Using quasi-experimental methods to select comparisons schools for an evaluation of the Northeast Tennessee College and Career Ready Consortium

- Contact email: mokherc@cna.org

- Title: Research Analyst

- First and second choice of conference section: 1) paper, 2) poster

- Full structured abstract as a PDF document (see following pages)

- Blinded version of the above PDF document (see separate document)
- Names of all authors in their intended order:
    - Christine Mokher, Ph.D.
    - Linda Cavalluzzo, Ph.D.
- Affiliations of all authors
    - CNA Education

- E-mail addresses of all authors
    - mokherc@cna.org
    - cavallul@cna.org

Using quasi-experimental methods to select comparisons schools for an evaluation of the Northeast Tennessee College and Career Ready Consortium

Abstract Body

Background / Context:

This presentation will focus on the quasi-experimental methods used to select comparison schools for an evaluation of a federal investing in innovation (i3) validation grant. The Northeast Tennessee College and Career Ready Consortium (NETCO) consists of 29 high schools participating in a five-year program to expand students' access to rigorous courses. NETCO's activities include (1) creating a regional coordinating body that will analyze course supply and demand in the region and determine course needs; (2) scaling up local promising practices to offer more rigorous courses, including distance and online learning, Advanced Placement, and dual enrollment; and (3) providing new resources on college preparedness, access, and application processes via college counselors. Key objectives include increasing college preparation, as well as college enrollment and persistence rates.

A quasi-experimental matched-control study will be conducted to examine differences in student outcomes among NETCO schools and a group of matched comparison schools. The confirmatory analysis will include two student-level outcomes: the likelihood that students will enroll in college by the fall following high school graduation and the likelihood that students will persist in college, measured by college enrollment in the first and second fall semesters following high school graduation.

Twenty-nine high schools in fifteen school districts were selected to participate in the intervention prior to the award of the grant. This was a convenience sample of districts in Northeast Tennessee who had previously worked in a partnership with grantee on other projects. A group of comparison schools will be selected using propensity score matching. Not all outcome data are available from state administrative records, which means that a group of comparison schools must be selected at the beginning of the evaluation for data collection.

Two sets of propensity score models will be used to identify the comparison group. First, a school-level propensity score model will be estimated to identify comparison schools that are similar to the intervention schools. Second, a student-level propensity score model will be estimated to identify students within the comparison schools who are similar to the students in the intervention schools.

Purpose / Objective / Research Question / Focus of Study:

The research question for this presentation is: How can propensity score matching be used to identify comparison schools for the evaluation of the NETCO intervention? We will describe the propensity score model used to select comparison schools, the techniques used to refine the model, and the recruitment of the comparison schools.

Setting:

Twenty-nine high schools in fifteen school districts from Northeast Tennessee were selected to participate in the intervention prior to the award of the grant. Thirty-eight percent of students are enrolled in schools located in rural locales. The student population is predominately white (95 percent) and low-income (55 percent receive free or reduced priced lunch). The median annual income in the surrounding community is $36,798 compared to $43,455 in non-NETCO schools in the state. Sixteen percent of population has a college degree in the NETCO region compared to 21 percent of the population in the rest of the state.

Population / Participants / Subjects:

The universe of public high schools in the state of Tennessee includes 351 schools. The following were excluded from the sample of potential comparison schools: 28 schools with a school type other than "regular" in the Common Core of Data, 13 adult and continuing education schools, and 22 magnet and charter schools. The remaining sample includes 288 regular public high schools (N=29 intervention schools and N=259 non-intervention schools).

Intervention / Program / Practice:

The intervention aims to scale up successful programs from the region in a cohesive way that can be sustained after the grant period ends and become a model for other areas of the state and the nation. Specifically, this grant will scale up: (1) the course supply and demand study conducted semi-annually conducted by the Northeast Tennessee Distance Learning Consortium; (2) the distance learning courses provided by the Northeast Tennessee Distance Learning Consortium; (3) the online courses provided by the Niswonger Learning Center; (4) AP courses which have been provided at scale in two participating districts; (5) dual enrollment programs modeled on the Educate and Grow program in Kingsport, Tennessee; and (6) the college counselor program in one of the participating districts. The timeframe for the intervention is Fall 2010 to Spring 2015.

Research Design:

A school-level propensity score model is used to identify 29 comparison schools. A binary logit regression model estimates the probability that a school is a member of the Northeast Tennessee College and Career Consortium. For each school, the propensity score is defined as:

$$P(Y = 1) = \frac{e^{\beta X}}{1 + e^{\beta X}} \tag{1}$$

In the formula, Y represents the outcome variable (which equals 1 if the school is a member of the Consortium and 0 otherwise). X represents the set of observed school characteristics representing demographic characteristics of the student population, test scores of the student population, career and technical education concentrations and AP course enrollments, attendance and graduation rates, school resources, and community characteristics. All covariates represent characteristics prior to the intervention, so their values are not influenced by treatment status.

Data Collection and Analysis:

All variables are measured at the school-level, except for the community characteristics, which are measured at the district-level. If a school is missing a value for one of the school-level variables, the average value for the district is used in place of the missing value. Data on the demographics of the student population, test scores of the student population, CTE concentrations, and attendance and graduation rates are from the Tennessee Department of Education school report cards[1] for the 2009/10 academic year. Data on school resources are from the National Center for Education Statistics' Common Core of Data[2] for the 2007/08 academic year (most recent data available). Data on community characteristics are from the 2000 Census, and were also downloaded from the Common Core of Data.

A propensity score is calculated for each school in the sample by substituting that school's set of observed characteristics (X), and the estimated coefficients (β) from the logit regression into the propensity score formula as defined in equation (1). The propensity score is therefore the estimated probability that the school was selected to receive the intervention.

The quality of the match between member *i* and member *j* is defined using the distance function:

$$D(i, j) \quad P_i^T \quad P_j^C {}^2 \tag{2}$$

where P_i^T represents the propensity score for member *i* of the treatment group and P_j^C represents the corresponding propensity score for member *j* of the control group. The closer the predicted propensity scores of the two schools are, the smaller the value of the distance function will be. If the two schools are identical in observed characteristics, the value of the distance function will be zero. Nearest-neighbor matching is used to assign to each treatment group member the one member of the comparison group who is closest in terms of predicted probability of being a member of the Consortium. In other words, for each treatment group member, we choose the comparison group member who results in the lowest absolute value of the distance function D(i, j). Matches are selected without replacement, so each comparison school can only be matched to a single intervention school.

Findings / Results:

Figure 1 illustrates the distribution of the propensity scores for all NETCO and non-NETCO schools prior to matching, while figure 2 shows the distribution after matching. One concern is that there is limited overlap among the propensity scores for the two groups of schools, which may threaten the accuracy of any causal inferences about the impact of the intervention. The range of propensity scores is 0.06 to 0.99 for the NETCO schools and 0.06 to 0.89 for the non-NETCO schools. There are 5 NETCO schools with propensity scores greater than 0.89 that lack common support. The average propensity score is 0.60 for the i3 schools and 0.04 for all non-i3

---

schools before matching, a difference of 0.56. The average propensity score for the matched comparison schools is 0.26, which reduces the difference in average propensity scores from the i3 schools to 0.26.

Table 1 illustrates the results from tests of the statistical significance of differences in characteristics of the NETCO schools and the non-NETCO schools before and after matching. To determine if the differences between the treatment and control groups are statistically significant, a series of t-tests are conducted for all the covariates in the propensity score model. We report any differences that are statistically significant at the 95 percent confidence level using a two-tailed test.

Before matching there are statistically significant differences between the i3 schools and the non-i3 schools in observed school characteristics representing demographic characteristics of the student population, test scores of the student population, school resources, and community characteristics. Among the student demographic characteristics, there are fewer minority students in i3 schools compared to the non-i3 schools (5 percent versus 27 percent) and fewer students with individualized education programs (7 percent versus 8 percent). For the test score of the student population, the i3 schools have higher scores than the non-i3 schools on the ACT reading, ACT science, and EOC Algebra assessments. When comparing school resources, the i3 schools have total expenditures per pupil that are $522.56 lower than the non-i3 schools. Community characteristics also differ in the i3 schools with fewer adults with a college degree (16 percent versus 21 percent), more of the population below the poverty line (15 percent versus 13 percent), and lower median annual income (a difference of $6,664.17). There are also more i3 schools in suburban areas (28 percent) than in the comparison group (10 percent). After matching there are no statistically significant differences between the i3 schools and the matched comparison schools on any of the observed school characteristics.

One limitation of using t-tests is that differences between groups may not be identified when the sample size is small. In order to further examine whether there are any differences between the i3 schools and matched comparison groups, we examine the absolute standardized bias on each of the covariates in the propensity score model. This is calculated using the equation:

$$\text{Absolute standardized bias} = \left| \frac{\overline{X}_t \quad \overline{X}_c}{\sigma_t} \right| \quad (3)$$

where $\overline{X}_t$ is the mean of the treatment group, $\overline{X}_c$ is the mean of the control group, and σ_t is the standard deviation of the control group. The absolute standardized bias is a diagnostic of the balance between the treatment and control group on the covariates of interest.

One way to visual changes in the balance of the individual covariates is to graph the absolute standardized difference of means before and after matching (Ridgeway, McCaffrey, & Morral, 2006; Stuart, 2007; 2010). Figure 3 shows the change in bias for each of the 26 covariates in the propensity score model. The greatest change is for the percent minority covariate, which has an absolute standardized bias of 4.19 before matching and 0.05 after matching. The two other covariates with an absolute standardized bias greater than 1.0 before matching are median annual

income in the community (reduced to 0.13 after matching) and percent of students with an individualized education program (reduced to 0.00 after matching).

Stuart (2007) recommends that absolute standardized bias values greater than 0.50 are "particularly problematic", and should ideally be less than 0.25. Before matching there are eleven covariates with an absolute standardized bias greater than 0.50 and six covariates with a value between 0.25 and 0.50. After matching, there are no covariates with an absolute standardized bias greater than 0.50. However, the following covariates have an absolute standardized bias between 0.25 and 0.50 after matching: EOC English score (0.28), and rural school locale (0.34).

Conclusions:

Overall, the preliminary propensity score analysis for the NETCO evaluation is able to substantially reduce differences in observable characteristics between the i3 schools and non-i3 schools in the state. There are no statistically significant differences in t-tests of the covariates between the treatment group and the matched comparison group, and the absolute standardized bias is less than 0.5 for all covariates. The presentation will also review several techniques used by the researchers to refine the propensity score model. These include making several changes to the covariates, identifying influential covariates, and imposing a caliper on the matches.

Tennessee State Collaborative on Reforming Education (SCORE), which has a strong statewide reputation across the state, will work with the Niswonger Foundation to secure the participation of control schools identified through the final propensity scoring model. Control schools will be given $15,000 over the course of the grant ($3,000 per year) to provide the appropriate data. An update on the recruitment of schools will be available at the time of the presentation.

After the control schools have been recruited, a second round of propensity score matching will be conducted at the student-level to match each student in the treatment group with a similar student in the control group. Of critical importance will be ensuring that there is baseline equivalence among treatment and control students on the EXPLORE or PLAN assessment for the two cohorts to be followed in the confirmatory impact analysis (students in grades 9 and 10 in the 2010/11 school year). The EXPLORE and the PLAN are part of ACT's Educational Planning and Assessment System (EPAS) and are used to measure student achievement on English, mathematics, reading, and science. According to ACT (2010), "the EXPLORE and PLAN benchmark scores are indicative of probably readiness for college-level work by the time the student graduates from high school" (p. 3). EPAS scores have also been used to predict students' grades in college courses corresponding to the subjects tested (Allen & Sconing, 2005; Coyle & Pillow, 2008; Noble & Sawyer, 2002). This suggests that EXPLORE and PLAN scores are important predictors of college enrollment and persistence, two outcomes which will be examined in the confirmatory impact analysis.

Appendices

Appendix A. References.

ACT Inc. (2010). "College readiness standards for EXPLORE, PLAN, and the ACT," ACT, Inc., Iowa City, IA.

Allen, J. and Sconing, J. (2005). "Using ACT assessment scores to set benchmarks for college readiness," ACT, Inc., Iowa City, IA.

Coyle, T.R. and Pillow, D.R. (2007). SAT and ACT predict college GPA after removing g*. *Intelligence, 36* (6): 719-729.

Noble, J. and Sawyer, R. (2002). "Predicting different levels of academic success in college using high school GPA and ACT composite scores," ACT, Inc., Iowa City, IA.

Ridgeway, G., McCaffrey, D., and Morral, A. (2006). "Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the TWANG package," RAND Corporation, Santa Monica.

Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25* (1): 1-21.

Stuart, E. (2007). Estimating causal effects using school-level data sets. *Educational Researcher, 36* (4): 187-198.

Appendix B. Tables and Figures

Figure 1. Propensity score histogram before matching for non-i3 schools and i3 schools

Propensity Score Histogram (Before Matching). Graphs by i3. Dependent variable is i3 school, 0=no, 1=yes
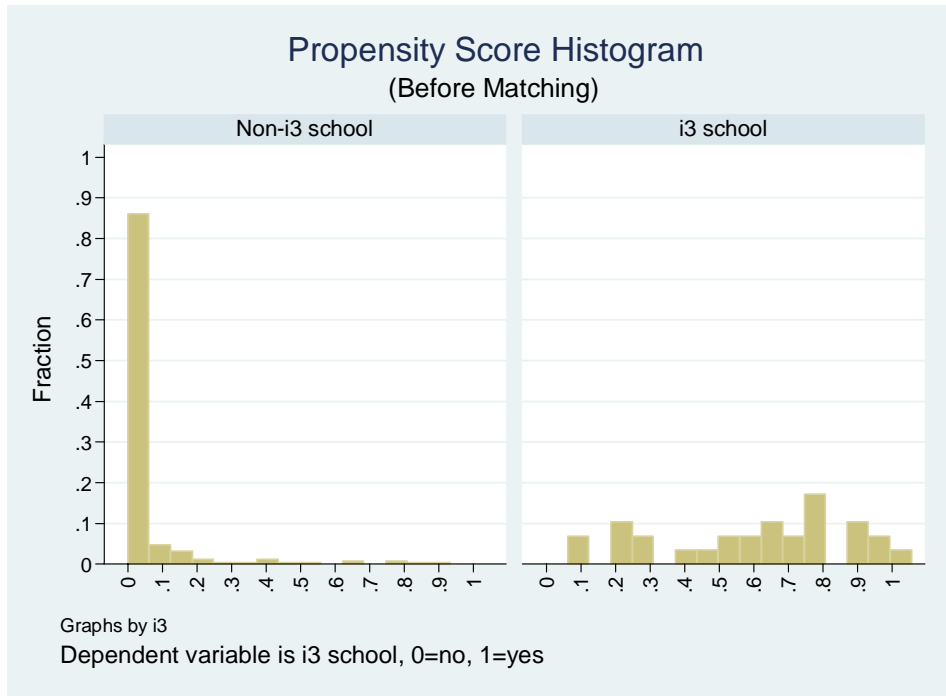
Figure 2. Propensity score histogram after matching for non-i3 schools and i3 schools

Propensity Score Histogram (After Matching). Graphs by i3. Dependent variable is i3 school, 0=no, 1=yes
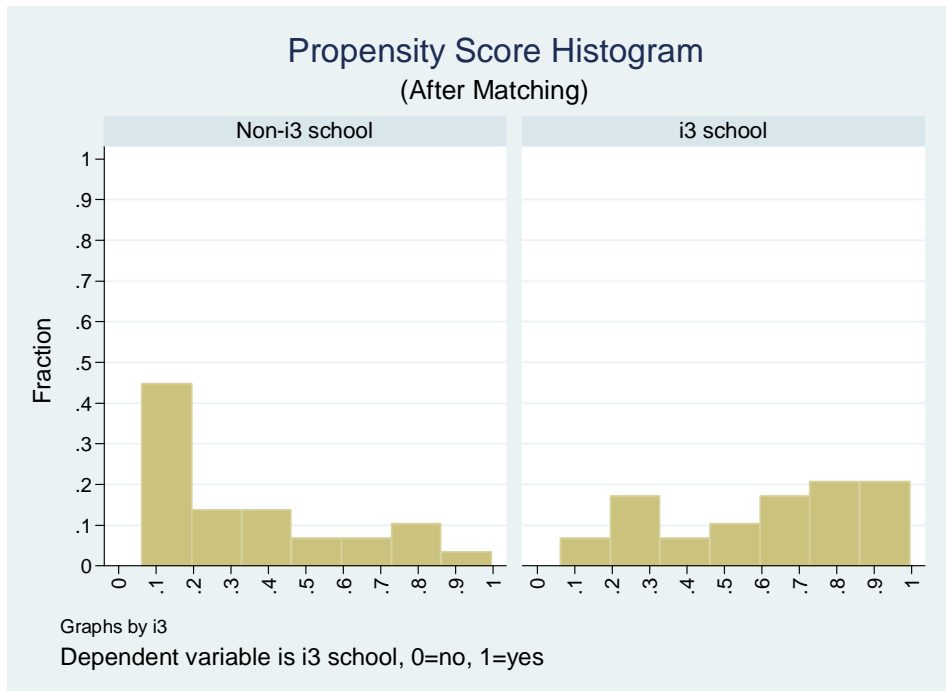
Figure 3: Absolute standardized bias of the 25 covariates in the propensity score model, before and after matching
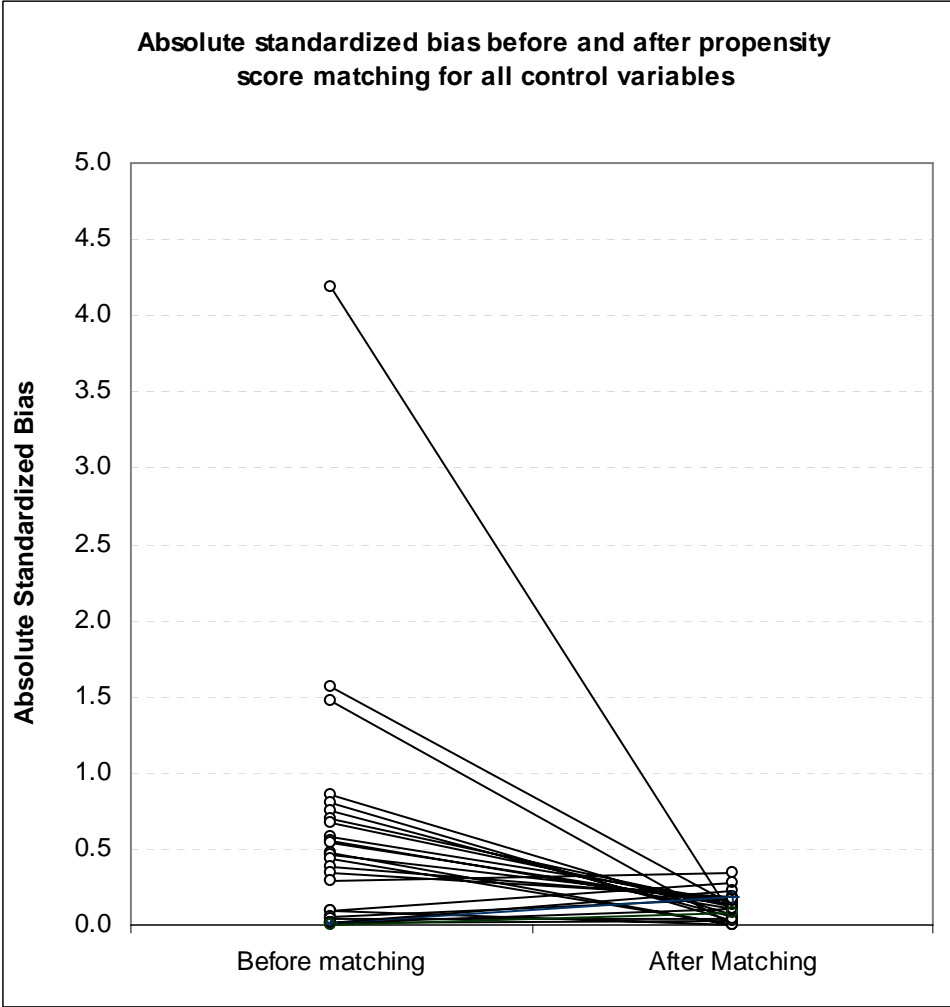
Table 1: Comparison of the baseline characteristics between the i3 schools and control schools before and after matching

	NETCO Schools		All Non-NETCO Schools				Matched Comparison Schools			
	Mean	Std. Dev.	Mean	Std. Dev.	Difference	p-value	Mean	Std. Dev.	Difference	p-value
Student demographics										
% Minority	0.05	0.05	0.27	0.31	-0.21 **	<0.01	0.05	0.05	0.00	0.85
% Male	0.52	0.02	0.51	0.02	0.01	0.17	0.52	0.02	0.00	0.53
% Free or reduced priced lunch	0.55	0.18	0.54	0.21	0.01	0.82	0.52	0.14	-0.03	0.54
% Limited English Proficiency	251.43	177.16	175.50	339.55	75.93	0.24	0.01	0.01	0.00	0.99
% Individualized Education Program	7.15	0.79	8.30	1.45	-1.16 **	<0.01	0.07	0.02	0.00	0.95
Test scores										
ACT English score (3yr avg)	20.36	1.19	19.70	2.17	0.66	0.11	20.49	1.42	0.13	0.70
ACT Math score (3yr avg)	19.55	1.41	19.02	1.72	0.53	0.11	19.80	1.39	0.24	0.51
ACT Reading score (3yr avg)	21.02	1.05	20.13	2.03	0.89 *	0.02	21.10	1.38	0.07	0.82
ACT Science score (3yr avg)	20.26	0.89	19.64	1.59	0.63 *	0.04	20.38	1.25	0.12	0.67
EOC Algebra score (3yr avg)	538.71	11.65	529.37	16.43	9.34 **	<0.01	538.99	11.31	0.28	0.93
EOC Biology score (3yr avg)	542.04	11.58	536.66	18.25	5.38	0.12	543.89	12.09	1.85	0.55
EOC English score (3yr avg)	533.07	9.70	532.15	13.15	0.92	0.72	535.73	7.04	2.67	0.24
EOC History score (3yr avg)	522.82	7.97	518.50	13.21	4.32	0.09	521.79	7.78	-1.03	0.62
CTE and AP										
Number of CTE program areas	6.14	1.43	6.00	1.04	0.14	0.51	6.14	1.03	0.00	1.00
% graduates with CTE concentrations	0.40	0.21	0.40	0.20	0.00	0.94	0.36	0.20	-0.04	0.48
Number of AP courses	4.55	4.99	4.47	5.20	-0.08	0.93	4.03	5.02	-0.52	0.70
Ratio of enrollments in AP to grades 11-12	0.22	0.25	0.23	0.36	0.01	0.88	0.21	0.30	-0.01	0.87
Attendance & Graduation										
Attendance rate (% days present)	0.93	0.02	0.93	0.03	0.00	0.97	0.94	0.01	0.00	0.23
Graduation rate	0.89	0.06	0.86	0.11	0.04	0.09	0.91	0.05	0.01	0.39

	NETCO Schools		All Non-NETCO Schools				Matched Comparison Schools			
	Mean	Std. Dev.	Mean	Std. Dev.	Difference	p-value	Mean	Std. Dev.	Difference	p-value
School resources										
Student/teacher ratio	16.47	2.75	16.45	2.28	0.01	0.97	16.66	2.76	0.20	0.79
Total expenditures per pupil	8031.62	782.36	8554.19	1083.41	-522.56 *	0.01	8122.44	928.61	90.83	0.69
School size (enrollment in grades 9-12)	974.69	547.10	982.99	516.39	-8.30	0.94	963.79	539.22	-10.90	0.94
Community characteristics										
% population w/college degree	0.16	0.06	0.21	0.12	-0.05 *	0.04	0.16	0.10	0.00	0.99
% population below poverty line	0.15	0.04	0.13	0.05	0.02 *	0.04	0.15	0.04	0.00	0.73
Median annual income	36798.48	4276.11	43462.65	10597.61	-6664.17 **	0.00	37339.83	5882.07	541.34	0.69
Locale=rural	0.38	0.49	0.53	0.50	-0.15	0.14	0.55	0.51	0.17	0.19

* The difference is statistically significant at the 95 percent confidence level using a two-tailed test.

Note: The p-values are based on a t-test of the difference between NETCO schools and groups of comparison schools