Educational Evaluation and Policy Analysis
March 2008, Vol. 30, No. 1, pp. 3–30
DOI: 10.3102/0162373707313781
© 2008 AERA. http://eepa.aera.net

Does NBPTS Certification Affect the Number of Colleagues a Teacher Helps With Instructional Matters?

Kenneth A. Frank
Gary Sykes
Dorothea Anagnostopoulos
Marisa Cannata
Linda Chard
Ann Krause
Raven McCrory
Michigan State University

In addition to identifying and developing superior classroom teaching, the National Board for Professional Teaching Standards (NBPTS) certification process is intended to identify and cultivate teachers who are more engaged in their schools. Here the authors ask, "Does NBPTS certification affect the number of colleagues a teacher helps with instructional matters?" If so, this could enhance the influence of NBPTS-certified teachers and their contributions to their professional communities. Using sociometric data within 47 elementary schools from two states, the authors find that NBPTS-certified teachers were nominated more as providing help with instruction than non-NBPTS-certified teachers. From analyses using propensity score weighting, the authors then infer that NBPTS certification affects the number of colleagues a teacher helps with instructional matters. The authors then quantify the robustness of their inference in terms of internal and external validity, finding, for example, that any omitted confounding variable would have to have an impact six times larger than that of their strongest covariate to invalidate their inference. Therefore, the potential value added by NBPTS-certified teachers as help providers has policy and practice implications in an era when teacher leadership has risen to the fore as a critical force for school improvement.

Keywords: *National Board Certification, causal inference, teacher help*

Introduction

The study reported here has a dual significance in addressing a question of import for educational policy and practice and demonstrating several cutting-edge analytic methods for making causal inferences. The substantive questions of interest concern whether teachers certified by the National Board for Professional Teaching Standards (NBPTS) provide more help to their colleagues with instructional matters than do non-NBPTS-certified teachers and whether such provision can be attributed to NBPTS certification. To answer these questions, we employ methodological

An earlier version of this article was presented at the annual meeting of the American Educational Research Association, San Francisco, California, April 10, 2006. The authors wish to thank the National Board for Professional Teaching Standards for their generous support of this research. Opinions expressed are solely the authors'.

advances concerning ways to estimate effects and quantify the validity and generalizability of causal inferences from observational data.

Estimates of total public and private expenditures on NBPTS certification include some $600 million in grants and fees together with $1 billion in salary incentives across the 50 states and 544 districts that offer such bonuses (Podgursky, 2001). Such investment naturally raises the question of whether these expenditures may be justified on the basis of the NBPTS's impact and outcomes (Boyd & Reese, 2006).

One way to calculate the impact of NBPTS-certified teachers is to ask whether they enhance the achievement of their students relative to comparable colleagues who are not NBPTS certified. If this were found to be the case, then NBPTS certification might be an important policy tool for identifying and utilizing highly accomplished teachers. Such a finding would have broad implications for such matters as teacher licensure, teacher evaluation, performance-based pay, professional development, selection for advanced positions, and deployment of teaching talent in schools. Therefore, research that considers the question of NBPTS-certified teachers' impact on student achievement is reviewed in the next subsection.

Beyond their direct effects in the classroom, the NBPTS also emphasizes the teacher's role in the larger school community. The proposal that spurred the establishment of the NBPTS called for teachers who "provide active leadership in the redesign of the schools and in helping their colleagues to uphold high standards of teaching and learning" (Carnegie Forum on Education and the Economy, 1986, p. 55), while the fifth proposition of NBPTS standards and assessments states, "Teachers are members of learning communities. Accomplished teachers contribute to the effectiveness of the school by working collaboratively with other professionals on instructional policy, curriculum development and staff development" (NBPTS, 1999, p. 31). Corresponding to this declaration of standards, the portfolio entries (submissions required as part of the certification process) include documentation of or reflections on one's interactions with colleagues and their impact on the school community (NBPTS, 1999).

The NBPTS's emphasis on help provided is supported by a recent general emphasis on social processes in schools (e.g., Bidwell, 2000, 2001). Most directly, help can contribute to the diffusion of the techniques that can improve achievement (Cavalluzzo, 2004; Goldhaber & Anthony, in press). This could range from pedagogy to modes of reflection. Furthermore, through the interaction associated with help, teachers can contribute to a norm of helping (Glidewell, Tucker, Todt, & Cox, 1983), through which expertise and resources can be distributed throughout a school (Spillane, 2006). Finally, the help and the resultant norm of providing help can contribute to the store of social capital on which teachers can draw to improve and innovate (Frank, Zhao, & Borman, 2004). Indeed, the provision of help is one of the long-sought components of professional community (Louis, Marks, & Kruse, 1996) that have been linked to organizational effectiveness (see, e.g., Bryk & Schneider, 2002; Bryk, Sebring, Kerbow, Rollow, & Easton, 1998).

Economists might phrase the specific help provided by NBPTS-certified teachers as "spillovers" or "positive externalities" associated with the presence of NBPTS-certified teachers in schools. These spillovers can be generally critical to production (Romer, 1990), and in particular, knowledge transfer has been linked to organizational performance through transfers of best practices (Szulanski, 1996), new product development (Hansen, 1999), and learning rates (Argote, Beckman, & Epoie, 1990). In fact, firms can be characterized by their capacity to contain and convey knowledge independent of market mechanisms (Arrow, 1974; Kogut & Zander, 1992; see Reagans & McEvily, 2003, for a review).

Not surprisingly, spillovers constitute a key aspect of the value of NBPTS certification that extends beyond individual classrooms to expert knowledge that circulates among teachers in a school. In this account, NBPTS certification holds potential for becoming a social resource directed to instructional improvement, enhancing the total stock of professional knowledge possessed not by teachers singly but by the school as a collective.

Motivating policy interest in NBPTS certification, then, are issues of cost and effectiveness but also a larger aspiration, expressed by many advocates for professionalism in teaching, to create a stronger leadership presence and structure for teachers. The policy hypothesis is that improved use of master teachers, as identified by NBPTS certification, has payoff for important long-term goals that include attracting, cultivating, and retaining talent in teaching; targeting recruitment of highly qualified teachers to high-need schools; filling emerging leadership-related positions in schools and districts; and mobilizing teacher expertise for school improvement.

Before we explore the potential of NBPTS-certified teachers to help others in their schools, we first review the evidence on the effects of NBPTS-certified teachers on student achievement. If NBPTS-certified teachers are not exceptionally effective in producing student achievement, then their helping behavior might actually be negligible, even counterproductive, to the extent they are supplying faulty guidance based on their own relatively ineffective practice.

Effects of NBPTS-Certified Teachers on Student Achievement

Evidence of the effects of NBPTS-certified teachers on student achievement currently yields a mixed record. In an early validation study, Bond, Smith, Baker, and Hattie (2000) found clear effects associated with certification on teaching self-reports and some student products. More recent studies using statewide student achievement data from North Carolina and Florida have confirmed that certification distinguished effectiveness over and above licensure status, with effects stronger in some subjects than in others (Clotfelder, Ladd, & Vidgor, 2007; Goldhaber & Anthony, in press; Harris & Sass, 2007). These statewide studies also raised questions about the possibility of cohort effects (with early cohorts showing stronger effects on achievement than later ones) and of the possibility that NBPTS effects on achievement might vary over time rather than being a constant increment above the results produced by non-NBPTS teachers. Finally, in one well-designed study based on a district rather than state sample, Sanders, Ashton, and Wright (2005) used HLM techniques to specify the effect of NBPTS certification at the level of the teacher (instead of the student) and found no significant differences between NBPTS-certified teachers and teachers who never applied, teachers who intended to apply, and teachers who applied but were not successful in becoming NBPTS certified.

Taken together, these studies do not yield firm, unambiguous conclusions. Indeed, the studies raise questions about the duration of the positive effects on student achievement of NBPTS certification. On some issues, however, evidence seems to converge. Most important to our own question, the certification process at minimum appears to identify teachers who are more effective prior to certification. It is then informative for policy if the NBPTS certification process affects how many others NBPTS-certified teachers help in their schools because the help that NBPTS-certified teachers provide their colleagues can have important direct and indirect effects on instruction and achievement.

Does NBPTS Certification Affect the Number of Colleagues a Teacher Helps With Instructional Matters?

There are preliminary answers to our question in existing literature. Interviews in Bond et al. (2000) showed that ". . . with rare exception they [NBPTS-certified teachers] have not noticed an increase in the use of their expertise since obtaining NBPTS certification" (p. 142). Harris and Sass (2007) indirectly tested for spillover effects, examining how the number of NBPTS-certified teachers in a school influenced the student achievement results of non-NBPTS-certified teachers in those schools. They found negligible to slightly negative effects, depending on the subject area, but found that results for students of non-NBPTS-certified teachers with NBPTS-certified teacher mentors were better on the FCAT-NRT, the norm-referenced version of the Florida state achievement test in reading and math, but not for the FCAT-SSS, the criterion-based version of the state achievement test. Given the lack of direct specifications of the mechanisms of spillover via teacher communication and collaboration in both studies, however, any causal links between the presence of NBPTS-certified teachers and the performance of others in their schools are not currently well understood.

In our effort to assess spillover effects of NBPTS-certified teachers through helping, we directly take on several previous methodological hurdles. To begin, to find out how helpful NBPTS-certified teachers are, we do not rely on reports from the NBPTS-certified teachers, as accurate assessments of the value of help more likely come from the recipient—not the provider—of help (Frank et al., 2004). Thus, we draw on sociometric data collected from entire rosters of each school, asking teachers to indicate those who were helpful to them. Anticipating our most basic result, NBPTS-certified teachers were nominated as helpful with instructional matters by about 1.5 colleagues, while non-NBPTS-certified teachers were nominated by about .9 colleagues, a difference of about .6.[1]

Given this descriptive result, we then consider whether the difference can be attributed to the experience of NBPTS certification. That is, does NBPTS certification affect the number of colleagues a teacher helps with instructional matters? To address this, we must engage three concerns regarding causal and statistical inference. First, because our goal is to inform policy, we use propensity score weighting to focus estimation on effects for those teachers most likely to respond to changes in policy. The policies might invoke state and school incentives for teachers to become certified, changes in the resources available to the NBPTS for certification, and changes in the certification process that could attract different numbers or types of teachers to pursue certification. But all teachers would not respond equally to policies related to certification. Teachers with extremely high propensity for pursuing NBPTS certification would likely pursue it or similar professional development regardless of incentives and resources for doing so. Conversely, teachers with extremely low propensities for pursuing NBPTS certification would be unlikely to pursue it or related professional development regardless of the incentives and resources for doing so. Therefore, our goal is to evaluate the effect of being NBPTS certified for those most likely to be responsive to incentives and policies to support NBPTS certification. Our corresponding counterfactual (Holland, 1986; Rubin, 1974) question is, For the type of teacher likely to respond to incentives to become NBPTS certified, how many colleagues does an NBPTS-certified teacher help in her school compared to how many she would have helped if she were not an NBPTS-certified teacher? This question is counterfactual because we cannot simultaneously observe a single teacher as an NBPTS-certified teacher and a non-NBPTS-certified teacher. As an approximation to the counterfactual, we will use propensity weighting to compare those who became NBPTS certified but who had low propensity for doing so (and may have responded to outside incentives) with those who were not NBPTS certified but had high propensity for doing so (and may respond to new incentives).

Second, absent random assignment to NBPTS certification (which would be fraught with logistical and ethical complications; see Rubin, 1974; Shadish, Cook, & Campbell, 2002), we rely on statistical control, in this case embedded within the propensity score, to achieve comparability between NBPTS-certified teachers and non-NBPTS-certified teachers. But the statistical control is only as good as the covariates used to construct the propensity scores (Heckman, 2005; Morgan & Harding, 2006; Rosenbaum, 2002; Shadish et al., 2002); there may be important omitted confounding variables. Therefore, we will characterize the robustness of our inference to the potential impact of confounding variables (Frank, 2000).

Third, our sample was drawn in 2003 from only two states; any results we find may not generalize to teachers in other cohorts and states. For example, the effect of NBPTS certification may be especially strong in states where there are few other alternative experiences that might cultivate more helpful teachers. The question is, how robust are inferences regarding the effect of NBPTS certification across contexts? To address this question, we will employ Frank and Min's (2007) analysis to quantify the conditions necessary to invalidate an inference if the sample were recomposed to be more representative of a given target population.

Thus, this article has two integrated purposes. The first is to answer the question of whether NBPTS certification affects the number of colleagues a teacher helps with instructional matters in her school. The answer to this question has important value for scientific and policy purposes. But in answering this question, we must employ several new techniques in the estimation and interpretation of effects from observational data. Thus, the second purpose is to call attention to methodological issues. Critically, the methodological issues are phrased not in statistical terms but in the scientific terms of alternative explanations for the findings and composition of the sample. Indeed, it is these scientific issues that give meaning to the quantitative expressions of robustness we employ.

Data and Method

This study is based on a survey of the full faculty in each of 47 elementary schools in two states. The states were chosen because of their relatively high proportion of NBPTS-certified teachers and their strong incentives for teachers to become NBPTS certified. Within states, we stratified our sample based on urbanicity, recognizing typical differences in educational processes and outcomes across this dimension. We began by sampling two urban districts in each state that had strong incentives and policies supporting NBPTS certification. In State A, we then sampled schools from one neighboring, nonurban companion district for each urban district. In State B, with smaller district sizes and fewer numbers of NBPTS-certified teachers, we sampled schools from multiple nearby nonurban districts for each urban district.

Once the districts were selected, a database including information on the school where each NBPTS-certified teacher (as of fall 2003) was located was obtained from the NBPTS. Schools with large proportions of NBPTS-certified teachers were then oversampled to ensure enough data to estimate the effects of NBPTS certification. We included 2 schools in each state that had no NBPTS-certified teachers to be sure to represent a range of conditions in our sample. Six schools in each urban district were included (a case study school, a school with no NBPTS-certified teachers, and 4 additional schools with varying numbers of NBPTS-certified teachers) as well as 6 schools in each neighboring district or group of neighboring districts (1 school with no NBPTS certified teachers and 5 other schools with varying numbers of NBPTS certified teachers), for a total of 48 elementary schools. One school declined to participate after data collection began, for a final sample of 47 schools.

A survey was administered in each school during regularly scheduled staff meetings. Researchers administered the survey in some schools, with local personnel distributing the survey in others. All schools were given $25 toward the purchase of refreshments for the staff meeting during which the surveys were distributed. Schools that participated only in the survey were given an additional $125 as an incentive. Case study schools were compensated with an additional $375. A total of 1,583 surveys were completed with an average school response rate of 84% per school.

The surveys included Likert response items regarding teacher attitudes, background information, and questions regarding conditions for NBPTS-certified teachers as well as each teacher's particular engagement with and perceptions of NBPTS certification. Critically, the surveys also included a sociometric question asking each teacher to indicate the other teachers (and other school actors) who were helpful with instruction.

Because our focus is on comparing NBPTS-certified teachers with similar others and because teachers can be eligible for NBPTS certification only after teaching in a classroom for at least 3 years, our analytic sample included only those teachers who had completed more than 3 years of teaching and who reported being either full- or part-time classroom teachers. This simple filter removed 375 non-NBPTS-certified teachers from the analysis. Furthermore, all but 2 of the NBPTS-certified teachers indicated in which grade levels they taught, their gender, and level of education. Thus, the 52 non-NBPTS-certified teachers who did not indicate information on one of these variables were removed from the analyses because their extent of comparability with the NBPTS-certified teachers could not be fully evaluated. Similarly, all NBPTS-certified teachers responded to questions about the perceived advantages of NBPTS certification. Therefore, those 23 non-NBPTS-certified teachers who did not respond to these questions were removed from the analysis. Our final sample size was 1,131.[2] We will quantify the robustness of our inferences to our sample filters in the results section.

Measures

Dependent variable. To construct our dependent variable, we asked each teacher in the school to nominate the other teachers who had been helpful with instruction. The dependent variable is then simply the total number of other teachers who nominated a teacher as helpful. Thus, if Lisa were nominated as helpful by Joe, Sue, and Lucy, then Lisa's value would be 3, because she was nominated by three other teachers. As simple as this is, we emphasize the importance of obtaining our measure from the recipients of help rather then the help providers (Frank et al., 2004). This holds because knowledge, especially the complex knowledge needed for teaching, likely has been transferred only if the recipient indicates such, regardless of reports of those who originally possess knowledge of attempts to transmit knowledge (Hansen, 1999).

Predictor of interest. Our focal predictor is whether a teacher indicated being NBPTS certified based on the question, Are you certified by the National Board for Professional Teaching Standards? Our predictor takes a value of 1 if NBPTS certified, 0 otherwise.

Selection Into NBPTS Certification

One of the great advantages of the propensity score approach is that it forces researchers to explicitly consider and model assignment to treatment (cf. Heckman, 2005; Rubin, 2004). To begin, we consider that teachers may be motivated to pursue NBPTS certification for instrumental reasons (e.g., additional stipends for becoming an NBPTS-certified teacher). Although we have some information regarding the incentives offered by each district from our interviews and surveys, there were formal and informal nuances between schools that we did not directly measure. Therefore, we account for differences in formal incentives and informal support for becoming an NBPTS-certified teacher across schools by controlling for schools using fixed effects (i.e., including a set of dummy variables representing schools in our models; see Wooldridge, 2002).

Teachers also may be motivated to participate in NBPTS certification through their identity as members of their profession in general and with members of their school in particular (Akerlof & Kranton, 2005; Talbert & McLaughlin, 1994). We measured motivation via identification with the general profession in terms of the extent to which the teacher perceived that involvement in leadership enhanced teaching as a profession. This was a composite of three items (4-point scale, *strongly disagree* to *strongly agree*; alpha = .89):

> My involvement in leadership activities makes me feel more significant in my profession.
> My involvement in leadership activities makes me feel like teaching has a lot to offer me.
> My involvement in leadership activities enhances my career satisfaction.

Approximately 16% (including 11 NBPTS-certified teachers) of our respondents had missing data on this variable, for whom we set the value to 0 and included a flag in our primary analyses indicating we had done so (Cohen & Cohen, 1983).

Teachers may also identify with their immediate colleagues representing the specific social system of the school (Bidwell, 2000; 2001). In particular, teachers may pursue NBPTS certification to retain standing and advance themselves within their school (Blau, 1967; Burawoy, 1979).[3] We measured motivation in terms of the perceived advantages of NBPTS-certified teachers in the school using a composite of three items (4-point scale, *strongly disagree* to *strongly agree*; alpha = .82):

> The principal includes NBPTS-certified teachers more than other teachers in school leadership.
> The principal encourages NBPTS-certified teachers more than other teachers to share ideas and innovations.
> NBPTS-certified teachers in our school are treated better than teachers who are not NBPTS certified.

As members of the school as a social organization, teachers may also respond to the norms of their school. In particular, teachers who received more help from others may be more likely to become NBPTS-certified teachers either because they are more motivated to excel or because the help enables them to satisfy the requirements of certification. Therefore, we measured the number of other teachers each teacher reported as helping her. Less directly, teachers who work in schools with many other NBPTS-certified teachers also may be more inclined to become NBPTS-certified teachers through a purely normative effect. Therefore, we measured the number of other teachers (besides the respondent) in the school who were NBPTS-certified teachers (cf. Levine and Painter, 2003).[4]

We also reasoned that teachers would reap more of the benefits of NBPTS certification the longer they planned to remain in the profession. Therefore, we measured each teacher's intent to leave teaching based on the response to the question, How long do you plan to continue teaching? (The responses were treated categorically: as long as I am able; until I am eligible for retirement; continue unless something better comes along; leave teaching as soon as I can; undecided at this time). The 16 cases (including 3 NBPTS-certified teachers) with missing values on intent to leave teaching were assigned to a separate category for our primary analyses.

Of course, a teacher's motivation for engaging in any work-related behavior may be a function of the teacher's relevant background and location in the division of labor. Therefore, we controlled for a teacher's gender, race (White or not), level of education (dummy variables indicating a bachelor's degree, bachelor's degree plus additional hours, a master's degree, master's degree plus additional hours, or a doctorate), number of years teaching in the school, and highest grade level at which the teacher taught (1 = *pre-K* through 14 = *12th grade*).

The means and standard deviations for each of our variables for teachers in our sample are listed in Table 1. Teachers on average helped one other teacher in the school. Note that the standard deviation for "help provided" of 1.09 implies a variance of 1.15, a slight but not extensive overdispersion (in a theoretical Poisson model of a dependent variable that is a count, the mean and variance are equal—here the variance is slightly greater than the mean). Overall, approximately 14% (or 160 out of 1,131) of the teachers in our sample were NBPTS-certified teachers.

Regarding motivation via identification with the profession or school, teachers were close to neutral on items linking leadership with an enhancement of teaching and tended to disagree with statements indicating strong advantages for being NBPTS certified. Hence, the traditional leveling assumptions described by Lortie (1975) among others continue to apply, at least with reference to teachers' nominal sentiments. Regarding the school norms, teachers reported receiving help with instruction from slightly less than 1 other on average, and there were 2.3 other teachers in a given school who were NBPTS certified. Furthermore, most teachers indicated high commitment to the profession (with the modal teacher intending to teach as long as he/she is able).

Regarding background, approximately 93% were female, and 85% of our teachers were White. The modal teacher had completed a bachelor's degree plus additional hours, and the average teacher had been teaching for about 16 years and taught between sixth and seventh grade.

Analytic Approach

Potential challenges to causal inferences demand that we employ two important new methods in the social sciences. First, we will use propensity score techniques to focus on an estimate of NBPTS certification for those most likely to be responsive to changes in incentives regarding certification. Second, we will use Frank's (2000) and Frank and Min's (2007) indices to quantify the robustness of our inference.

Propensity score techniques. Propensity score techniques function in the spirit of the counterfactual (Morgan & Harding, 2006) by facilitating direct comparisons between individuals in a "treatment group" (e.g., the NBPTS-certified teachers) and those in a "control group" (e.g., non-NBPTS-certified teachers). As such, propensity scores allow researchers to relax the typical assumptions of regression concerning linear relationships among the dependent and independent variables and of homogeneous effects (Morgan & Harding, 2006).

The propensity score, $e(x)$, of receiving the treatment given the covariates is defined as follows (Rosenbaum & Rubin, 1983, p. 42):

$$\mathrm{e}(x) \equiv \Pr\{t = 1|x\}, \qquad (1)$$

where t indicates whether a subject received the treatment, in our case being NBPTS certified, and $\Pr(t = 1|x)$ is the probability of receiving the treatment given the covariates, x.[5] As is typical, we will use a logistic regression to estimate the propensity that a teacher became NBPTS certified given covariates x and then use the propensity to define comparisons between those who

TABLE 1

Descriptive Statistics for Variables Related to NBPTS Certification

Variable	Mean	Count	Standard Deviation
Number of others helped with instruction	.98		1.09
NBPTS certified	.14		.35
Enhancement of teaching through leadership (n = 850) (scale 1–4)	2.38		1.18
Missing on enhancement of teaching through leadership	.16		.36
Perceived advantage of certification (scale 1–4)	1.95		.56
Number of other teachers who helped respondent with instruction	.92		.77
Number of other NBPTS-certified teachers in school	2.36		2.44
Number of other NBPTS-certified teachers in school, squared	6.47		11.76
Plans to leave teaching:			
As long as I am able		504	
Until I am eligible for retirement		341	
Continue unless something better comes along		53	
Leave teaching as soon as I can		14	
Undecided at this time		203	
Missing		16	
Female	.93		.25
White	.85		.36
Level of education:			
Bachelor's degree		79	
Bachelor's degree plus additional hours		348	
Master's degree		202	
Master's degree plus additional hours		45	
PhD		7	
Years teaching	15.99		8.62
Highest grade level taught	8.64		3.84

Note. n = 1,131 unless otherwise specified. NBPTS = National Board for Professional Teaching Standards.

did and did not become NBPTS-certified (Morgan & Harding, 2006). One way to use propensity scores in our case would be to match each NBPTS-certified teacher with one or many teachers close to her in terms of the propensity of becoming an NBPTS-certified teacher and then analyze differences on outcomes (e.g., number of nominations of help provided) between the matched pairs (Morgan, 2001; Morgan & Harding, 2006). But recent critiques of this approach note that considerable data are lost in the matching—analyses are only of those in the treatment and their matches, removing all those who were not in the treatment or who were not matched (Hirano & Imbens, 2001; Hirano, Imbens, & Ridder, 2003; Morgan & Harding, 2006; Robins, 1987; Robins, Hernán, & Brumback, 2000; Robins & Rotnitzky, 1995; Robins, Rotnitzky, & Scharfstein, 2000).

Robins and colleagues (Robins, 1987; Robins, Hernán, & Brumback, 2000; Robins & Rotnitzky, 1995; Robins, Rotnitzky, & Scharfstein, 2000) suggest an elegant alternative use of propensity scores that preserves many of the benefits of matching while making full use of the data and improving efficiency in estimation. In particular, they suggest conducting a simple regression of outcome on treatment, using weights of

$$\omega(t, x) = \frac{t}{e(x)} + \frac{1-t}{1-e(x)}. \qquad (2)$$

Thus, in our example, a teacher who is NBPTS certified is weighted by $\frac{1}{e(x)}$ and a teacher who is not NBPTS certified is weighted by $\frac{1}{1-e(x)}$. As a result, NBPTS-certified teachers would be weighted more, the lower their propensity of being certified, while non-NBPTS-certified

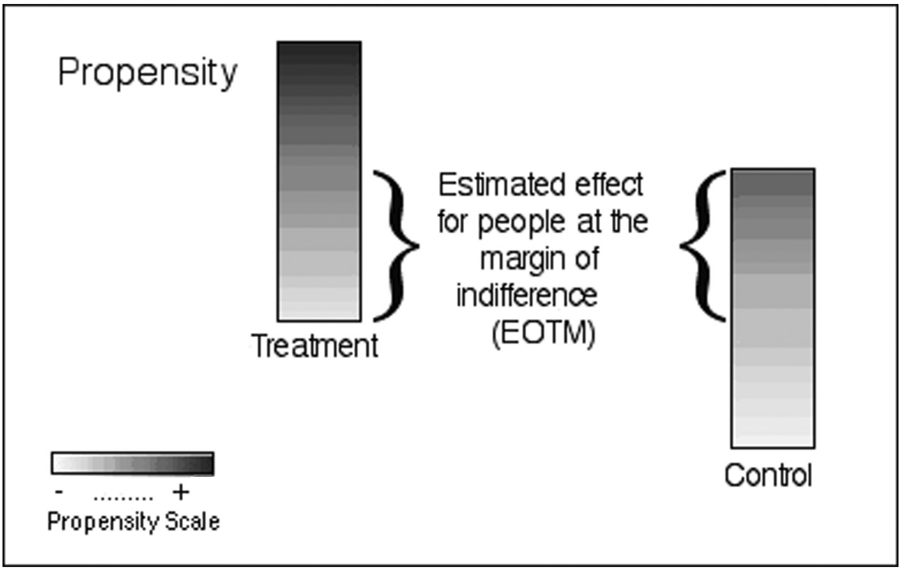
Propensity
Treatment
Estimated effect for people at the margin of indifference (EOTM)
Control
- .... +
Propensity Scale

FIGURE 1. *Overlap in propensity between treatment and control groups.*

teachers would be weighted more, the higher their propensity for being NBPTS certified. As such, the comparison focuses on the strongest overlap (or support) in propensity: those who received the treatment yet had low propensity with those who did not receive the treatment but had high propensity (see Figure 1).

Critically, the weighting as in Equation 2 reflects the policy considerations described in the introduction. If policies focus on changes in incentives and resources for becoming NBPTS certified, then estimates of effects should focus on those most likely to respond to changes in policies: those who became NBPTS certified but who had low propensity for doing so and therefore might not have become certified if there were fewer incentives for pursuing certification, and those who did not become NBPTS certified but who had high propensity for becoming NBPTS certified and therefore might respond to increases in incentives for doing so. Thus, the estimate using the weights in Equation 2 is referred to as the effect of the treatment for people at the margin of indifference (EOTM) (Heckman, 2005).

The great advantages of the weighting approach are that (a) the weighting scheme is relatively simple and intuitive; (b) estimates using the weights are easy to obtain (e.g., using weighted least squares) and can be implemented within the context of simple or more complex models; (c) the standard errors can be calculated as in any weighted analysis; (d) because the estimand is a smooth function of the data (as in the weighted regression), bootstrapping techniques can be employed to calculate standard errors that reflect uncertainty in estimating the propensity; and (e) all subjects contribute to the analysis (though not equally, by definition). In fact, all matching estimators can be considered examples of weighting approaches (Morgan & Harding, 2006), but only Robins's approach has been proven to improve the efficiency of estimation using propensity scores (Hirano et al., 2003).

The greatest concern with the weighting approach is that extreme weights could exert undue influence on the estimates. This is easily addressed by examining the distribution of weights and trimming extreme values. In our study, a box plot of the weights indicated five extreme values ranging from 28 to 68 (relative to the median of 1.14 and the upper quartile defined by 1.32). We trimmed the extreme weights to a value of 20, which was 1 greater than the next most extreme weight in the data.

Although we emphasize the use of propensity weights to focus on the EOTM, they can also be used to differentiate among treatment effects. In particular, Hirano and Imbens (2001) note that Equation 2 can be easily modified to estimate the treatment effect for the treated:

$$\omega(t, x) = t + \frac{1 - t}{1 - e(x)} . \tag{3}$$

Using the weights in Equation 3, any person who received the treatment (i.e., was NBPTS certified) receives a weight of unity, whereas those who received the control is weighted as in Equation 2. Thus, the estimand compares all those who were NBPTS certified with those who were not NBPTS certified, where the latter group is weighted by propensity. As a complement, one can also estimate the treatment effect for the control group using weights $\omega(t, x) = \frac{t}{e(x)} + 1 - t$.

Once we obtain propensity scores, we use the propensity weights in evaluating whether the covariates are balanced between the NBPTS-certified teacher (treatment) and non-NBPTS-certified teacher (control) groups. When balance is achieved, one has adjusted for potential bias that could be attributed to the observed variables (Morgan & Harding, 2006). We first evaluate balance in terms of differences between NBPTS-certified teachers and others on our covariates once the weights are employed.[6] We also evaluate balance by quantifying the reduction in the impact of covariates (controlling for schools as fixed effects) on the treatment effect as a result of employing the weights.[7]

After establishing the balance of the covariates, we use several different models to estimate the effect of NBPTS certification on help provided to other teachers. First, using a general linear model, we estimate the effect of NBPTS certification on help provided, weighted by propensity as in Equation 2. Then, we compare with estimates from models employing weighting for the treatment effect for the treated (as in Equation 3), weighting for the treatment effect for the control, unweighted ordinary least squares (OLS) with covariates, and unweighted OLS with and without adjustments for schools. Furthermore, we explore whether the effect of NBPTS certification interacts with propensity. Because our outcome is a count variable with many (approximately 37%) teachers not receiving any nominations, we also confirm our results using a Poisson model with correction for overdispersion.[8]

When using Robins and colleagues' (Robins, 1987; Robins, Hernán, & Brumback, 2000; Robins & Rotnitzky, 1995; Robins, Rotnitzky, & Scharfstein, 2000) technique for weighting by propensity, the uncertainty in the estimates of the propensity scores should be taken into account in calculating standard errors. Although Hirano and Imbens (2001) provide asymptotic standard errors analytically, most current applications use bootstrap procedures to account for the uncertainty in the estimates of the propensity scores.[9] The bootstrap approach consists of repeatedly applying the estimation procedure, including the propensity score model and the final regression, to random samples (with replacement) from the original sample. This technique was employed here for each estimate of NBPTS certification on help provided, using 500 bootstrap replications of data sets with size equal to the input data set (e.g., if there were 1,131 observations in the original data set, then there were 1,131 observations in each replication). The standard errors we report were the standard deviations of the bootstrap estimates (the distributions of bootstrap estimates were all quite close to normal, with no concerns for heavy tails). *t* ratios were then constructed from the estimated effect in the full sample and its corresponding bootstrap standard error.[10]

To further narrow our research question, we note that a teacher must apply to be NBPTS certified and then becomes NBPTS certified only if she satisfies the portfolio requirements and passes the assessment procedures. Thus, there is selection into the application process and certification status, and we explore two comparisons. First, we compare NBPTS-certified teachers with other teachers who applied ($n = 280$, 160 of whom were NBPTS certified), thus removing concern about selection in the application process. But if NBPTS certification is a valid measure of teacher quality, then perhaps those who became NBPTS certified were different in important but unobserved ways from those who applied but were not NBPTS certified. Following this logic, we make a final comparison of NBPTS-certified teachers to only those who did not apply. For each of these latter two cases, we estimated separate propensity scores (for the reduced samples) and used the scores to obtain weighted regression estimates.

Instrumental variables. In estimating parameters in a general linear model, we allow for the possibility that each of the factors that predict the propensity to be an NBPTS-certified teacher could also directly affect the number of others a teacher helps in her school. This has high face validity for most of the measures (e.g., gender, years teaching, level of education, enhancement of teaching through leadership). One factor, however—perceived advantage of NBPTS certification—may be related to number of others helped only indirectly, mediated by whether a teacher becomes NBPTS certified. As such, it satisfies the exclusion restriction and could prove an ideal instrument (Heckman, 1978; Heckman & Robb, 1985; Winship & Morgan, 1999) for addressing concerns regarding selection bias or endogeneity in estimating the effect of NBPTS certification on help provided.[11] Unfortunately, perceived advantage of NBPTS certification was correlated only at .03 with whether a teacher was NBPTS certified. Thus, it hardly satisfies the criterion for an effective instrument, nor could we compensate with an extremely large sample size. The implication is that an instrumental variables estimate would have an extremely large standard error and be severely inconsistent (see Wooldridge, 2002, p. 102); in exploratory analyses not reported below, the standard errors were approximately 10 times larger using instrumental variables than when using other estimation procedures and estimates of the effect of NBPTS certification on number of others helped ranged from –10 to +10 depending on only slight modifications in the model. No other covariate theoretically satisfied the exclusion restriction (i.e., related to the provision of help only through whether a teacher became NBPTS certified). Therefore, we do not report results from models using instrumental variables.[12]

Quantifying the robustness of the inference. Having used propensity weighting to estimate the effect of NBPTS certification, we then turn to the robustness of our inference of the effect of NBPTS certification on the provision of help. The approach in this article can best be considered an extension of sensitivity analysis (cf. Copas & Li, 1997; Holland, 1989; Robins, Rotnitzky, & Scharfstein, 2000; Rosenbaum, 1986; Rosenbaum & Rubin, 1983; Scharfstein & Irizarry, 2003). Sensitivity analyses represent a set of possible estimates given a broad set of alternative conditions, helping researchers nuance interpretations and inferences. As in sensitivity analysis, we will consider how unknown quantities could affect estimates. But rather than reporting how violations of assumptions produce a range of estimates, we focus on exactly how much an assumption must be violated to invalidate an inference. As a result, the indices quantify the robustness of the original inference.

There have been important recent developments in quantifying the robustness of inferences. As one example, Rosenbaum (2002, chap. 4) extended Rosenbaum and Rubin (1983) to develop an index of the robustness of inferences to possible selection bias (see also Copas & Li, 1997). For matched cases, Rosenbaum (2002, p. 114) shows that, "to attribute the higher rate of death from lung cancer to an unobserved covariate *u* rather than to the effect of smoking, that unobserved covariate would need to produce a six-fold increase in the odds of smoking, and it would need to be a near perfect predictor of lung cancer." Diprete and Gangl (2004) apply Rosenbaum's approach to the use of instrumental variables (see also Altonji, Elder, & Taber, 2005; Lin, Psaty, & Kronmal, 1998), and Gastwirth, Krieger, and Rosenbaum (1998) extend the work of Rosenbaum and Rubin (1983) by expressing sensitivity in terms of the relationship between an unmeasured confounding variable and the outcome (called "dual" sensitivity analysis) as well as between the unmeasured confounding variable and the treatment (called "primal" sensitivity analysis).

Although the above robustness indices have put an important new spin on sensitivity analyses, there are two features that limit their accessibility for social scientists. First, confounding is defined by a relationship between the confound and the predictor of interest (e.g., NBPTS certification) and between the confound and the outcome (e.g., number of others helped) (Shadish et al., 2002). But the sensitivity analyses above treat each component separately.

Failure to incorporate both into sensitivity results in expressions such as "that unobserved covariate would need to produce a six-fold increase in the odds of smoking, *and* [italics added] it would need to be a near perfect predictor of lung cancer." Similarly, Diprete and Gangl (2004) write that "the effect of unemployment benefits on post-unemployment wage change remains statistically significant at the .05 level as long as the effect of the omitted variable on the outcome is below +.10 *and* [italics added] the correlation between the omitted variable and the treatment is less than .25" (p. 297). Each of these expressions requires the social scientist, as interpreter, to cognitively balance the relationship between the confounding variable with the predictor of interest as well as with the outcome. In the example from Rosenbaum, what would the association between the unobserved covariate and smoking have to be if the unobserved covariate were a good but not "near perfect" predictor of lung cancer? Rosenbaum's language and formulae place a high cognitive demand on the social scientist to obtain an answer, thus distracting from the scientific debate about the likelihood of observing and the nature of a confound that could invalidate the inference.

Approaches that focus on only one component of confounding also ignore the dependencies between the two components induced by the presence of a treatment effect. That is, when there is a nonzero treatment effect, covariates that are correlated with the outcome are more likely to be correlated with the predictor of interest. This will inflate the magnitudes of the impacts of the covariates on the effects of interest (Frank, 2000), and the distribution of impacts will be skewed (Pan & Frank, 2004a, 2004b), making sensitivity expressed in terms of one component even more difficult to interpret.

Second, many sensitivity analyses are expressed in terms of concordance or nonparametric statistics (e.g., Rosenbaum, 2002), when, in fact, social scientists often employ forms of the general linear model. In fact, the general linear model is so ubiquitous that expressions in terms of correlation, ranging from –1 to +1, have been internalized by most social scientists and can be used for calculations of power or effect size (e.g., Cohen & Cohen, 1983) as well as for other graphical interpretations of relationships (Rodgers & Nicewander, 1988).

To express robustness that simultaneously accounts for the relationship between a confounding variable and the predictor of interest and between the confounding variable and the outcome in terms relatively accessible to social scientists, we employ Frank and colleagues' (Frank, 2000; Pan & Frank, 2004) impact threshold for a confounding variable. Frank (2000) begins by defining the *impact* of a confounding variable on an estimated regression coefficient as $r_{v \cdot y} \times r_{v \cdot t}$, where $r_{v \cdot y}$ is the correlation between a covariate, *v*, and the outcome, *y*, and $r_{v \cdot t}$ is the correlation between *v* and *t*, a predictor of interest (for example, *t* is an indicator of whether a teacher is NBPTS certified; see Figure 2). Critically, the product $r_{v \cdot y} \times r_{v \cdot t}$ captures both the relationship between the confounding variable and the outcome and between the confounding variable and the predictor of interest. Moreover, it is through the *impact* that multiple regression adjusts for covariates as in the following expression for a correlation between *t* and *y*, partialling for *v*:

$$r_{t \cdot y \mid v} = \frac{r_{t \cdot y} - r_{v \cdot y} \times r_{v \cdot t}}{\sqrt{1 - r^2_{v \cdot y}}\,\sqrt{1 - r^2_{v \cdot t}}}\,. \tag{4}$$

Equation 4 shows that any reduction in the partial correlation must be attributed to $r_{v \cdot y} \times r_{v \cdot t}$ because the correlations in the denominator will serve only to increase $r_{t \cdot y|v}$ relative to $r_{t \cdot y}$.

To obtain the impact necessary to invalidate an inference, begin by defining $r^{\#}$ as a quantitative threshold for making inferences from a correlation. For example, $r^{\#}$ can be defined by a correlation of a specific magnitude (e.g., an effect size). Here, we will define $r^{\#}$ by statistical significance. We are well aware that statistical significance is not sufficient for causal inference (Wilkinson and Task Force on Statistical Inference, 1999). But statistical significance is often the first threshold in a two-step procedure for making causal inferences, "where first the likelihood of an effect (small *p* value) is established before discussing how impressive it is" (Wainer & Robinson, 2003, p. 25). That is, most social scientists are uncomfortable making causal inferences if their estimated effect (or something more extreme) could have occurred more than a small percentage (e.g., 5%) of the time by the chance of sampling when in fact the null hypothesis is true.

Given the definition of $r^{\#}$, the inference from r_{xy} is invalid if $r_{xy|v} < r^{\#}$. Therefore, to obtain the impact necessary to invalidate the inference, set $r_{xy|v} < r^{\#}$ in Equation 4. Next, to give maximal credence to the challenge that the inference is invalid because of exclusion of a confounding variable, maximize Equation 4 with respect to *impact*. Frank (2000) shows that this maximum

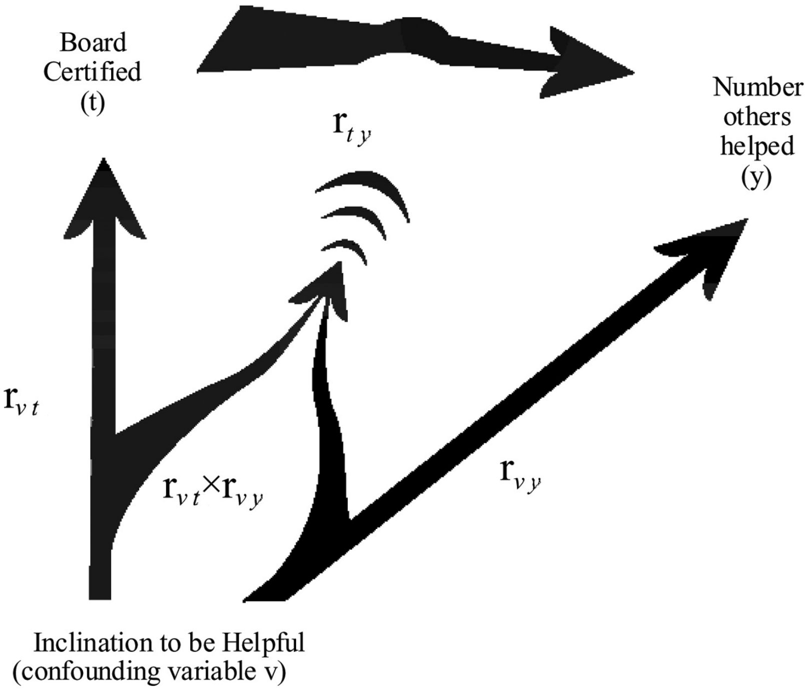

FIGURE 2. *The impact of a confounding variable on a regression coefficient.*

occurs when $r_{v\cdot y} = r_{v\cdot t}$ (given the constraint: *impact* = $r_{v\cdot y} \times r_{v\cdot t}$). Given this maximum, *impact* can be substituted for $r_{v\cdot y} \times r_{v\cdot t}$, $r^2_{v\cdot y}$ and $r^2_{v\cdot t}$ in Equation 4 and solving for *impact* yields that $r_{xy|v} < r^\#$ if

$$impact > (r - r^\#) / (1 - |r^\#|). \qquad (5)$$

Thus, the quantity $(r - r^\#) / (1 - |r^\#|)$ defines the impact threshold for a confounding variable; if there is a confounding variable with *impact* greater than $(r - r^\#) / (1 - |r^\#|)$, then the relationship between the treatment and outcome, given the confound ($r_{xy|v}$), would fall below the threshold ($r^\#$) for making a causal inference.[13]

Critically, because the impact is defined by correlation coefficients (*impact* = $r_{v\cdot y} \times r_{v\cdot t}$), it can be readily understood by social scientists comfortable with correlation and the general linear model. This makes it an ideal complement to our use of propensity weighting, which is applied to a general linear model.

Even if an inference is robust with respect to concerns about unmeasured confounding variables, there still may be challenges to the inference based on the generality or external validity of the findings. For example, the effect we find for NBPTS certification on help provided in our study may not generalize to effects likely to be observed among other cohorts of NBPTS-certified teachers or in other states. The optimal response would be to randomly sample data from the entire population to which we hope to generalize. But in this study, as in most moderately sized studies, such a sampling scheme was not feasible. This generates a quandary: If we can generalize our results only to the immediate population from which we sampled, then our study has limited value for general policy. Does our study have no meaning for those considering policies related to NBPTS certification in cohorts or geographic regions other than those directly represented in the sample?

To quantify the robustness of an inference with respect to sample representativeness, Frank and Min (2007) conducted a thought experiment in which part of a sample is replaced with cases from some unobserved subpopulation. Assuming the means and variances of predictors and the outcome would be the same in the unobserved sample as in the observed sample, a correlation from a combined sample including the original cases as well as cases from the unobserved population would be a simple weighted combination of the observed and unobserved correlations:

$$r_{ty}^{combined} = (1 - \pi)r_{ty}^{ob} + \pi r_{ty}^{un}, \qquad (6)$$

where π is the proportion of the combined sample that is constituted by the unobserved cases, r_{ty}^{ob} is the correlation in the observed sample, and r_{ty}^{un} is the correlation in the unobserved sample that replaces some of the original sample. Then an index of robustness can be calculated by assuming that the null hypothesis holds for the unobserved data ($r_{ty}^{un} = 0$) and solving for π to determine what proportion of the original sample must be replaced with $r_{ty}^{un} = 0$ to invalidate the inference. Frank and Min's (2007) calculations then show that the inference is invalid if

$$\pi > 1 - r^{\#}/r_{ty}^{ob}. \qquad (7)$$

Thus the quantity $\pi > 1 - r^{\#}/r_{ty}^{ob}$ defines the index of external validity (IEV) for $r_{ty}^{un} = 0$ or IEV($r_{ty}^{un} = 0$).

Treatment of missing data. Two of our covariates had nontrivial missing data for both NBPTS-certified teachers and non-NBPTS-certified teachers: "plans to leave teaching" (missing = 16) and "enhancement of teaching through leadership" (missing = 281). For our primary analyses, we set those with missing data on plans to leave teaching to a separate category and those with missing data on enhancement of teaching through leadership to a value of 0 and included a flag in our models indicating missing data. This approach is advocated by Cohen and Cohen (1983).

A more recent treatment of missing data is to use a multiple imputation procedure (Allison, 2000), whereby plausible values on predictors are imputed as a function of other predictors. This step is repeated multiple times to produce new data sets, each time with random error (with variance based on the observed data) introduced. A model is then estimated separately for each data set, and an average estimate is obtained across the data sets, with standard errors based on the average standard error as well as the variance in estimates across the data sets. In this study, we used the SAS PROC MI procedure to impute five data sets, using a Markov chain Monte Carlo method (with a Jeffrey prior), constraining values for each variable to the range that occurred in the observed data. The imputed values for each predictor were based on all other predictors in our models.

Although multiple imputation can reduce bias and increase efficiency, the properties and procedures for multiple imputation using categorical variables are less well known. In our imputation procedure, we represented plans to leave teaching with four dummy variables, each of which was assigned to be missing if the original variable was missing. Similarly, we represented level of education with four dummy variables (although there were no missing data on education level among the 1,138 final cases). Finally, because the estimate from the imputation procedure was larger (and therefore more liberal from a policy perspective) than the comparable estimate designating missing values with dummy variables, we used the latter procedure for all other analyses.

Results

Propensity Model

The results of the logistic regression for the propensity to become an NBPTS-certified teacher are given in Table 2. Those who believed that leadership could enhance their teaching were statistically more likely to become NBPTS-certified teachers. Furthermore, being an NBPTS-certified teacher was statistically related to current level of education (χ^2 = 25.788, $p \leq .0001$) with those with master's plus additional hours of school and PhDs more likely to become NBPTS-certified teachers. Females were borderline more likely to become NBPTS-certified teachers ($p \leq .06$). Although other predictors were not statistically significant, we retain them in the model because they were of a priori importance and contribute to the best prediction possible (Rosenbaum & Rubin, 1983). The logistic function correctly classified 61% of cases using a propensity of .13 as the threshold for classification as an NBPTS-certified teacher (recalling that 14% of teachers were NBPTS-certified teachers).

Table 3 shows results of statistical tests for differences between NBPTS-certified teachers

TABLE 2

Logistic Regression for Propensity of Being NBPTS Certified

Independent Variable	Estimate	Standard Error	Wald χ^2	Pr > χ^2
Intercept	–4.871	.880	30.63	<.0001
Enhancement of teaching through leadership	.655	.158	17.144	.0001
Missing on enhancement of teaching through leadership	.960	.576	2.780	.095
Perceived advantage of certification	.111	.159	.484	.489
Number of other teachers who helped respondent with instruction	.188	.111	2.872	.090
Number of other NBPTS-certified teachers in school	.109	.069	2.53	.112
Number of other NBPTS-certified teachers in school, squared	–.012	.014	.700	.403
Plans to leave teaching:				
As long as I am able	.041	.254	.027	.870
Until I am eligible for retirement	.212	.262	.651	.420
Continue unless something better comes along	.157	.406	.149	.700
Leave teaching as soon as I can	–.338	.895	.143	.706
Undecided at this time	–.540	.320	2.84	.092
Missing	reference			
Female	.937	.488	3.69	.055
White	–.063	.251	.062	.803
Level of education:				
Bachelor's degree	–.234	.335	.490	.485
Bachelor's degree plus additional hours	–.787	.251	9.87	.002
Master's degree	–.530	.264	4.03	.045
Master's degree plus additional hours	.287	.207	1.916	.166
PhD	reference			
Years teaching	–.003	.011	.061	.806
Highest grade level taught	–.003	.024	.020	.900

Note. *n* = 1,131. NBPTS = National Board for Professional Teaching Standards; reference = the reference category for that particular variable.

and non-NBPTS-certified teachers on each covariate, weighting for the propensity using the EOTM as in Equation 2. When using the propensity weights, there were no statistically significant differences between NBPTS-certified teachers and non-NBPTS-certified teachers on the predictors used in the propensity model. Note that there were significant differences between NBPTS-certified teachers and non-NBPTS-certified teachers on several covariates (perceived enhancement of teaching through leadership, missing data on perceived enhancement of teaching through leadership, plans to leave the profession, teacher's level of education)[14] before weighting, indicating the importance of accounting for propensity to achieve balance (Morgan & Harding, 2006).

We also express the value of the weighting scheme in terms of the reduction in the impacts of those covariates for which the differences between NBPTS-certified teachers and others was statistically significant prior to employing the weights. Employing the weights reduced the magnitude of the impact of the enhancement of teaching through leadership from .0110 to .0025, a reduction of 77%; reduced the impact of number of other teachers who helped respondent with instruction from –.0207 to –.0030, a reduction in magnitude of 86%, and reduced the impact of female from .0023 to .0008, a reduction of 66%. Thus, on average, more than three fourths of the impact of the most important covariates was reduced by employing the weights. This confirms the value of the weights in accounting for alternative explanations that can be attributed to the measured covariates.

Having established that weighting by propensity achieves balance on our covariates, we next estimate the effect of NBPTS certification on the

TABLE 3

Testing for Balance Between NBPTS-Certified and Non-NBPTS-Certified Teachers, Weighting for Propensity (EOTM)

	NBPTS	Non-NBPTS				
	Certified	Certified			t Ratio	χ^2
Variable	(n = 160)	(n = 971)	t Ratio[a]	χ^2	(unweighted)	(unweighted)
Number of others helped	1.39 (3.80)	.90 (1.06)	6.62		5.89	
Enhancement of teaching through leadership	2.86 (1.68)	2.83 (1.68)	.75		5.24	
Missing on enhancement of teaching through leadership	.15 (.93)	.16 (.39)	–.21		–3.33	
Perceived advantage of certification	1.95 (1.23)	1.95 (.62)	.18		.90	
Number of other teachers who helped respondent with instruction	.90 (2.10)	.92 (.82)	–.46		2.60	
Number of other NBPTS-certified teachers in school	2.44 (6.47)	2.36 (2.64)	.59		1.76	
Number of other NBPTS-certified teachers in school, squared	6.86 (33.22)	6.48 (12.60)	.52		.95	
Plans to leave teaching:				9.03		1069
As long as I am able	457	502				
Until I am eligible for retirement	357	341				
Continue unless something better comes along	38	52				
Leave teaching as soon as I can	20	14				
Undecided at this time	173	203				
Missing	26	17				
Female	.95 (.59)	.93 (.27)	.88		1.95	
White	.82 (1.00)	.85 (.39)	–1.29		–.19	
Level of education:				6.45		696
Bachelor's degree	65	79				
Bachelor's degree plus additional hours	291	348				
Master's degree	227	202				
Master's degree plus additional hours	483	494				
PhD	6	6				
Years teaching	15.94 (18.05)	15.98 (9.58)	–.08		.70	
Highest grade level taught	8.55 (9.91)	8.64 (4.16)	–.37		.25	

Note. NBPTS = National Board for Professional Teaching Standards; EOTM = effect of the treatment for people at the margin of indifference.

[a] Positive value indicates NBPTS-certified teachers have higher mean than non-NBPTS-certified teachers.

number of others helped, accounting for covariates by weighting by propensity as well as via traditional statistical control.

Estimating the Effect of NBPTS Certification on Help Provided With Instructional Matters

Table 4 shows the estimated effects of NBPTS certification on the number of others helped with instructional matters from our various models. When weighted by propensity to become an NBPTS-certified teacher using the EOTM, NBPTS-certified teachers were nominated, on average, by 0.569 more other teachers than were non-NBPTS-certified teachers ($p \leq .001$). By comparison, the estimated effect of NBPTS certification is 0.603 in the unweighted analysis with covariates.[15] Thus, the estimated effect was reduced only a small amount when we focused on the region of overlap in propensity (see

TABLE 4

Estimated Effect of NBPTS Certification on Number of Colleagues Helped With Instructional Matters in the School

Model[a]	Coefficient	Standard Error	*t* Ratio	*p* Value
Weighted by propensity (EOTM)	.569	.138	4.12	≤.001
Weighted by propensity (treatment effect for the treated)	.598	.130	4.60	≤.001
Weighted by propensity (treatment effect for the control)	.562	.138	4.07	≤.001
Unweighted, with covariates[b]	.603	.092	6.56	≤.001
Unweighted with covariates, using multiple imputation	.621	.092	6.75	≤.001
Unweighted, no covariates	.583	.092	6.35	≤.001
Unweighted, no control for school	.540	.092	5.88	≤.001
NBPTS-certified teacher versus other teachers who applied, EOTM (n = 280, NBCT = 160, non-NBCT = 120)	.577	.167	3.46	≤.001
NBPTS certified teacher versus other teachers who did not apply EOTM (n = 1,017, NBCT = 160, non-NBCT = 857)	.562	.139	4.04	≤.001

Note. n = 1,131 unless otherwise stated. Standard errors based on 500 bootstrap replications. NBPTS = National Board for Professional Teaching Standards; EOTM = effect of the treatment for people at the margin of indifference; NBCT = National Board Certified Teacher.

a. Schools controlled for with fixed effects in all models unless otherwise stated.

b. R^2 = .21 for standard model with covariates.

Figure 1). Consistently, there was little difference in the effect of NBPTS certification for NBPTS-certified teachers (the treatment effect for the treated: .583) and non-NBPTS-certified teachers (the treatment effect for the control: .569).

Although the difference in help provided by NBPTS-certified teachers and non-NBPTS-certified teachers is modest, we note that the estimate of about .57 is about one half of a standard deviation (1.08) of the outcome (number of others helped), suggesting a healthy effect size. Furthermore, translating into the diffusion process of the school, an effect of 0.57 above the baseline of about .92 implies that NBPTS-certified teachers helped about 1.5 other teachers on average. Extrapolating to a school with 6 NBPTS-certified teachers out of 25 teachers, an additional 9 non-NBPTS-certified teachers would benefit indirectly from the NBPTS certification process. Thus, the effect of NBPTS certification has limited but discernable capacity to spill over to those who did not experience the certification process.

In outlining our analytic approach, we argued that there was a double selection process into NBPTS certification. First, a teacher had to apply to become NBPTS certified, and then a teacher had to satisfy the portfolio requirements and pass the assessment procedures to become an NBPTS-certified teacher. To eliminate differences between those who applied and those who did not, we estimated a separate model based on the comparison of NBPTS-certified teachers only with those who applied but were not successful in becoming an NBPTS-certified teacher (n = 280). This begins with a separate propensity model estimated from only this subsample. Although we do not present the full model, being White was the only statistically significant predictor. Adjusting for the other covariates, the odds that a White teacher was certified were 2.06 greater than the odds that a non-White teacher was certified. (This finding may be suggestive either of unequal opportunities to become NBPTS certified or of racial and cultural biases in the NBPTS certification processes; further investigation on this issue is warranted.)

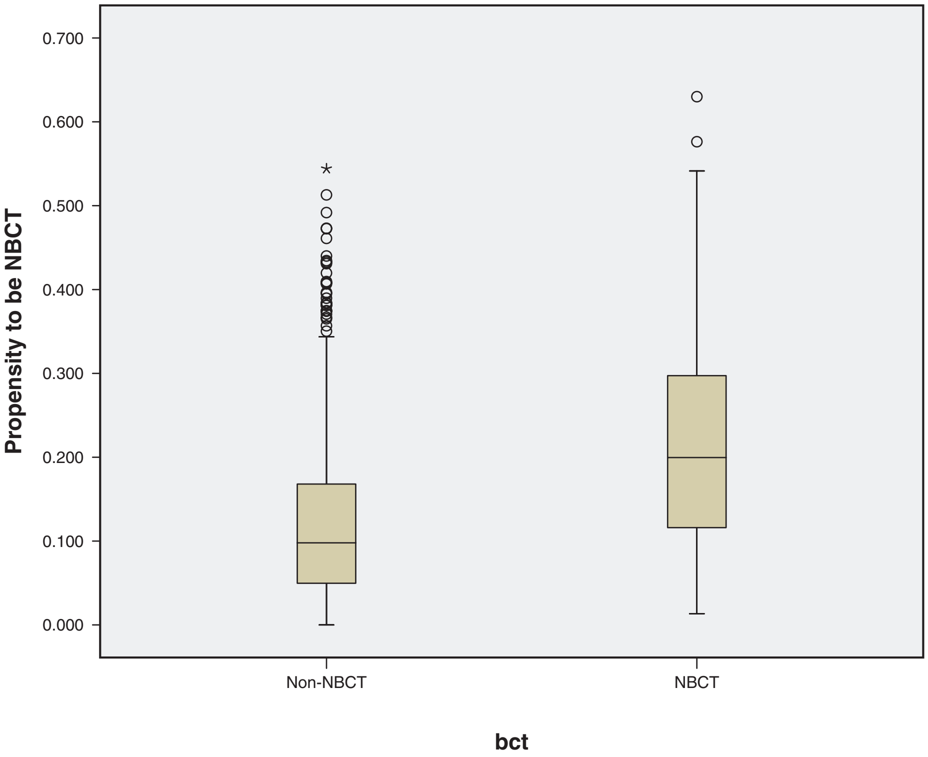

FIGURE 3. *Box plot comparison of distributions of propensity between NBPTS-certified teachers and non-NBPTS-certified teachers: The region of common support.*
Note. NBPTS = National Board for Professional Teaching Standards; NBCT = National Board Certified Teacher.

Weighting the cases using the propensity of being an NBPTS-certified teacher given application for certification, NBPTS-certified teachers were nominated as helpful by 0.577 more other teachers than were non-NBPTS-certified teachers who applied but who did not satisfy the portfolio requirements and assessment procedures. Thus, when the first selection criterion of application is removed, the effect of NBPTS certification appears about the same as when it is not.

Some of the difference between NBPTS-certified teachers and others who applied may be due to actual differences in teacher characteristics (e.g., ability to reflect on teaching practices), not an effect of NBPTS certification. Therefore, perhaps the better comparison is between NBPTS-certified teachers and other teachers who did not apply but who had similar propensity to become NBPTS certified. This comparison is shown in the last row of Table 4, with an estimate of .562 of NBPTS certification on help provided (using the EOTM based on a propensity estimated from NBPTS-certified teachers and others who did not apply).

The most striking aspect of Table 4 is that all the estimates that control for school effects and use the final sample (n = 1,131) are between .569 and .621 (using the multiple imputation procedure), or within .052 of one another.[16] The difference is less than the standard error for any of the estimates, and therefore none of the differences between estimates would even approach statistical significance.

The similarity of the estimates between the propensity models and the OLS with covariates is likely due to the overlap in propensities, as shown in Figure 3. The region of common support, where there are enough teachers who are and are not NBPTS certified to obtain estimates, is fairly large, ranging from .05 to .5. Given this region of common support, estimates from OLS regression that apply across the range of propensity are likely to be similar to those that focus the estimate on certain ranges of propensity (Morgan & Harding, 2006). In fact, the largest differences in Table 4 are not due to how or whether the propensity scores are used but to model specification (e.g., use of covariates or not) and treatment of missing data (via multiple imputation or flags for missing data).

In the end, the model estimated, use of propensity scores, treatment of missing data, or subsample used does not affect our inference: NBPTS certification affects the number of others a teacher helps with instructional matters in her school. That is, having controlled for the covariates available in our data, we now move past correlation to statistical and causal inference. In the next subsection, we use robustness indices to explicitly recognize that there could be alternative causes that invalidate our inference or that the inference may not generalize across contexts.

The Robustness of the Inference

Even given the bias reduction through propensity weighting (or inclusion of covariates in a regression model), there may still be bias in an estimate due to unobserved confounding variables (Morgan & Hardin, 2006; Rosenbaum, 2002; Shadish et al., 2002). For example, those who were successful in pursuing NBPTS certification may have had a stronger inclination and ability to be helpful. In fact, as the NBPTS certification process requires teachers to write commentaries about their teaching and reflect on the evidence they provide about their teaching,[17] those who are inclined to be helpful may be better at the reflective practices required to become NBPTS certified than those who are not. Optimally, we would be able to assess the effect of NBPTS certification on help provided to other teachers through a randomized experiment or by measuring and controlling for all possible confounds, such as inclination to be helpful. But instead, we must estimate an effect controlling only for the measured confounds, as is often the case in the social sciences (Frank, 2000).

While we may be close to exhausting our ability to reduce bias that can be attributed to confounding variables measured in our data, we use Frank's (2000) indices to quantify how much the impact of an unobserved confound must be to invalidate the inference that NBPTS certification affects the number of others a teacher helps with instructional matters. Here, we base the analysis on the estimate and inference using propensity weighting to estimate the EOTM, the most conservative of the estimates that used the full sample and controlled for covariates.

Given the sample size of 1,131, the threshold for statistical significance, $r^{\#}$, is .058. The observed t ratio of 4.13 (4.13 = .57 / .138) translates to a correlation between being an NBPTS-certified teacher and number of others helped of r = .122.[18] From Equation 5, the impact of an unmeasured confound (recall *impact* = $r_{v \cdot y} \times r_{v \cdot x}$; see Figure 2) would have to be greater than .068 to invalidate our inference; the *impact threshold* = $(r - r^{\#}) / (1 - |r^{\#}|) = (.122 - .058) / (1 - |.058|) = .068$. Correspondingly, each component correlation would have to be equal to .26 (see technical appendix for calculations).[19] Thus, to invalidate the inference that NBPTS certification increases the help provided by a teacher, a confounding variable would have to be correlated with NBPTS certification at 0.26 and with help provided at 0.26. These are moderate correlations by social science standards (Cohen & Cohen, 1983). Moreover, these are zero-order correlations, assuming that the unmeasured confound is uncorrelated with the measured covariates (see Frank, 2000). The relevant partial correlations from which the impact of an unobserved confound would be constructed would be smaller than the zero-order correlations because of correlations with existing covariates.[20]

Although the magnitude of the impact threshold for an unmeasured variable can be interpreted in terms of general findings in the social sciences, it is also helpful to compare the threshold to the impacts of measured covariates. The extent to which a teacher believes leadership will enhance teaching has the strongest impact of the measured covariates. Its impact on the coefficient for NBPTS-certified teachers on help provided is 0.011, which is the product of the correlation with being an NBPTS-certified teacher (0.17) and the correlation with number of other teachers helped (0.06). Thus, the impact of an unmeasured confound necessary to invalidate the inference, .068, would have to be more than six times greater than the strongest impact of the measured covariates, .011.

Having quantified the robustness of the inference with respect to internal validity in terms of the impact of an unmeasured confounding variable, we now turn to quantifying the robustness with respect to external validity. Using Frank and Min's (2007) approach, the question is how much of the sample would have to be replaced with unobserved cases for which the null hypothesis is true (i.e., NBPTS certification has no effect on the number of colleagues helped with instruction) to invalidate the inference. This thought experiment helps quantify the concern that the sample may not have been representative of some important target population (e.g., teachers in other cohorts or states). Using Frank and Min's (2007) calculations as in Equation 7, the $\text{IEV}(r_{ty}^{un} = 0) = 1 - r^{\#}/r_{ty}^{ob} = 1 - .058\ /\ .122 = .52$. Thus, more than 52% of the original sample would have to be replaced to invalidate the original inference that NBPTS certification affects the number of colleagues a teacher helps with instructional matters. That is, 52% of the teachers in our study would have to be considered not representative of teachers in other states or cohorts to invalidate our inference. Further calculations based on Frank and Min (2007, n. 15) indicate that to invalidate our inference, the correlation between being NBPTS certified and number of others helped would have to be between –1 and –.9 among those 75 cases that were excluded due to missing values (on grade level, gender, level of education, and perceived advantages of certification).[21] Similarly, if all of the original cases were used, including those too junior to be NBPTS certified, the correlation between NBPTS certification and number of others helped would have to be –.1 to invalidate our inference for the original sample. Given that it is unlikely that becoming an NBPTS-certified teacher *reduces* a teacher's helpfulness, we claim our inference is reasonably robust with respect to the filters applied to the sample because of missing data and eligibility.

Discussion

We have asked a very simple question: Does NBPTS certification affect the number of colleagues a teacher helps with instructional matters? The simple observation is that NBPTS-certified teachers helped about 1.5 others in contrast to non-NBPTS-certified teachers who helped about .9 others, a difference of about .6. But can we attribute the difference to the process of NBPTS certification? If so, then there is an extended influence of the NBPTS certification process, and NBPTS certification could be one way to cultivate social capital in schools that generally contributes to effective teaching and innovation.

From our data and analyses, we infer that NBPTS certification increases the number of others a teacher helps with instructional matters. We recognize that making causal inferences from observational data is controversial (Shadish et al., 2002; Wilkinson and Task Force on Statistical Inference, 1999) and our particular inference might be wrong. Such is the nature of causal inference; it is inference, not certainty. In fact, the counterfactual implies that causality for an individual is never certain (Holland, 1986).

Substantive Interpretation

We interpret our inference that NBPTS certification affects the number of colleagues a teacher helps with instructional matters in some historical context. Teachers for the most part have worked alone, drawing sparingly on their colleagues for assistance and support. While Lortie (1975) did not discount the importance of colleagues, he argued that the norm governing assistance was voluntarism and individual choice. Teachers are willing to respond to their colleagues but do not normally offer unbidden assistance, even to novice or junior colleagues. In part, this may be due not simply to teaching's culture but also to the social meanings more generally ascribed to helping behavior.

As Huberman (1993) argues in his chapter on teacher as artisan or craftsman, helping behavior opens up difficult interactions because help seeking may be interpreted as a sign of weakness or incompetence, while offers of help may be spurned, yielding hurt feelings and resentment. Hence, helping behavior in professional settings may be regarded, relatively speaking, as an unnatural act. Glidewell et al. (1983) elaborate, noting that helping behavior is less likely when workers are relatively independent of one another and operate under norms of autonomy and status equality. Insofar as advice may imply higher status (Blau, 1967), the absence of status distinctions among workers is likely to reduce helping behavior. Such a perspective suggests that the introduction of a new status (e.g., NBPTS certification) that signifies or signals special competence might alter help giving and seeking among teachers. Still, social patterns deeply rooted in the structure, culture, and traditions of teaching work are unlikely to be overturned in so brief a compass as one decade. Findings to the contrary, such as those in this study, then must be regarded with some weight.

Methodological Interpretation

From a methodological standpoint, standard regression techniques are the workhorse of the social sciences. But they assume linear relationships and homogeneous effects. It is these limitations that have historically motivated extension of regression into logistic regression (for nonlinear relationships for dichotomous outcomes) and multilevel models (for heterogeneous effects). Importantly, each of these techniques can be considered an application of weighted least squares (with weights based on cell sample sizes for logistic regression or sampling error within level two units for multilevel models). Furthering the trend, the weights based on propensity scores employed in this study allow for estimation of the NBPTS certification effect without assuming linear relationships between covariates and NBPTS certification and allowed us to differentiate among treatment effects (e.g., for the treated and for the control). This methodological point stands in general, although in this particular case, there was little difference between estimates that employed propensity weighting and the unweighted OLS. This is likely due to the extensive region of common support or overlap in propensity between treatment and control groups in these data.

Although employing propensity weights allowed us to relax the typical regression assumptions of linearity and homogeneous effects, the conceptualization and implementation of statistical control through regression still elegantly represents concerns regarding confounding and spurious effects. Specifically, regression controls for confounding variables through the product of the two correlations associated with the confounding variable: the correlation with the predictor of interest and the correlation with the outcome. We then leveraged this conceptualization to express the robustness of our inferences in terms of the product.

The information provided by the robustness for confounding variables and sample representation indices calls attention to the need to quantify the sensitivity of inferences to model specification and sample representation (Holland, 1989). This is especially relevant given current ideas in the philosophy of science that emphasize how new ideas emerge out of and are debated within a social and cultural context (Abbott, 1998; Ben-David, 1984, 1991; Schofer, 2003). Translating into contemporary social science, the robustness indices help to ensure that debate about scientific inferences is phrased in the precise assumptions of statistical and causal inference—possible alternative explanations or variable effects (Holland, 1986). This should help researchers and policy makers discern immediate and pragmatic debates about the validity of specific inferences from academic debates about the general and theoretical conditions necessary for causal inference (e.g., Cook & Campbell, 1979, versus Cronbach, 1982).

Limitations

If our results are judged against the deep-seated and prevailing pattern of teacher work, there are some obvious limits to our study. First, we are inferring an effect of NBPTS certification, but the observed coefficient could be due to preexisting differences between those teachers who became NBPTS certified and those who did not. The gold standard would be to conduct a randomized experiment of assignment to NBPTS certification (U.S. Department of Education, 2003). But given the intense process and objective standards of NBPTS certification, it is not ethical or practical to randomly assign teachers to be NBPTS certified or not. We could have also employed a quasi-experimental design, measuring teachers pre- and post-NBPTS certification. But our dependent variable, helping behavior, is likely slow to change and thus would have required a time lag of several years to detect effects. Such a study would not reflect the immediacy and urgency of policies related to NBPTS certification. Therefore, in the absence of a randomized or quasi experiment, we have analyzed our cross-sectional data using statistical controls, allowing that our inference may be incorrect, and then quantifying the robustness of our inference.

Second, given the differences in other effects across cohorts of NBPTS-certified teachers

(Harris & Sass, 2007), it is possible that the effects we observe for the NBPTS-certified teachers in this study would not generalize to others who were induced to become NBPTS certified via new incentives. But because the findings are reasonably robust with respect to changes in the composition of the sample, we speculate that NBPTS certification would affect the number of others helped by most new cohorts, even if the effects were not as large as observed here.

Third, we need to know in more qualitative terms about the nature and content of helping behavior together with its ultimate effects on instruction. Helping behavior is a complex phenomenon in social–psychological and organizational terms (Hansen, 1999). What we have uncovered is a brute fact—that NBPTS-certified teachers provide more help than comparable peers and that such behavior appears causally related to certification status. Such a fact is provocative for advocates of the NBPTS and more generally for champions of professionalism in teaching, but many questions remain. Why should NBPTS certification status produce more helping behavior? What school circumstances and situations elicit useful help from NBPTS-certified teachers? What is the actual content of the exchanges of help (cf. Coburn & Russell, 2007)? How does such help influence important aspects of curriculum, instruction, and assessment?

Policy Implications

Policy implications stem from one of our initial premises and the findings in this study. First, our initial premise drawn from the literature suggests that schools are more effective when teachers in general and highly trained teachers such as NBPTS-certified teachers in particular provide instructional help to others in their schools (Bryk et al., 1998; Bryk & Schneider, 2002). Second, the crux of the causal inference from this study is that NBPTS certification affects the number of others a teacher helps with instructional matters in her school.

Given the combination of our initial premise and our findings, this study has two key implications for policy. First, incentives for NBPTS certification might be used particularly to employ NBPTS-certified teachers in a variety of roles outside their own classrooms. To date, the value that NBPTS-certified teachers add to a school's faculty has not been of paramount concern either to policy makers or to researchers, although it was integral to the definition of an NBPTS-certified teacher. But if managed skillfully and well, helpful NBPTS-certified teachers may enhance overall school performance and help justify the expenditures associated with NBPTS certification. Along these lines, one important piece of missing information is the effect of NBPTS certification on teacher retention. Obviously, if NBPTS-certified teachers tend to leave teaching shortly after certification, then their long-term value as a human capital investment will be reduced. Research must address this issue in the near future.

Second, as so-called distributed perspectives on leadership gain currency (see, e.g., Spillane, 2006), uses of NBPTS certification can play a role in both identifying and signaling experienced teachers' willingness and capability to undertake more leadership. Furthermore, the evolving political economy of school districts also is implicated, as one critical decision facing districts today concerns what functions to manage directly and what to "contract out" to external providers (see Supovitz, 2006, for discussion). NBPTS certification can assist districts in strategic staffing of instructional support, potentially saving expenditures on outside consultants by making better use of an internal leadership cadre.

Conclusion

As a major reform in American education, NBPTS certification has yet to prove itself in certain terms. But we interpret the evidence so far as indicating for the most part that NBPTS-certified teachers are effective teachers and that the status of NBPTS certification can serve a number of potentially useful functions in schools and districts. If NBPTS certification status promotes helping behavior among teachers, it is one important indicator of their leadership potential in such formal roles as mentor teacher, instructional coach, cooperating teacher (with university-based teacher education), team- or grade-level leader, and others. Such leadership is increasingly important because many schools across the country are developing teacher leader positions intermediate between the principal and a school's staff (see, e.g., Mangin & Stoelinga, in press). NBPTS certification is one natural device for "certifying" a teacher's capability in filling these new roles; evidence indicating both that NBPTS-certified teachers provide help more than comparable peers and that certification status enjoys a causal relationship with such help is an important finding in the evolving social organization of the teaching occupation.

Uses of NBPTS-certified teachers in schools and districts ultimately will depend on developments in the policy and administrative spheres because, on their own, NBPTS-certified teachers are unlikely to realize their potential as social resources for instructional improvement. Teaching's traditional ethos of independence (Lortie, 1975), together with the "zone of discretion" allocated to teachers, militates against the systematic emergence of teacher leadership in schools. But as instructional leadership increases in importance, the likelihood that traditional staffing patterns—one principal, perhaps an assistant principal, and a staff of classroom teachers—will supply such leadership appears slight. New organizational designs will be needed to better support new instructional designs, and the NBPTS-certified teacher can play an important role at their intersection. The result reported here is a slender but unmistakable indication of their potential.

This study is as much about causal inference as about policies related to NBPTS certification because the conclusions of the preceding two paragraphs are premised on the statement "If NBPTS certification status promotes helping behavior among teachers . . ." If the inference were not valid, the resulting policy implications would be irrelevant and could even have harmful effects. As it is, our attention to the validity of the inference establishes a baseline for a debate regarding the policy implications. It is in this sense that the dual purposes of this article are convergent.

Technical Appendix
Calculating Impact Threshold

t critical	*n*	*r*#	Observed *t*	r(x,y)	ITCV	r(x,cv)	r(y,cv)
1.96	**1,131**	= +A2 / SQRT(A2 * A2 + B2 – 3) .058	**4.13**	= +D2 / SQRT(B2 – 2 + D2 * D2) .122	= +(E2 – C2) / (1 – C2) .068	= +SQRT(F2) .26	= +SQRT(F2) .26
Multivariate (with other covariates, **z**, in model)							
t critical	num **z**	*r*#	R^2 (x,**z**)	R^2 (y,**z**)	ITCV	r(x,cv)	r(y,cv)
1.96	**48**	= +A7 / (SQRT(A7 * A7 + B2 – B7 – 3)) .060	**0.163**	**0.124**	= +F2 * SQRT((1 – D7) * (1 – E7)) .058	= SQRT(+F7 * SQRT((1 – D7) / (1 – E7))) .238	= SQRT(+F7 * SQRT((1 – E7) / (1 – D7))) .244

Note. R^2 (x,z) and R^2 (y,z) only need to be entered to correct impact threshold for a confounding variable (ITCV) calculations in F7–H7. Can be downloaded from http://www.msu.edu/~kenfrank/.

Notes

[1] Controlling for differences between schools using fixed effects.

[2] Note, however, that for our dependent variable based on the number of times nominated as helpful with instruction, nominations could come from anyone in the school, regardless of whether the respondent was in the final sample.

[3] We recognize that a teacher may pursue National Board for Professional Teaching Standards (NBPTS) certification as a form of professional development, but we emphasize the instrumental value of NBPTS certification because earlier findings suggested that NBPTS certification was more effective at identifying effective teachers than as a mechanism for professional development.

[4] Because the effect of number of others who were NBPTS-certified teachers on becoming NBPTS certified may be nonlinear, we also included in our models a quadratic term based on the square of the number of others who were NBPTS-certified teachers.

[5] Note $e(x)$ is called a "propensity" and not a "probability" because the treatment condition for each subject in the data set is discrete and known, making the term *probability* less applicable.

[6] Our tests are relatively conservative because we did not reduce the Type I error rate for multiple tests in determining statistical significance.

[7] See Frank (2000) and the subsection titled, "Quantifying the robustness of the inference," for a discussion of impact.

[8] Equivalent to a negative binomial.

[9] Personal e-mail, Keisuke Hirano, May 1, 2006.

[10] Code for the statistical software SAS for this procedure is available from the first author on request.

[11] One can perform sensitivity analysis for the assumption that the instrument is related to the outcome only through the predictor of interest (Diprete & Gangle, 2004).

[12] Estimates using instruments defined by perceived advantage of NBPTS and our other covariates were very similar to those reported in the main text. Glazerman, Levy, & Myers (2003) also found serious limitations when comparing estimates based on instrumental variables with those based on randomized experiments in a meta-analysis of effects of welfare, job training, and employment service programs on earnings.

[13] The expressions can be easily adapted to focus on one component correlation when researchers have specific prior beliefs about the strength of the other correlation. The expressions can also be modified to account for the presence of other covariates in the model. See Frank (2000).

[14] And gender had a p value of .06.

[15] The estimate from the Poisson model (weighted by propensity and corrected for overdispersion, equivalent to a negative binomial) was 0.502 with standard error of 0.066 and $p \leq .0001$. Taking e^{502}, the estimate is 1.65, implying that NBPTS-certified teachers are nominated about 1.65 times more frequently than non-NBPTS-certified teachers. Given that those who were not NBPTS-certified teachers were nominated by about .9 others, $1.65 \times .9 = 1.49$. Thus, the difference between NBPTS-certified teachers and others is $1.49 - .9 = .59$, which is quite consistent with estimates in Table 3.

[16] The smallest estimate is from the model with no controls (.540), indicating that schools are a suppressor for the NBPTS-certified teacher effect—there are some schools with small numbers of NBPTS-certified teachers who are especially helpful.

[17] See http://www.nbpts.org/candidates/guide/1_portflo.html (retrieved January 15, 2008).

[18] $r = t / (t^2 + df)^{.5} = 4.13 / (4.13^2 + 1{,}128)^{.5} = .122$.

[19] These calculations assume the impact of the unobserved variable is maximized (Frank, 2000) by setting the two component correlations equal to one another. The product would have to be greater if the two components were not equal. Furthermore, these calculations are for a bivariate relationship because the covariates are accounted for via the weighting (see Frank, 2000, for calculations that are multivariate, accounting for observed covariates).

[20] Frank (2000) refers to this as absorption of the impact of an unmeasured confound by existing covariates.

[21] This calculation is just an example of how the indices can be applied to attrition because none of the NBPTS-certified teachers had missing data on the relevant variables.

References

Abbott, A. (1998, November). The causal devolution. *Sociological Methods & Research*, *27*(2), 148–181.

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the economics of organizations. *Journal of Economic Perspectives*, *19*(1), 9–32.

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods and Research*, *28*, 301–309.

Altonji, J. G., Elder, T. E., & 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.

Argote, L., Beckman, S., & Epoie, D. (1990). The persistence and transfer of learning in industrial settings. *Management Science*, *36*, 140–154.

Arrow, K. J. (1974). *The limits of organization.* New York: Norton.

Ben-David, J. (1984). *The scientist's role in society*. Chicago: University of Chicago Press.

Ben-David, J. (1991). *Scientific growth: Essays on the social organization and ethos of science*. Berkeley: University of California Press

Bidwell, C. (2000). School as context and construction: A social psychological approach to the study of schooling. In M. T. Hallinan, (Ed.), *Handbook of the sociology of education* (pp. 15–36). New York: Kluwer Academic/Plenum.

Bidwell, C. (2001). Analyzing schools as organizations: Long-term permanence and short-term change [Extra issue]. *Sociology of Education*, 100–114.

Blau, P. M. (1967). *Exchange and power in social life*. New York: John Wiley.

Bond, L., Smith, T., Baker, W., & Hattie, J. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*. Greensboro: Center for Educational Research and Evaluation, University of North Carolina.

Boyd, W., & Reece, J. P. (2006). Great expectations. *Ed next*. Palo Alto, CA: Hoover Institution. Retrieved January 6, 2008, from http://www.educationnext.org/20062/50.html

Bryk, A. S., & Schneider, B. (2002). *Trust in schools*. New York: Russell Sage.

Bryk, A. S., Sebring, S. B., Kerbow, D. D., Rollow, S. S., & Easton, J. Q. (1998). *Charting Chicago school reform: Democratic localism as a lever for change*. Boulder, CO: Westview.

Burawoy, M. (1979). *Manufacturing consent: Changes in the labor process under monopoly capitalism*. Chicago: University of Chicago Press.

Carnegie Forum on Education and the Economy. (1986). *A nation prepared: Teachers for the 21st century*. New York: Carnegie Corporation.

Cavalluzzo, L. C. (2004). *Is NBPTS certification an effective signal of teacher quality?* Arlington, VA: CNA Corporation.

Clotfelder, C., Ladd, H., & Vigdor, J. (2007, March). *How and why do teacher credentials matter for student achievement?* (Working Paper No. 2). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, Urban Institute.

Coburn, C. E., & Russell, J. L. (in press). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis.*

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral science*. Hillsdale, NJ: Lawrence Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago: Rand McNally.

Copas, J. B., & Li, H. G. (1997). Inference for nonrandom samples. *Journal of the Royal Statistical Society, Series B (Methodological)*, *59*(1), 55–95.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Diprete, T., & Gangl, M. (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology*, *34*, 271–310.

Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods and Research*, *29*, 147–194.

Frank, K. A., & Min, K.-S. (2007). Indices of robustness for sample representation. *Sociological Methodology*, *37*, 349–392.

Frank, K. A., Zhao, Y., & Borman, C. (2004). Social capital and the diffusion of innovations within organizations: Application to the implementation of computer technology in schools. *Sociology of Education*, *77*, 148–171.

Gastwirth, J. L., Krieger, A. M., & Rosenbaum, P. R. (1998). Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika*, *85*, 907–920.

Glazerman, S., Levy, D., & Myers, D. (2003). Nonexperimental versus experimental estimates of earnings impacts. *The ANNALS of the American Academy of Political and Social Science*, *589*, 63–93.

Glidewell, J. C., Tucker, S., Todt, M., & Cox, S. (1983). Professional support systems: The teaching profession. In A. Nadler, J. Fisher, & B. DePaulo (Eds.), *New directions in helping* (pp.189–212). New York: Academic Press.

Goldhaber, D., & Anthony, E. (in press). Can teacher quality be effectively assessed? NBPTS certification as a signal of effective teaching. *Review of Economics and Statistics.*

Hansen, M .T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, *44*, 82–111.

Harris, D., & Sass, T. (2007, March). *The effects of NBPTS-certified teachers on student achievement* (Working Paper No. 4). Washington, DC: Center for Analysis of Longitudinal Data in Education Research, Urban Institute.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equations system. *Econometrica*, *47*, 153–161.

Heckman, J. (2005). The scientific model of causality. *Sociological Methodology*, *35*, 1–99.

Heckman, J., & Robb, R. (1985). Using longitudinal data to estimate age, period and cohort effects in earnings equations, in William M. Mason and Stephen E. Feinberg (Eds.), *Cohort Analysis in Social Research Beyond the Identification Problem*. New York/Berlin: Springer-Verlag.

Hirano, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology*, *2*, 259–278.

Hirano, K., Imbens, G. W., & Ridder G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, *71*, 1161–1189.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–970.

Holland, P. W. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training: Comment. *Journal of the American Statistical Association*, *84*(408), 875–877.

Huberman, M. (1993). The model of the independent artisan in teachers' professional relations. In J. Little & M. McLaughlin (Eds.), *Teachers' work: Individuals, colleagues, and contexts* (pp. 11–50). New York: Teachers College Press.

Kogut, B., & Zander, U. (1992). Knowledge in the firm, combinative capabilities and the replication of technology. *Organization Science*, *3*, 383–397.

Levine, D. I., & Painter, G. (2003). The costs of teenage out-of-wedlock childbearing: Analysis with a within-school propensity score matching estimator. *Review of Economics and Statistics*, *85*(4), 884–900.

Lin, D. Y., Psaty, B. M., & Kronmal, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*, *54*(3), 948–963.

Lortie, D. C. (1975). *Schoolteacher*. Chicago: University of Chicago Press.

Louis, K. S., Marks, H., & Kruse, S. (1996). Teacher's professional community in restructuring schools. *American Educational Research Journal*, *33*(4), 757–798.

Mangin, M. M., & Stoelinga, S. R. (in press). *Instructional teacher leadership roles: Using research to inform and reform teaching*. New York: Teachers College Press.

Morgan, S. L. (2001). Counterfactuals, causal effect heterogeneity, and the Catholic school effect on learning. *Sociology of Education*, *74*, 341–374.

Morgan, S. L., & Harding, D. J. (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods and Research*, *35*, 3–60.

National Board for Professional Teaching Standards. (1999). *What teachers should know and be able to do*. Arlington, VA: Author.

Pan, W., & Frank, K. A. (2004a). A probability index of the robustness of a causal inference. *Journal of Educational and Behavioral Statistics*, *28*, 315–337.

Pan, W., & Frank, K. A. (2004b). An approximation to the distribution of the product of two dependent correlation coefficients. *Journal of Statistical Computation and Simulation*, *74*, 419–443.

Podgursky, M. (2001). Defrocking the National Board: Will the imprimatur of "Board Certification" professionalize teaching? *Education Matters*, *1*(2), 79–82.

Reagans, R., & McEvily, W. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly*, *48*(2), 240–267.

Robins, J. (1987). A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Journal of Chronic Diseases*, *40*(2), 139S–161S.

Robins, J., Hernán, M., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550–560.

Robins, J., Rotnitzky, A., & Scharfstein, D. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In E. Halloran & D. Berry (Eds.), *Statistical models for epidemiology, the environment, and clinical trials* (pp. 1–95). New York/Berlin: Springer-Verlag.

Robins, J. M., & Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *JASA*, *90*, 122–129.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42, 59–66.

Romer, P. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5), S71–S102.

Rosenbaum, P. (2002). *Observational studies*. New York: Springer.

Rosenbaum, P. R. (1986). Dropping out of high school in the United States: An observational study. *Journal of Educational Statistics*, *11*(3), 207–224.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, *66*, 688–701.

Rubin, D. B. (2004). Teaching statistical inference for causal effects in experiments and observational studies. *Journal of Educational and Behavioral Statistics*, *29*(3), 343–368.

Sanders, W. L., Ashton, J., & Wright, S. P. (2005). *Comparisons of the effects of NBPTS-certified teachers with other teachers on the rate of student academic progress* (Final Report). Arlington, VA: National Board for Professional Teaching Standards. Retrieved January 6, 2008, from http://www.nbpts.org/UserFiles/File/ SAS_final_report_Sanders.pdf

Scharfstein, D. A., & Irizarry, R. A. (2003). Generalized additive selection models for the analysis of studies with potentially non-ignorable missing data. *Biometrics*, *59*(3), 601–613.

Schofer, E. (2003). The global institutionalization of geological science, 1800 to 1990. *American Sociological Review*, *68*(5), 730–759.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Spillane, J. (2006). *Distributed leadership.* San Francisco: Jossey-Bass.

Supovitz, J. (2006). *The case for district-based reform.* Cambridge, MA: Harvard Education Press.

Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, *17*, 27–43.

Talbert, J., & McLaughlin, M. (1994). Teacher professionalism in local school contexts. *American Journal of Education*, *102*, 123–153.

U.S. Department of Education. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington, DC: Author.

Wainer, H., & Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance testing. *Educational Researcher*, *32*, 22–30.

Wilkinson, L., and Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604.

Winship, C., & Morgan, S. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659–707.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT Press.

Authors

KENNETH A. FRANK is a professor in Counseling, Educational Psychology and Special Education as well as in Fisheries and Wildlife at Michigan State University, Room 462 Erickson Hall, East Lansing, MI 48824-1034; http://www.msu.edu/~kenfrank/. His substantive interests include the diffusion of innovations, study of schools as organizations, social structures of students and teachers and school decision-making, social capital and resource flow. His substantive areas are linked to several methodological interests: social network analysis, causal inference and multi-level models. His publications include quantitative methods for representing relations among actors in a social network, robustness indices for inferences, and the effects of social capital in schools and other social contexts. He teaches general introductory courses in research methods and quantitative methods as well as advanced courses in multivariate analysis and seminars in social network analysis and causal inference.

GARY SYKES is a professor in the department of teacher education at Michigan State University, 410-A Erickson, Michigan State University, East Lansing, MI 48824; garys@msu.edu. He specializes in policy research on topics ranging from teacher policies to effective school districts to school choice issues.

DOROTHEA ANAGNOSTOPOULOS is an associate professor in the Department of Teacher Education at Michigan State University. Her research interests include socio-cultural analyses of policy implementation, teaching and learning in urban high schools, and teacher education.

MARISA CANNATA is a postdoctoral fellow and research associate at Vanderbilt University, GPC #414, 230 Appleton Place, Nashville, TN 37203; marisa.a.cannata@vanderbilt.edu. Her areas of specialization include teacher career decisions, teacher policies, charter schools, and professional community. She has a PhD in educational policy from Michigan State University.

LINDA CHARD is an associate psychometrician in K-12 Research and Development at the Educational Testing Service, 666 Rosedale Road MS 13P, Princeton, NJ 08541.

ANN KRAUSE is an assistant professor in the Department of Environmental Sciences at the University of ToledoMailstop #604, 2801 West Bancroft, Toledo, OH 43606-3390; ann.krause@utoledo.edu.

RAVEN McCRORY is an associate professor in the Department of Counseling, Educational Psychology and Special Education in the College of Education at Michigan State University, 513G Erickson Hall, East Lansing, MI 48824; mccrory@msu.edu.

Manuscript received July 17, 2006
Final revision received July 2, 2007
Accepted September 15, 2007