is correct, this sorting will be efficient. But this sorting has been understood to have more to do with preferences for expenditures and not "fit" in the personal sense, as described in this paper.

Identifying the effects of a good fit in education ultimately calls for at least three new avenues for research. First, it would be of great value to attempt to replicate these findings in better data, and to tease out the differences between school-effects and fit-effects (student-school specific). The findings in this paper support the general spirit of earlier findings which see particular effects for particular student-school groupings. At the same time, more attention needs to be paid in understanding the mechanisms behind certain interaction effects. Why do Catholic schools appear to confer greater benefits on some students than on others? Answering that question may help identify more general effects which may be repeatable outside of the Catholic school setting.

Second, alternative ways of observing and measuring fit might be explored—especially models of fit which allow for gradients instead of the binary measure used here. Recent work on models of peer effects may provide a platform for thinking about variable degrees of fit (Hoxby and Weingarth 2005). Certainly, peer effects models help answer what degree of fit between a student and her peers is most beneficial to growth in scores. One generalization from this work is that students do seem to benefit from a healthy dose of "sameness" or "fit". Too much diversity in scores—large differences between the top and the bottom—seems to have deleterious effects.

Finally, future work may obviously extend the logic of this paper to other publicly provided social services for which a "good fit" might matter—such as health care, welfare, job training, substance abuse programs, and the broader class of public goods which may be characterized as social investments into human persons. The basic question remains the same: whether goods that aim to form (or transform) human beings can be expected to perform equally well for all human beings. While most would say that such goods have an objective component and a subjective component, it is clear that for reasons of expediency research has been almost entirely focused on identifying the objective component and its covariates. These findings, however, and all those which have found significant interaction effects, suggest that the subjective components may be much larger than previously thought. But this is good news: effects of any sort that are larger than expected means that better reforms may be within reach.

NOTES

1 Paragraph 53A-1a-802 of Utah 2007 H.B. 148

2 According to one reporter describing the Obamas choice of school in Washington, D.C., "A number of great schools were considered. In the end, the Obamas selected the school that was the best fit for what their daughters need right now." (Leiby 2008)

3 This paper uses the language of "fit" and "match" somewhat interchangeably, but in general, "fit" is conceived as the true continuous variable, and "match" is conceived of as a binary approximation of fit.

4 Tables 35 and 71 from the 2012 NCES Digest of Education Statistics reports that in 2007-8 the average tuition for a Catholic school student was $6,018, and for a non-Catholic religious school student the same figure was $7,117. Meanwhile, the average tuition at a non-sectarian private school was $17,316, and public schools spent on average $12,759 per pupil (NCES 2012).

5 Note that the NLSY97 oversampled minority populations. I do not use weights in any of my analyses, since I do not make proportional claims from my analysis to the general population.

REFERENCES

Akerlof, G. A., and Kranton, R. E. 2002. Identity and Schooling: Some lessons for the economics of education. *Journal of Economic Literature,* 40 (4):1167-1201.

Altonji, J. G., Elder, T. E., and Taber, C. R. 2005. An evaluation of instrumental variable strategies for estimating the effects of Catholic schooling. *Journal of Human Resources,* 40 (4):791-821.

Angrist, J. D., and Krueger, A. B. 2001. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives* , 15 (4):69-85.

Angrist, J. D., and Pischke, J.-S. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Bordin, E. 1979. The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice,* 16:252-260.

Bound, J., Jaeger, D. A., and Baker, R. M. 1995. Problems with Instrumental Variables Estimation When the Correlation Between the. *Journal of the American Statistical Association,* 90 (430): 443-450.

Chaves, M. 1994. Secularization as Declining Religious Authority. *Social Forces,* 72 (3):749-774.

Coleman, J. S., and Hoffer, T. 1987. *Public and private high schools: The impact of communities.* New York: Basic Books.

Coleman, J., Hoffer, T., and Kilgore, S. 1982. Cognitive outcomes in private and public schools. *Sociology of Education,* 55 (2):65-76.

Dee, T. S. 2004. Teachers, Race, and Student Achievement in a Randomized Experiment. *The Review of Economics and Statistics,* 86 (1):195-210.

________. 2005. The Effects of Catholic Schooling on Civic Participation. *International Tax and Public Finance,* 12 (5):605-625.

Eide, E. R., Goldhaber, D. D., and Showalter, M. H. 2004. Does Catholic High School Attendance Lead to Attendance at a More Selective College? *Social Science Quarterly ,* 85 (5):1335-1352.

Elder, T., and Jepsen, C. 2014. Are Catholic primary schools more effective than public primary schools? *Journal of Urban Economics,* 80 (1):111-132.

Epple, D., and Romano, R. E. 1998. Competition between Private and Public Schools, Vouchers, and Peer-Group Effects. *The American Economic Review ,* 88 (1):33-62.

Evans, W. N., and Schwab, R. M. 1995. Finishing high school and starting college: Do Catholic schools make a difference? *Quarterly Journal of Economics,* 110 (4): 941-974.

Figlio, D. N., and Stone, J. A. 1999. Are Private Schools Really Better? *Research in Labor Economics,* 18:115-140.

Grogger, J., and Neal, D. 2000. Further Evidence on the Effects of Catholic Secondary Schooling. *Brookings-Wharton Papers on Urban Affairs*, pp. 151-201.

Hess, F. M., and Leal, D. L. 1997. Minority teachers, minority students, and college matriculation: A new look at the role-modeling hypothesis. *Policy Studies Journal,* 25: 235-248.

Hoffer, T., Greeley, A. M., and Coleman, J. S. 1985. Achievement growth in public and Catholic schools. *Sociology of Education,* 58 (2):74-97.

Horvath, A. O., and Symonds, B. D. 1991. Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology,* 38, 139-149.

Hoxby, C. M. 1994. Do private schools provide competition for public schools? *NBER Working Papers,* 4978.

________. 2000. Does competition among public schools benefit students and taxpayers? *The American Economic Review,* 90 (5):1209-1238.

________. 2002. Would school choice change the teaching profession? *Journal of Human Resources,* 37 (4):846-891.

Hoxby, C., and Weingarth, G. 2005. Taking Race out of the Equation: School Reassignment and the Structure of Peer Effects. *Working Paper.*

Hurst, G. A., Williams, M. G., King, J. E., and Viken, R. 2008. Faith-based intervention in depression, anxiety, and other mental disturbances. *Southern Medical Journal,* 101 (4):388-392.

Kim, Y.-J. 2011. Catholic schools or school quality? The effects of Catholic schools on labor market outcomes. *Economics of Education Review,* 30 (3):546-558.

LaVigne, N. G., Small, K., and Brazzell, D. 2007. *Evaluation of the Florida department of correction's faith and character-based institutions.* Washington DC: Urban Institute Justice Policy Center.

Lee, V., Bryk, A., and Holland, P. B. 1993. *Catholic schools and the common good.* Cambridge, MA: Harvard University Press.

Martin, D. J. 2000. Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology,* 68, 438-450.

McClellan, M., McNeil, B. J., and Newhouse, J. P. 1994. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *Journal of the American Medical Association,* 272 (11):859-866.

McDonald, D., and Schultz, M. 2008. *The annual statistical report on schools, enrollment and staffing: United States Catholic elementary and secondary schools 2007-2008.* Arlington: National Catholic Educational Association.

Murray, M. P. 2006. Avoiding Invalid Instruments and Coping with Weak Instruments. *The Journal of Economic Perspectives,* 20 (4):111-132.

Neal, D. 1997. The effects of Catholic secondary schooling on educational achievement. *Journal of Labor Economics,* 15 (1):98-123.

Noell, J. 1982. Public and Catholic schools: A reanalysis of "public and private schools". *Sociology of Education,* 55 (2):123-132.

Sloan, F. A., Picone, G. A., Taylor, D. H., and Choue, S.-Y. 2001. Hospital ownership and cost and quality of care: is there a dime's worth of difference? *Journal of Health Economics,* 20 (1):1-21.

Stock, J. H., Jonathan, W. H., and Yogo, M. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *American Statistical Association,* 20 (4):518-529.

Taylor, R. J., Chatters, L. M., Jayakody, R., and Levin, J. S. 1996. Black and White Differences in Religious Participation: A Multisample Comparison. *Journal for the Scientific Study of Religion,* 35 (4):403-410.

Taylor, R. J., Mattis, J., and Chatters, L. M. 1999. Subjective Religiosity among African Americans: A Synthesis of Findings from Five National Samples. *Journal of Black Psychology,* 25 (4):524-543.

The Association of Religious Data Archives. n.d. *American denominations: Family trees*. Retrieved from The association of religious data archives: http://www.thearda.com/Denoms/Families/trees.asp

Tiebout, C. M. 1956. A Pure Theory of Local Expenditures. *Journal of Political Economy,* 64 (5):416-424.

Trivitt, J. R., and Wolf, P. J. 2011. School Choice and the Branding of Catholic Schools. *Education Finance and Policy,* 6 (2):202-245.

Tyler, B. S. 1994. An analysis of public and Catholic secondary education and the earnings of men.

United States Department of Education Institute of Education Sciences. 2012. *Digest of education statistics: 2012 digest tables directory*. Retrieved from National Center for Educational Statistics: http://nces.ed.gov/programs/digest/d12/tables/dt12_071.asp

United States Department of Labor. 2012. *National longitudinal survey of youth 1997*. Retrieved from Bureau of Labor Statistics: http://www.bls.gov/nls/nlsy97.htm

Witte, J. F. 1996. School choice and student performance. In: H. F. Ladd (Ed.), *Holding Schools Accountable: Performance-Based Reform in Education* (pp. 149-165). Washington, DC: Brookings Institution Press.

Woodberry, R. D., Park, J. Z., Kellstedt, L. A., Regnerus, M. D., and Steensland, B. 2012. The Measure of American Religious Traditions: Theoretical and Measurement Considerations. *Social Forces,* 91 (1):65-7.