ACHIEVEMENT OUTCOMES OF PROJECT LEAD THE WAY: A STUDY OF THE IMPACT OF PLTW IN IOWA

Tom Schenk Jr.[1]

David Rethwisch[2]

Melissa Chapman[3]

Frankie Santos Laanan[4]

Soko S. Starobin[5]

Leaf Zhang[6]

WORKING PAPER

September 29, 2011

[1] Iowa Department of Education, Division of Community Colleges & Workforce Preparation tom.schenk@iowa.gov
[2] Department of Chemical and Biochemical Engineering, The University of Iowa, david-rethwisch@uiowa.edu
[3] Professional Data Analysts, MChapman@pdastats.com
[4] Department of Educational Leadership & Policy Studies, Iowa State University, laanan@iastate.edu
[5] Department of Educational Leadership & Policy Studies, Iowa State University, starobin@iastate.edu
[6] Department of Educational Leadership & Policy Studies, Iowa State University, lyzhang@iastate.edu

The authors would like to thank Xiaoping Wang, Diane Chadwick, Mary Grinstead, and Peter Orazem who provided invaluable feedback. We would also like to thank participants of the American Educational Research Association 2010 in Denver, Colorado, Association of Institutional Researchers 2010 in Chicago, Illinois, P-12 Summit in Seaside, Oregon and Evaluation 2010 in San Antonio, TX. This research was partially funded by the Iowa Math and Science Educational Partnership and the Kern Family Foundation. All opinions expressed in this document are solely those of the author's and may not necessarily reflect those of the Iowa Department of Education, Iowa State University, University of Iowa, nor the State of Iowa.

Abstract

Background

Iowa has implemented the secondary engineering curriculum Project Lead The Way (PLTW) in an effort to create a more seamless transition for students from secondary school into science, technology, engineering, and mathematics post-secondary programs. PLTW has been implemented in all fifty states; however, there has been sparse research to-date that has rigorously measured the impact of PLTW on mathematics and science achievement.

Purpose (Hypothesis)

Our first purpose is to report on the socio-demographic, academic, and achievement of PLTW students. Second, we hypothesize that cognitive improvement for PLTW students is greater than that for non-PLTW students.

Design/Method

We used Iowa's statewide longitudinal data system to follow multiple cohorts of PLTW participants and nonparticipants from 8th grade into higher education. We derived a comparable treatment and control group by matching students based on their propensity to enter PLTW, permitting a stronger interpretation of the program's impact than prior studies.

Results

We found that PLTW enrolls a disproportionate number of white males and students who performed in the upper quartile in mathematics and science prior to PLTW enrollment. Further, we found statistically significant evidence that PLTW increases mathematics or science scores

on the Iowa Test of Educational Development by 5 points after controlling for selection bias. The effect size (f^2) for mathematics was 0.15 and 0.05 for science—a moderate and small effect size, respectively.

Conclusions

PLTW attracts high-achieving students; however, after controlling for this using propensity score matching we found participation in PLTW produces a small or moderate effect on test score growth. Further studies will also need to properly account for pre-existing ability in mathematics and science when determining achievement outcomes to ensure results are not being driven by pre-existing ability.

1. Introduction

Nationally, the United States is facing an international decline in performance in mathematics and science. In 2003, the Program for International Student Assessment (PISA), found that nearly 30 percent of U.S. college students required remedial mathematics and sciences courses in their first year of college. Similar results were found by the 2003 International Mathematics and Science Study (TIMSS). According to Kuenzi (2008) American students, while outscoring the international average in mathematics and science, still scored lower in composite scores than nine other developed countries.

Some scholars assert that international comparisons may not be justified since test results may not accurately reflect other, namely rural, parts of the country (Wildavsky, 2011; Rotherham, 2011). Even so, if we take a step back and analyze how American students are performing according to our own benchmarks, we find similar shortcomings. While proficiency in mathematics skills among 8th graders has increased on the National Assessment of Educational Progress (NAEP), basic skill levels have remained unchanged over the same time period. Twelfth-grade tests results have actually declined since 1990, with fewer students scoring as either proficient or basic in their mathematics skills. In addition, anywhere from 10 to 35 percent of all American students score under the basic skill-level in mathematics. In 2005, nearly a third of all students tested below basic skill levels in mathematics. In essence, a significant portion of American students are failing to achieve the standards set by American educators, much less remain competitive with their international peers.

One reason for this decline may be that the traditional teaching methods for mathematics and science have come under scrutiny. Often, the "chalk-and-talk" teaching model has become associated with declining student engagement with mathematics and science (Mills & Treagust, 2003). School leaders and administrators have sought to engage students through hands-on

instruction. Problem-based learning (PBL) has become a growing field of interest, especially within the science, technology, engineering, and mathematics (STEM) fields. The instructional approach offers students the opportunity to engage a topic in "real-world" circumstances, where students are presented ill-structured problems which require critical thinking and reasoning skills to solve (Barrows & Kelson, 1995; Hmelo-Silver, 2004).

Project Lead the Way (PLTW) is a fast-growing PBL program emerging in high schools. Nationally, PLTW has been implemented in all 50 states and enrollment has grown to more than 350,000 students to 17% of all high schools (Misko, 2011). PLTW's explicit goal is to increase student interest in engineering and other STEM-related majors and careers (PLTW, 2006; Maguire 2006).

The curriculum provides foundational instruction in pre-engineering and design while also providing specialized instruction in a range of specific topics from aerospace engineering to biotechnical engineering. Each course is hands-on and provides the opportunity for students to, for example, work with computer-aided drafting software or build mechanical robots.

In addition, PLTW provides mandatory professional development to teachers, counselors, and administrators. Each professional development course is team taught by a university professor and master teacher. The goal of these opportunities is to ensure all PLTW teachers are proficient with the curriculum.

While PLTW has multiple goals, several papers have begun to explore the relationship between PLTW participation and achievement on standardized exams (Blais & Adelson, 1998; Bottoms & Anthony, 2005; Bottoms & Uhn, 2007; Rogers, 2006; Tran & Nathan, 2009, 2010; and Walcerz, 2007). However, few of these studies accounted for pre-existing ability prior to enrolling in the PLTW curriculum. Since most states provide PLTW as an optional curriculum, it

is likely that the curriculum attracts students who are already interested in STEM courses. Since pre-existing achievement often portends future performance, it is also likely that the reported differences between PLTW participants and non-participants are not entirely attributable to the program itself.

This paper attempts to control for pre-existing achievement levels by using administrative data from Iowa's statewide longitudinal data system (SLDS). Iowa tests all students in 8th grade in mathematics and science, which is the year prior to when PLTW is available to students. Iowa then tests all students again in 11th grade as part of their No Child Left Behind obligations.[7] We used propensity score matching to derive a smaller, but balanced dataset of comparable PLTW participants and non-participants. We used the balanced treatment and control groups to measure PLTW's impact on the relative growth of science and mathematics scores between 8th and 11th grade.

2. PLTW Program Description

In 2005 Iowa implemented the PLTW curriculum to create a seamless transition for students from secondary school into STEM majors at two- and four-year postsecondary institutions (Maguire, 2006). Since then, enrollment has grown to over 2,000 students and it is currently offered in 101 of 260 Iowa school districts. The increase in Iowa's enrollment coincided with growth of PLTW nationwide (e.g., Brandt, 2009; Cech, 2007; Spellman, 2007).

The program is a sequence of year-long courses designed to teach engineering and problem solving concepts to high school students. The curriculum is divided into two strata (Taylor, Foster, & Ratcliff, 2006)—foundation courses (Introduction to Engineering Design and Principles of Engineering) and specialization courses (Aerospace Engineering; Biotechnical

[7] Iowa, however, has been using standardized exams in elementary, junior high, and high schools for several decades prior to the No Child Left Behind legislation.

Engineering; Civil Engineering and Architecture; Computer Integrated Manufacturing; and Digital Electronics). The sequence of courses ends with a capstone course (Engineering Design and Development) which requires students to take their own idea from design through development. In addition, Gateway to Technology is offered in middle school in selected school districts.

The PLTW curriculum is optional in Iowa school districts as it is in most states. Each course is one full Carnegie unit (e.g., full year) and is offered to anyone between ninth and twelfth grade. The curriculum requires students to enroll in mathematics and science as pre- or co-requisite courses in conjunction with the PLTW curriculum. PLTW is open to any student who meets the minimum requirements, but Misko (2011) and other PLTW administrators have noted the program is generally targeted toward the top 80% of a school's population. In addition, PLTW courses can qualify students for high school and college credit. Students are also encouraged to enroll in college preparatory mathematics, biology, chemistry, and physics courses.

The foundation courses provide an overview of engineering and introduce students to various engineering aspects, such as design, and manufacturing processes. These courses involve students learning 3-D computer modeling, designing and reverse engineering objects, applying the fundamentals of physics, and using electronics and computer programs to build robotic machines. Specialization courses allow students to explore a specific engineering discipline in more detail.

The PLTW courses offer projects which would seem engaging for a variety of students. For example, one segment of the curriculum requires students to design and build a small mechanical robot that sorts marbles made from various materials (such as metal, wood, and

glass) into bins to mimic sorting of recyclable materials. The marbles are sorted based on their opacity, which is determined by shining a light on the marble and determining the amount of ohms received by a sensor on the other side. Students utilize software to adjust the sensitivity of the sensor, which is crucial for performance.

PLTW's curriculum contains detailed daily lesson plans and is disseminated through rigorous professional development courses. All PLTW teachers must attend a two week summer training institute for *each* course to be taught (cf. Taylor et al., 2006). The summer training involves a university professor and experienced PLTW teacher (master teacher) for both theory and application with a heavy emphasis on the pedagogical approach of project based learning. In Iowa, most of the training originates from engineering departments at The University of Iowa or Iowa State University. Eventual teachers learn the same software, theory, and applications that their students will use.

3. Literature Review

PLTW is an interesting STEM program to evaluate given the recent, rapid growth across the United States, including Iowa. Further, it is a highly structured PBL curriculum with a strong focus on STEM, and potential PLTW teachers must participate in a rigorous professional development model. The role of PLTW's two week professional development may be crucial as the teacher (facilitator) is regularly identified as a crucial role in PBL curriculums (Hmelo-Silver, 2004; Holmes & Kaufman, 1994).

Several studies have examined outcomes of students in PBL curriculum. Most of these studies have focused on medical schools, where PBL grew in popularity throughout the 1960s (Jonassen, 2000; Spector, 2004; Hmelo-Silver, 2004). In randomized studies, students in PBL curriculum scored lower on assessments which utilized multiple choice responses and focused on fact-based knowledge (Albanese & Mitchell, 1993; Dochy et al., 2003; Goodman et al., 1991;

Mennin et al., 1993; Vernon & Blake, 1993). Students scored better on assessments which related to problem solving, clinical performance, and explanation of clinical problems (Albanese & Mitchell, 1993; Hmelo, 1998; Hmelo Grotter, & Bransford, 1997; Patel, Groen, & Norman, 1991, 1993; and Vernon & Blake, 1993). The totality of the evidence suggests students in PBL curriculum perform better on assessments related to problem solving and interpretation, but are weaker on assessments related to factual knowledge.

However, some scholars claim PBL can have adverse effects on student learning. Kirschner, Sweller, & Clark (2006) argue that problem-based learning does not develop long-term memory, and thus, is ineffective as soon as the student leaves the classroom. Brown & Campione (1996) found instruction with minimal guidance, such as problem-based learning, can let student's misconceptions be left uncorrected.

In addition to medical schools, PBL curriculum has expanded into engineering schools. Williams & Williams (1994) noted that instruction in engineering design had already been very similar to PBL practice and could easily be accepted amongst engineering faculty. Mills & Treagust (2003) reviewed several PBL curriculums in Australian universities noted the results in engineering are similar to those findings from medicine in that students show a lower grasp of engineering fundamentals, but understand "the application of their knowledge in practice and the complexities of other issues involved in professional practice" (p. 12) .

PLTW is a more recent engineering PBL program, and although numerous studies have attempted to explore the impact of PLTW on various educational outcomes (Blais & Adelson, 1998; Bottoms & Anthony, 2005; Walcerz, 2007), a serious limitation of these studies is the lack of control for pre-existing ability. However, some researchers have begun to address this issue. A research brief by the Southern Regional Education Board (SREB) matched PLTW participants

with career and technical education students with similar demographics and fields of study. SREB found that PLTW students who enrolled in two or more PLTW courses did significantly better in mathematics and science on the High Schools that Work (HSTW) assessment than career/technical students in comparable fields (Bottoms & Anthony, 2005). Differences between PLTW students and similar career/technical students were also found for subsequent course-taking behavior, with PLTW students more likely to complete the four years of mathematics and science (Bottoms & Uhn, 2007). Yet, SREB's study was limited to matching on students' race and gender.

In a follow-up to an earlier study, Tran & Nathan (2010) collected transcript data from a school district in Wisconsin which was heavily represented by racial minorities as well as a high proportion of students eligible for free or reduced lunch prices. Their study matched PLTW participants with nonparticipants using results of prior achievement scores in mathematics and science, gender, and free and reduced lunch eligibility. Tran & Nathan measured relative change of mathematics and science scores between the state-mandated tests, a part of the No Child Left Behind requirements between 8th and 10th grade. They found that PLTW had no measurable impact on science scores while PLTW participants actually scored lower in mathematics compared to similar students.

It is important to note that there were some limitations in the Tran and Nathan study, such as the short time frame between the two standardized assessments. The structure of the study only permitted students to realistically enroll in a single class, although a very small number enrolled in two courses. Additionally, the paper's focus was on a single school district with a limited sample size. Yet, the study supported crucial questions about the integration of math in the PLTW curriculum. In a curriculum review, Prevost et al. (2009) and Nathan et al. (2008)

found some, but minimal alignment between PLTW curriculum and standards set by the National Council of Teachers of Mathematics.

Following Tran and Nathan's prior work, we test three hypotheses in this study. First, the *enriched integration hypothesis* proposes that enrollment in PLTW will increase cognitive performance in mathematics and science standardized exams. The integration between PLTW, mathematics, and science content in a PBL context means students will better understand that information and demonstrate that knowledge in achievement assessments. Under this hypothesis, PLTW students would have greater growth in cognitive development for mathematics and science between 8th and 11th grade as compared to non-PLTW students.

An alternative hypothesis is the *insufficient integration hypothesis*, where mathematics and science content is insufficiently integrated into the PLTW curriculum. In this scenario, the PLTW treatment would not lead to any relative improvement in cognitive development for mathematics and science.

A review of the PBL literature indicates a potential third hypothesis where curriculum could "lead to changes in attitude, confusion, or even misconceptions that hinder student performance" (Tran & Nathan, 2009, p. 10). Prior studies, such as Brown & Campione (1996), suggest inadequate guidance in PBL curriculum can let student's misconceptions be left uncorrected and lead to lower relative growth between 8th and 11th grades.

4. Data

The Iowa Department of Education maintains a longitudinal data set that tracks students through high school into college and the workforce. Since 2005, each student has been assigned a unique student identifier, which is retained as they progress through high school, including if they transfer to any other secondary institution within Iowa. Figure 1 shows the path of students from middle school through the workforce. Our focus in this paper is the short-term impact on

high school achievement tests, so the main source of data will be limited to the secondary school data set.

[FIGURE 1 ABOUT HERE]

Iowa's administrative records also contain socio-demographic data, achievement scores in the area of mathematics, science, and reading, institutional-level factors, and course enrollment information. Socio-demographic information includes the student's race/ethnicity, gender, eligibility for free and reduced lunch (an underreported proxy for economic status, Harwell & LeBeau, 2010), and whether the student is homeless. The data set contains scores from the Iowa Test of Basic Skills (ITBS) for 8th grade and Iowa Test of Educational Development (ITED) for 11th grade, including scores by subject exam (reading, science, and mathematics). It also contains information on whether students have a special need, indicated by a Section 504 disability or individual education plan (IEP). Finally, the school district is recorded, which we can use to control for institutional-level factors.

We limited our analysis to all students enrolled in school districts offering PLTW, which yielded over 35,000 students. Since the focus of this paper is the relative growth in test scores between the 8th grade ITBS and 11th grade ITED, the timeframe of the data means we had to limit our data to the class of 2009 and the class of 2010. Further, students were only included if their records were available in 8th and 11th grade. Consequently, the sample size included 26,030 students.

PLTW participants were identified through course records. All PLTW programs are assigned a program number using the Classification of Instructional Programs (CIP), which indicates a list of PLTW courses offered at each school district. Students were considered a PLTW participant if they enrolled in at least one PLTW course in either 9th or 10th grade. For the

purposes of this study, nonparticipants are students who did not enroll in a PLTW course in 9th or 10th grade, but were enrolled in a school district that offered the program.

Table 1 shows the descriptive statistics from the PLTW dataset for the 2009 and 2010 cohorts. Total enrollment was 1,321 students, compared to 24,709 students in the control group. The data strongly suggests the presence of selection bias. Participants were disproportionally white males compared to nonparticipants. Eighty-five percent of the PLTW participants were male and 91 percent of the participants were white. By contrast, 49% of non-participants were male 80% were white. The proportion of females entering PLTW, however, is approximately the same amount of females entering mechanical and electrical engineering programs nationwide (Walcerz, 2007).

[TABLE 1 ABOUT HERE]

Further, the economic status proxies indicated PLTW participants were less likely to come from low-income families. Sixteen percent of participants were eligible for free or reduced lunch, compared to 32 % of the control group. PLTW participants were also twice as likely to be identified as a part of a gifted & talented program (21%) and were seven times less likely to have an IEP (2%).

Finally, we found PLTW participants had higher achievement in mathematics and science than non-PLTW participants prior to entering in the program. Eighth grade mathematics scores show mean mathematics scores were in the 80th percentile, compared to the 60th percentile for nonparticipants. Science scores show a similar pattern. The mean participant score was the 84th percentile compared to the 64th percentile for nonparticipants.

5 Methodology

Selection Bias

Since enrollment in PLTW is not random, we must be control for pre-existing achievement not attributable to the program. Data shown in Table 1 demonstrates that PLTW participants already exhibit higher mathematics and science scores prior to enrolling into PLTW. Thus, a simple comparison between expected values (e.g., mean differences) between the treated and control group is highly biased.

It is important to note that there are two levels of selection bias: for school districts and for students. Most noticeably, school districts that offer PLTW have a larger number of students, and are more likely to be in an urban setting. The average size of a PLTW district in 2007 was 1,830 students compared to the 339 students for non-PLTW districts. While it is possible to control for selection bias with observed variables, such as size, we also know that PLTW districts were also selected on non-observables, such as relationships between PLTW program officers and administrators and the willingness of administrators to pursue "innovative" programs. In order to limit bias from school district-level selection bias, the control group consists only of students from districts offering PLTW, but who did not enroll in any PLTW courses.

In addition to school district selection bias, we have already highlighted data which suggest selection bias at the student level. We can determine if there is any selection bias by estimating the likelihood of entering PLTW while conditioning on data prior to enrollment. We use socio-demographic and testing data from 8th grade—the year prior to any PLTW enrollment—to calculate the conditional probability of enrolling in PLTW in 9th or 10th grade given the observed covariates (p. 296, Rosenbaum, 2002)

$$\rho(\mathbf{X}) = E(\tau = 1|\mathbf{X}) = \varphi(\alpha_i + \overleftarrow{\beta}\mathbf{X} + \epsilon_i) \quad (1)$$

Where $\rho(\mathbf{X})$ is the propensity score, τ indicates treatment (enrolling in PLTW = 1), $\boldsymbol{X}$ is a matrix of historical socio-demographics (gender, race, free/reduced lunch, and homeless status) and testing data (reading, mathematics, and science) from 8th grade—a year prior to PLTW entry—and $\varphi(\bullet)$ is the logit function.[8]

Matching

With selection bias, the distribution for the outcome of interest (Y, Y') is unequal across the treatment variable (τ'). Specifically, Table 1 shows the outcome of interest (11th grade test scores) are much higher for participants, but participants are also more likely to be male, white, high-achievers in mathematics and science, are more likely to be in gifted and talented programs, and are less likely to be eligible for free or reduced lunch. The unequal distribution or "imbalanced" data set is formally noted by:

$$Y, Y' \not\perp \tau | \mathbf{X}$$

Rosenbaum and Rubin (1983, 1984) prove that the distribution of outcomes can be balanced using propensity scores. Namely, students with similar propensity scores, $\rho(\boldsymbol{X})$, are matched while unmatched students are discarded from the data. Rosenbaum and Rubin (1983, 1984) demonstrate that

$$Y, Y' \perp \rho(\mathbf{X})$$

where $\rho(\mathbf{X})$ is derived from equation 1.

We can estimate the impact of PLTW on student outcomes by comparing the expected values of the treatment and control group once the propensity of PLTW entry, $\rho(\mathbf{X})$, is estimated. Specifically,

[8] This leaves the question if participation was determined by an unobserved variable. This cannot be answered with the data and is a weakness of propensity score analysis and we must rely on the stable unit treatment value assumption (SUTVA), which ignores any unobserved covariate that affects the probability of treatment. Some of this concern is mitigated since unobserved covariates are likely correlated with observed variable, thereby limiting the reliance on SUTVA (Stuart, 2009).

$$\mathrm{E}[Y|\tau = 1, \rho(\mathbf{X})] - \mathrm{E}[Y'|\tau = 0, \rho(\mathbf{X})]$$

can be denoted as the average treatment effect.

The quality of matching is dependent upon the matching algorithm. Figure 2 illustrates the trade-off between matching algorithms. Imbalanced data is often present in large observational data sets, but provides biased estimates of the regression coefficients. Matching will reduce bias to various degrees. Some matching algorithms may retain large sample sizes, but only provide modest balance. Stricter matching algorithms can provide precise balance, but greatly reduce the sample size, which could increase the incidence of Type II errors.

[FIGURE 2 ABOUT HERE]

We used three different matching algorithms to match students. The most common matching algorithm is "nearest neighbor", where the treated student (PLTW participant) is matched with the nearest untreated student. Nearest neighbor is a flexible technique because it allows the modeler to impose additional criteria for the match. We forced matched students to be from the same school district and we also forced a single control student to be matched with only one PLTW participant.

One disadvantage to this matching method is that it only chooses a minimal distance between each treatment and control, but does not minimize the total distance. Since we are trying to achieve overall balance, a better approach seeks to minimize the total difference between treatment and control groups.

We also chose to use genetic algorithms to find the minimal distance between treatment and control units. Genetic algorithms are based on the principles of population biology which use selection, recombination, and mutation of estimates to derive optimal solutions (Duffy, 2006; Holland, 1975). These algorithms are computationally more efficient than other algorithms to

find solutions to optimization problems. In this case, we used genetic algorithms from Ho, Imai, King, & Stuart (2011) to minimize the distance (difference) between the treatment and control group.

The distance between these two groups are defined by the generalized Mahalanobis distance measure (Diamond & Sekhon, 2010). The genetic algorithms search to find an optimal mix of students and weights to minimize the difference in characteristics between these two groups. Once an optimal solution is derived, the matched students and corresponding regression weights are used in the subsequent analysis.

We matched one participant to one nonparticipant and also matched one participant to two nonparticipants with genetic algorithms. All three matching methods—nearest neighbor, one-to-one matching with genetic algorithms, and one-to-two matching with genetic algorithms—were used in this study.

Impact on Test Scores: Regression Model

Since propensity score matching does not deliver estimates of program impact, we derived a parametric model of PLTW's impact on test scores between 8th and 11th grade using a difference-in-differences (DID) estimator. Students in Iowa typically have the option of taking the foundational PLTW courses beginning in 9th grade. Thus, we can use the 8th grade test scores as a baseline, pre-program measure of students' mathematics and science achievement prior to the program, and the 11th grade test scores as an indicator of change in achievement in relation to PLTW participation.

A participant realizes a pre-treatment test score in 8th grade $Y(\tau,8)$ and post-treatment test score in 11th: $Y(\tau,11)$. Meanwhile, a non-participate realizes the 8th grade test score, $Y(\tau',8)$, and 11th grade score, $Y(\tau',11)$, without treatment in the intervening period. The first difference is the

respective growth *within* participants and non-participants, that is, Y(τ',11)-Y(τ',8) and Y(τ,11) – Y(τ,8), respectively. We are interested if participants see a larger growth in test scores, that is, Δ = [Y(τ',11) - Y(τ',8)] - [Y(τ,11) – Y(τ,8)]. The parameter, Δ, is the relative gain for participants against nonparticipants.

Even though the data set contains individual information, students are grouped within school districts where performance is likely correlated. The relationship can be dichotomized into two levels: the first consists of individual students who are nested in the second level, the school districts themselves. Ordinary least squares (OLS) regression should not be used on nested data sets since it can increase the risk of committing Type I errors (Kreft and de Leeuw, 1998; Barcikowski, 1980; Steenbergen and Jones, 2002). An OLS approach will overestimate the degrees of freedom since each student will be counted as an independent observation. In reality, the error term for each student observation within a district is correlated. Second, standard errors will be underestimated when student observations are correlated within each district (Goldstein, 2002). As a result, both of these conditions will cause statistical significance to be overstated (see Schenk, 2007, Appendix A for a formal proof).

We use a multilevel model which accounts for student variation within each school district (Gelman & Hill, 2006). Equation 2 allows the intercept and slope to vary by the jth school district:

$$Y_{ij} = \left(\alpha_j + u_j\right) + \beta_1 T + \beta_2 \tau + \Delta\tau_{\mathrm{j}} \cdot \mathrm{T_j} + \sum_{\mathrm{k=3}}^{\mathrm{m}} \beta_{\mathrm{k}} \mathrm{x_k} + \epsilon_{\mathrm{ij}} \qquad (2)$$

The parameter of interest, Δ, represents the relative gain (or loss) of PLTW students compared to their matched peers. We are not making any direct comparison between the 8th grade ITBS and 11th grade ITEDS, which is an invalid use of the standardized scores. Equation 2 estimates a

slope for the gain in test scores for participants and another slope for the control group. The difference between these slopes is summarized in the Δ coefficient.

There is a pair of error terms, the first, u_j, is an error term for the school districts intercept and second is the total error term, ε_{ij}, is modified to incorporate student- and school-specific residuals. The error terms, u and ε, have a standard normal distribution with a zero mean, i.e.:

$$u_j \sim N(0, \sigma_u)$$

$$\epsilon_j \sim N(0, \sigma_\epsilon)$$

Subsequently, the expected value and variance of the regression can be written:

$$E(Y_{ij}) = \alpha_i + \beta_1 T + \beta_2 \tau + \Delta\tau \cdot \mathrm{T} + \sum_{\mathrm{k=3}}^{\mathrm{m}} \beta_{\mathrm{k}} \mathrm{x}_{\mathrm{k}}$$

$$Var(Y_{ij}) = \sigma_u^2 + \sigma_\epsilon^2$$

The regression provides estimates of the fixed and random effects. The fixed effects are the student-specific estimates which are of primary interest in this paper. The random effects are the school-level variation captured by the multilevel model.

6. Results

Propensity Score Matching

Table 2 shows the results of the propensity score estimates. Males, gifted and talented, and ITBS scores on mathematics and science were positively correlated with entry into PLTW, while free lunch eligibility and individualized education plans (IEP) were negatively correlated with PLTW participation.

[TABLE 2 ABOUT HERE]

We generated predicted probabilities for program entry using the coefficients and matched students based on the results. Table 3 shows diagnostic information for each matching

algorithm used in this study. The nearest neighbor matching technique substantially reduced bias from the raw data set while reducing the sample size from 26,030 to 1,500 students. However, there was still some imbalance in the matched group. In particular, the nearest neighbor matching algorithm yielded a data set where participants had average 8th grade test scores which were between 0.1 and 0.15 standard deviations higher than the control group.

[TABLE 3 ABOUT HERE]

We also tested variants of matching using genetic algorithms. One variant matched up to one control student with each participant. The dataset was reduced to 1,477 students, but the mean difference was essentially zero. We also used matching parameters that matched up to two observations for each participant. The balance was similar, near zero, but the total sample size was increased to 2,111 students. Of the various matching algorithms attempted, we found the genetic matching algorithm with a maximum ratio of two non-participants to be matched with each participant produced the best balance between sample size and covariate balance.

Table 4 shows the descriptive statistics for the matched data set using a genetic algorithm with a maximum ratio of two control students for every treatment case. PLTW participants comprise approximately 38 % of the sample, which was almost mostly white (90%) and male (85%). Average test scores are much higher for the matched data set than the full sample (see Table 1). The average 8th grade mathematics score was 284 points and science was 289 points. The demographic profile of the balanced data set now mirrors the typical profile of a PLTW participant demonstrated in Table 1.

[TABLE 4 ABOUT HERE]

Test Scores

Table 5 shows the subsequent estimate of PLTW's impact on mathematics scores. The DID estimator (Δ) shows participants had a relative gain of 5.2 points ($p = 0.01$) on the national scale score. The magnitude of impact does seem to depend on the matching algorithm and the sample size. The match performed using genetic algorithms, but limiting one participant to be matched with one non-participant showed a lower estimate and a lower t-statistic. Likewise, nearest neighbor matching, which also had a small sample size and greater imbalance, produced an estimate which was not statistically significant at any practical threshold. Unless otherwise noted, the remainder of this section will refer to the one-to-two genetic matching since it produces the best mixed of balanced covariance and large sample size.

[TABLE 5 ABOUT HERE]

Table 6 shows the estimated impact of PLTW on science scores to also be 5.2 points ($p = 0.02$) on the national scale score. The point-estimates show PLTW participants gained an average of 36 points ($T + \Delta$) compared to a 30.8 point gain for non-participants (T).

[TABLE 6 ABOUT HERE]

While holding other variables constant, a PLTW participant scored in the 91st percentile in mathematics compared to the 81st percentile for similar nonparticipants by the 11th grade.[9] The gap in science scores is less pronounced. The mean science percentile for a participant is the 83rd percentile compared to the 80th percentile for the control group.

Additional covariates portray other important information that is not directly attributable to PLTW. Demographic covariates show African-American and Hispanic students' score significantly lower than white student on the 8th grade ITBS. Asian students, meanwhile, score

[9] The percentiles are based on national percentile rank. The percentiles are translated from the raw scores by summing the intercept (α), PLTW participant (τ), ITED (T), and interaction term (Δ) coefficients. Thus, these percentiles are being computed for white, female students who are not eligible for free or reduced lunch, nor homeless, nor a member of any special population.

equivalently to white students in both mathematics and science. Males score around 3 points higher than females in mathematics, but both genders had statistically equivalent ITBS scores on the science subtest.

Economic status also impacts test scores with students eligible for free or reduced lunch scoring lower on both subtests. Homeless students had mixed results, scoring lower on the science exam, but having equivalent mathematics scores to non-homeless students. Special populations also had statistically different test performance; students with an IEP or Section 504 plan scored lower on both subtests, the former with a large disparity in scores. Gifted and talented students scored significantly higher on mathematics and science exams.

Effect Size

The results show PLTW leads to a 5.2 increase for mathematics and science scores. Statistical significance indicates that the estimates from the impact are unlikely a statistically anomaly of a program that provides no or even negative impact. The evidence from the 2009 and 2010 cohorts suggest PLTW participation leads to an average increase of 5 points in math and science scores.

However, these results do not yet show the relative effect size—whether they are small or large. Literature on this topic provides several options (cf. Cohen, 1988). The most straightforward calculation of effect size is Cohen's *d*: Δ/σ_{Δ}. That is, the estimated impact of PLTW divided by the standard deviation. Using the results from the data matched using genetic algorithms (one-to-two) shows the estimated effect size is 0.16 for mathematics and 0.14 for science. Typically, effect sizes below 0.2 are considered "small" impacts.

The calculation of the effect size, however, is more complicated in this case. Effect sizes from results that employ regression are sometimes expressed from Cohen's f^2:

$$f^2 = \frac{R^2_{X,\Delta,\tau} - R^2_X}{1 - R^2_{X,\Delta,\tau}}$$

Where $R^2_{X,\Delta,\tau}$ denotes the pseudo R^2 for the regression model which includes socio-demographics (X), the interaction term (Δ), and PLTW participation (τ). The recommended thresholds for f^2 are slightly different than *d*. A small effect is around 0.02, a moderate effect is approximately 0.15, and large effect is 0.35.

Using McFadden's adjusted R^2 formula, we determined the approximate effect size was 0.15 for mathematics and 0.05 for science. Thus, the impact on mathematics scores could be considered a "moderate" effect while the impact on science seems relatively "small."

Multiple Carnegie Units

Tables 5 or 6 treat PLTW as a binary variable with students who enrolled in any PLTW classes labeled as a participant, regardless whether it was 1, 2, or even 3 Carnegie units. Other papers have limited their analysis to students completing or enrolling in two or more Carnegie units of instruction, arguing that PLTW must be analyzed only as a holistic sequence of courses (Bottoms & Anthony, 2005).

Most of the students in the matched data set only enrolled in one PLTW course in 9th or 10th grade. Sixty-three percent of matched participants only enrolled in a single PLTW course, 35.3% enrolled in two courses, and 1.4% enrolled in three courses—indicating they simultaneously enrolled in two PLTW courses in a single year.[10]

[FIGURE 3 ABOUT HERE]

Figure 3 shows PLTW estimates using various different controls for participation: binary, continuous/binary, and continuous. The binary results reflect the coefficient and standard error

[10] These statistics are similar for the entire cohort, including unmatched students. Of the entire 2009 and 2010 cohort, 65.4% of students enrolled in a single PLTW courses in their freshman or sophomore year, 29.1% enrolled in two courses, and 5.4% enrolled in three courses.

previously discussed. Continuous/binary uses a binary variable to denote the cumulative PLTW units and then representing the total as a set of binary variables. Point-estimates show that test growth corresponds to higher growth in test scores. The scores grew at an increasing rate for mathematics scores and a decreasing rate for science scores. However, the confidence intervals show the results are statistically insignificant at the 5% level.

Bottoms & Anthony (2005) conducted their review of PLTW by isolating PLTW students who enrolled in two or more PLTW courses. Since PLTW is a sequence of courses, the authors' motivation was to use students who had the opportunity for PLTW's approach to take effect. We also analyzed PLTW's impact for students who only enrolled in two or more PLTW courses. Again, the coefficient estimates a positive impact of 7.3 and 8.8 points for mathematics and science scores, but it is not statistically significant. A continuous variable shows each course increases test scores for mathematics and science by 3.8 and 4.9 points, respectively. Yet, these estimates were also statistically insignificant.

Project Lead The Way and Additional Mathematics & Science Courses

The PLTW curriculum also requires several pre- or co-requisite courses. For instance, Introduction to Engineering Design, typically the first PLTW courses, usually requires algebra as a co-requisite. Subsequent foundation courses recommend algebra 2 as pre-or co-requisites. Thus, entry into PLTW also means required enrollment in additional mathematics courses.

So far, we have not controlled for additional mathematics and science courses, but are the slight increase in test scores associated with PLTW attributable to the program or additional mathematics and science courses? Our data set also contains other course information that was coded with NCES or SCED identifiers that allows us to explore this question.[11] We counted the

[11] Appendix A contains the coding scheme used to aggregate math and science courses.

cumulative Carnegie units enrolled in a variety of mathematics and science courses in 9th and 10th grade for our analysis.

Table 7 contains descriptive data of enrollment in additional mathematics and science courses from the matched data set. Despite the matching performed on 8th grade data, there is still substantial difference between the number of Carnegie units in mathematics and science. In Algebra 2, for instance, the average number of units for a PLTW participant in the treatment group was 0.38, compared to 0.14 for a nonparticipant in the control group. Meanwhile, the average number of Carnegie units in chemistry for an individual in the treatment group was 0.33 units by the 10th grade compared to 0.12 in the control group.

[TABLE 7 ABOUT HERE]

Enrollment in these courses also has a large impact on mathematics and science standardized scores. Table 8 shows the estimates of PLTW while controlling for 10 mathematics courses and four science courses. Including these other courses lead to a significant reduction in the correlation between PLTW and test score growth. The interaction term for mathematics drops from +5 points (p-value of 0.01) to +0.5 points (p-value of 0.42); similarly, the coefficient for science scores falls from +5 (p-value of 0.02) points to +0.8 points (p-value of 0.38).

[TABLE 8 ABOUT HERE]

The results show seven courses with a statistically significant (p-value < 0.05) impact on mathematics and science growth: chemistry, physics, algebra 1 (negative), algebra 2, algebra 3/trigonometry, precalculus, and business/technical mathematics (negative). These results are consistent with other research, especially with respect to algebra 2 which is shown to have an enormous association with educational and workforce outcomes (Aldeman, 2006; Harvill, 2010).

The 14 additional course variables provided very little additional explanation power, instead, the mathematics and science courses seems to supplant the explanatory power of PLTW participation. The adjusted McFadden pseudo-R^2 for mathematics scores only grows from 6.1% to 6.6% when courses are included. The respective statistics for science grows from 4.7% to 5.7% while the PLTW coefficient is reduced to nearly zero.

7 Conclusions

There are several points of interest based on these results. First, PLTW in Iowa tends to attract white males and students who have higher achievement in mathematics and science in junior high. The percentages of women who enter PLTW are approximately the same percentage that enters mechanical and electrical engineering programs (Walcerz, 2006). Overall, the selection bias was rather significant, with participants scoring between 0.72 and 0.9 standard deviations higher on mathematics and science scores prior to entering PLTW. We strongly suggest that future research on PLTW needs to reflect selection bias in their analysis of the program.

We controlled for selection bias using propensity score matching that predicted entry into the program based on demographics, economic status, and whether the student was a part of a special population. We used a differences-in-differences growth model to measure the relative growth of students between 8th and 11th grade. We found that PLTW increase both mathematics and science scores by 5.2 points.

It is not immediately clear if 5.2 points is a large, modest, or small increase. A measure of effect size—Cohen's f^2—showed the impact was approximately 0.15 for mathematics and 0.05 for science—a moderate and small impact, respectively. Our results show a stronger impact than similar studies utilizing matching. Using the intercept and participation coefficients, we found

this roughly equates PLTW students performing in the 91st percentile in mathematics while non-participants scored in the 81st percentile. PLTW participants scored in the 83rd percentile compared to the 80th percentile for similar non-participants.

Our findings contradict Tran & Nathan (2010) who found PLTW had insufficient or adverse integration for science and mathematics, respectively. However, our findings support the conclusions of Tran & Nathan (2009) who showed positive impacts on mathematics with weaker impact on science. Our statistical significance was higher, which likely related to the greater statistical power from the larger sample.

We also conducted some analysis that included additional mathematics and science courses (e.g., algebra 2, chemistry). PLTW students were much more likely to enroll in higher level mathematics and science courses, which support the findings from Bottom & Uhn (2007). It is possible, but we cannot conclusively determine, that PLTW's pre- and co-requisites lead students to enrolling in more mathematics and science courses in high school. It is also possible that PLTW's program encourage students to enroll in additional coursework due to other factors, such as self-efficacy or student engagement.

Notwithstanding the reason, it is clear that these additional mathematics and science courses play an important role in the growth of test scores. Including these courses led to a substantial reduction in the estimated impact of PLTW. The pseudo R^2 suggests the additional courses simply supplant the variation explained by PLTW. These results strongly suggest that researchers need to also consider the role of other mathematics and science course, especially whether PLTW leads to increased enrollment in mathematics and science.

Assessment Instrument

The current policy environment has emphasized growth and measurable gains in core subjects within standardized testing. Nevertheless, the use of statewide standardized exams may be inappropriate for the evaluation of a PBL curriculum (Tran & Nathan, 2010; Brophy, Klein, Portsmore, & Roger, 2008; Hmelo, et al., 1997). In studies already cited, PBL students perform at par or below on assessments which emphasize recollection of factual knowledge, but perform moderately better on assessments measuring applied skill.

While the achievement measure used in this study, the ITBS and ITED, include items that range in assessment of factual knowledge to application of skills to real-life scenarios, its purpose and content is not necessarily aligned with that of PLTW. For example, the range of content covered in the science sub-test of the ITBS includes scientific inquiry, an area more aligned with PLTW, but the other three content areas are less related to PLTW: life science, earth and space science, and physical science. It will be less likely that a change is captured in achievement if the test is less closely aligned with the content of a program, such as PLTW.

Limitations

We have conducted a statewide evaluation of PLTW's impact on test scores. There are a few limitations that we would like to address in future research and evaluation. Namely, this study uses administrative data, which is distinct from transcript data as it only captures course enrollment, not completion. We may overestimate the enrollment in PLTW or other courses by not being able to remove students who dropped a course. Likewise, we also do not have access to student grades or grade point averages.

This study does not provide any controls for individual classrooms or teachers since the data was not available. There are multiple ways to view this issue and its impact. Within the context of a multi-level model, we essentially ignored another level where we observe students

within a classroom that is within a school district. The effects of PLTW within the classroom may be positively correlated and vary within the school district. The variance of effects within the classroom can be influenced by teachers and their qualifications. Earlier, we noted the varying proportions of teachers with advanced degrees in each school district. Also, we noted the challenges in which PLTW can either be taught by teachers licensed in mathematics, science, or industrial technology. We are uncertain if teacher qualifications or licensures have a differential impact on student performance.

Future Directions

Test scores are relevant to the literature given the current policy environment's emphasis on measureable gains on standardized tests. Yet, gains on standardized tests are not the only goals of the program. Future iterations of this research will explore other outcomes that should be considered, such as college attendance, choice of majors, and college completion. The authors are currently using Iowa's SLDS to follow outcomes for the treatment and control group beyond high school, which will enable research of PLTW's long-term outcomes.

Future research is also needed on PLTW's impact on problem solving through testing instruments that are oriented toward evaluating problem-solving and critical thinking. Prior research on PBL has shown little success of students on fact-based exams, but it is possible PLTW improves critical thinking and problem solving abilities.

The results of this study also suggest a strong relationship between PLTW and enrollment in other mathematics and science courses. These other mathematics and science courses play an even larger role in mathematics and science achievement scores than PLTW. Future research is needed to explore this relationship and whether PLTW increases enrollment in these areas

through the co- and pre-requisites, increased self-efficacy, increased student engagement, or other mechanisms.

References

Adelman, C. (2006). The Toolbox Revisited: Paths to Degree Completion from High School Through. *U.S. Department of Education*.

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine*, *68*(1), 52-81.

Barcikowski, R. S. (1981). Statistical Power with Group Mean as the Unit of Analysis. *Journal of Educational and Behavioral Statistics*, *6*(3), 267-285.

Barrows, H., & Kelson, A. C. (1995). *Problem-based Learning in Secondary Education and the Problem-Based Learning Institute (Monograph 1)*. Springfield, IL: Problem-Based Learning Institute.

Blais, R. R., & Adelson, G. I. (1998). Project Lead the Way(SM) Models a Program for Changing Technology Education. *Tech Directions*, *58*(4), 40-43.

Bottoms, G., & Anthony, K. (2005). *Project Lead the Way: A Pre-engineering Cirriculum That Works* (Research Brief). Atlanta, GA: Southern Regional Education Board.

Bottoms, G., & Uhn, J. (2007). *Project Lead The Way Works: A New Type of Career and Technical Program* (Research Brief). Atlanta, GA: Southern Regional Education Board.

Brandt, D. (2009, January). Feeding the need for industry: Zulma Toro-Ramos targets grade-school students for pre-engineering program. *Industrial Engineer*.

Brophy, S., Klein, S., Portsmore, M., & Rogers, C. (2008). Advancing Engineering Education in P-12 Classrooms. *Journal of Engineering Education*, *97*(3), 369-387.

Brown, A. L., & Campione, J. C. (1996). Guided Discovery in a Community of Learners. In K. McGilly (Ed.), *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice* (pp. 229-272). MIT Press.

Cech, S. J. (2007, September 26). Engineering a Blueprint for Success. *Education Week*.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge Academic.

Diamond, A., & Sekhon, J. S. (2010). Genetic Matching for Estimating Causal Effects: A General Matching Method for Achieving Balance in Observational Studies. *Working Paper*.

Dochy, F., Segers, M., Bossche, P. V. den, & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, *13*, 533-568.

Duffy, J. (2006). Agent Based Modeling and Human Subject Experiments. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of computational economics: Agent-based computational economics* (pp. 949-1012). Elsevier.

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (1st ed.). Cambridge University Press.

Goldstein, H. (2002). *Multilevel Statistical Models* (3rd ed.). Wiley.

Goodman, L. J., Brueschke, E. E., Bone, R. C., Rose, W. H., Williams, E. J., & Paul, H. A. (1991). An Experiment in Medical Education. *JAMA: The Journal of the American Medical Association*, *265*(18), 2373 -2376.

Harvill, E. L. (2010). *High School Math Curriculum, Students' Course Selection and Education Outcomes*. Working Paper, University of Pennsylvania.

Harwell, M., & LeBeau, B. (2010). Student Eligibility for a Free Lunch as an SES Measure in Education Research. *Educational Researcher*, *39*(2), 120-131.

Hmelo, C. E. (1998). Problem-Based Learning: Effects on the Early Acquisition of Cognitive Skill in Medicine. *The Journal of the Learning Sciences*, *7*(2), 173-208.

Hmelo, C. E., Gotterer, G. S., & Bransford, J. D. (1997). A Theory-Driven Approach to Assessing the Cognitive Effects of PBL. *Instructional Science*, *25*, 387-408.

Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How do Student Learn? *Educational Psychology Review*, *16*(3), 235-266.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, *42*(8).

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor: University of Michigan Press.

Holmes, D. B., & Kaufman, D. M. (1994). Tutoring in problem-based learning: a teacher development process. *Medical Education*, *28*(4), 275-283.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *48*(4), 63-85.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75.

Kreft, I. G. G., & Leeuw, J. de. (1998). *Introducing Multilevel Modeling* (1st ed.). Sage Publications Ltd.

Kuenzi, J. J. (2008). *Science, technology, engineering and mathematics (STEM) education: Background, federal policy and legislative action* (No. Congressional Research Service Paper Order Code RL33434). Washington D.C.: Federation of American Scientists.

Maguire, K. (2006). *Statewide system to foster the intregration of academics into Career and Technical Education through Project Lead The Way* (Ph.D. Capstone). Iowa State University, Ames, IA.

Mennin, S. P., Friedman, M., Skipper, B., Kalishman, S., & Snyder, J. (1993). Performances on the NBME I, II, and III by medical students in the problem-based learning and conventional tracks at the University of New Mexico. *Academic Medicine: Journal of the Association of American Medical Colleges*, *68*(8), 616-624.

Mills, J. E., & Treagust, D. F. (2003). Engineering Education - Is problem-based or project-based learning the answer? *Australasian Journal of Engineering Education*.

Misko, T. (2011). The "E" in STEM: Clarifying What Engineering Education Means for K-12 + The Opportunity Equation. Retrieved September 26, 2011, from http://opportunityequation.org/school-and-system-design/e-stem-clarifying-what-engineering#thor-misko

Nathan, Mitchell J., Tran, Natalie, Phelps, A., & Prevost, A. (2008). The Structure of High School Academic and Pre-Engineering Curricula: Mathematics. Presented at the American Society for Engineering Education 2007, Honollulu, Hawaii.

Patel, V L, Groen, G J, & Norman, G R. (1991). Effects of conventional and problem-based medical curricula on problem solving. *Academic Medicine*, *66*(7), 380-389.

Patel, Vimla L., Groen, Guy J., & Norman, Geoffrey R. (1993). Reasoning and Instruction in Medical Curricula. *Cognition and Instruction*, *10*(4), 335-378.

Prevost, A., Nathan, Mitchell J., Stein, B., Tran, Natalie, & Phelps, A. (2009). Integration of Mathematics in Pre-college Engineering: The search for explicit connections. Presented at the American Society for Engineering Education Annual Conference 2009, Austin, TX.

Project Lead The Way. (2006). Project Lead The Way: Colleges & Universities FAQ's. Retrieved November 1, 2009, from http://www.pltw.org/faqs/colleges-universities.html

Rogers, G. E. (2006). The Effectiveness of Project Lead The Way Curriculua in Developing Pre-engineering Competencies as Perceived by Indiana Teachers. *Journal of Techonology Education*, *18*(1).

Rosenbaum, P. R. (2002). *Observational studies*. Springer.

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.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, *79*(387), 516-524.

Rotherham, A. J. (2011, January 20). Shanghai Surprise: Don't Sweat Global Test Data. *Time*. Retrieved from http://www.time.com/time/nation/article/0,8599,2043312,00.html

Schenk Jr., T. (2007). *The Effects of Graduate-Student Unionization* (Masters Thesis). Iowa State University.

Schultz, P. W., Hernandez, P. R., Woodcock, A., Estrada, M., Chance, R. C., Aguilar, M., & Serpe, R. T. (2011). Patching the Pipeline. *Educational Evaluation and Policy Analysis*, *33*(1), 95 -114.

Spector, J. M. (2004). Problems with problem-based learning: Comments on model-centered learning and instruction in Seel (2003). *Technology, Instruction, Cognition and Learning*, *1*(4), 359–374.

Spellman, K. (2007, February 1). Opportunities Come at the Strangest Times. *Techniques*.

Steenbergen, M. R., & Jones, B. S. (2002). Modeling Multilevel Data Structures. *American Journal of Political Science*, *46*(1), 218.

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

Taylor, K. D., Foster, G. N., & Ratcliff, M. M. (2006). Project Lead The Way - Raising the Technical Literacy of High School Graduates. *Proceedings of ICEE 2006* (p. R3D-19 - R3D-21). Presented at the 9th International Conference on Engineering Education, San Juan, PR.

Tran, N., & Nathan, M. J. (2010a). An investigation of the relationship between pre-college engineering studies and student achievement in science and mathematics. *Journal of Engineering Education*, *99*(2), 1–15.

Tran, N., & Nathan, M. J. (2010b). The effects of pre-engineering studies on mathematics and science achievements for high school students. *International Journal of Engineering Education (Special Issue on Applications of Engineering Education Research)*, *26*(5), 1049-1060.

Vernon, D. T., & Blake, R. L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine: Journal of the Association of American Medical Colleges*, *68*(7), 550-563.

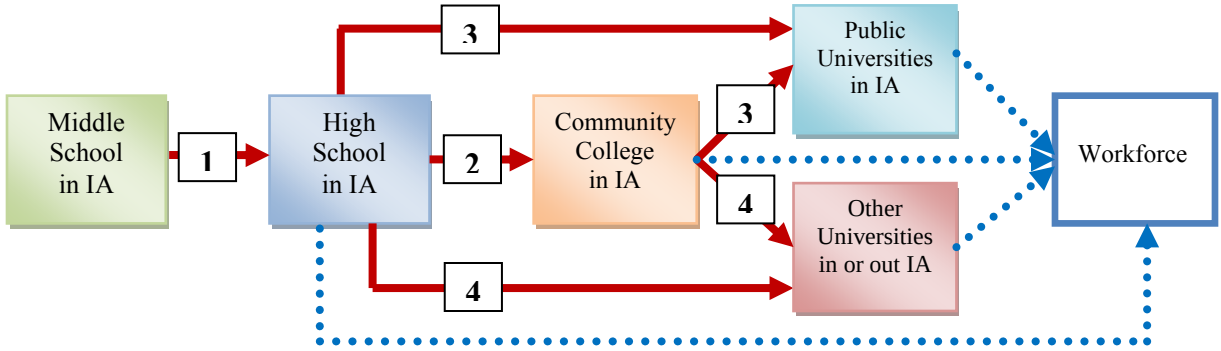
Walcerz, D. (2007). *Report on the Third Year of Implementation of the TrueOutcomes Assessment System for Project Lead The Way*. TrueOutcomes.

Wildavsky, B. (2011, April). Think Again: Education - By Ben Wildavsky | Foreign Policy. *Foreign Policy*. Retrieved from http://www.foreignpolicy.com/articles/2011/02/22/think_again_education

Williams, A., & Williams, P. J. (1994). An approach to teaching technology. In M. Ostwald & A. Kingsland (Eds.), *Research and development in Problem Based Learning*, Research and

development in problem based learning ; v. 2, 1994 (pp. 355-367). Sydney: Australian Problem Based Learning Network.

Figure 1: Iowa Statewide Longitudinal Dataset: Tracking Project Lead The Way Students



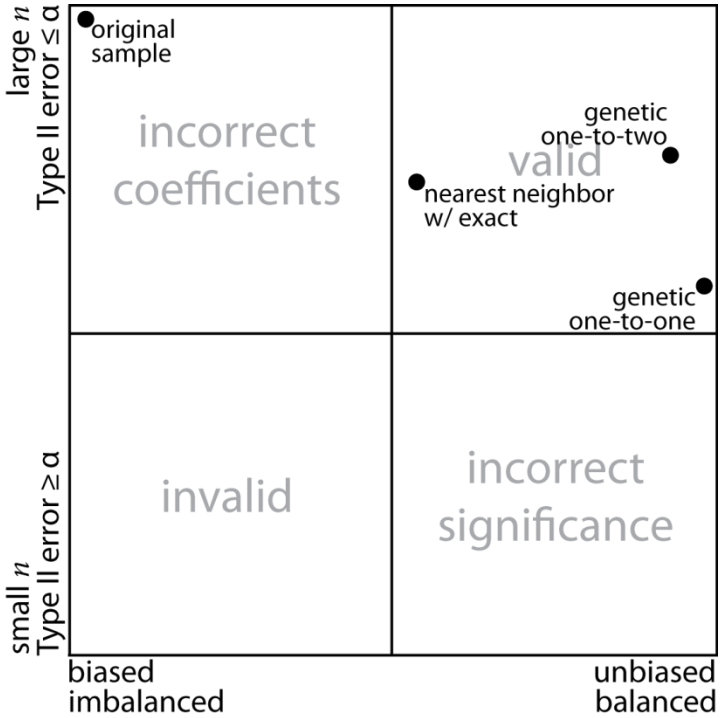
Note: The transition shown in [1] is tracked by Iowa's secondary data system (Project EASIER); the transition into community colleges [2] is from Iowa's community college data system (MIS), we obtain the transition into public universities, [3] through partnership with the State Board of Regents; and the transition into other higher education institutions [4] is from the National Student Clearinghouse.

Table 1: Descriptive statistics of Project Lead The Way data set: 2009 & 2010 Cohorts

	Nonparticipants (N = 24,709)		PLTW Participants (N = 1,321)	
	Mean	Standard Deviation	Mean	Standard Deviation
Demographics				
Male	0.49	0.50	0.85	0.36
American Indian	0.01	0.09	0.00	0.06
Asian	0.03	0.16	0.03	0.17
Black	0.09	0.29	0.03	0.16
Hispanic	0.07	0.26	0.03	0.17
White	0.80	0.40	0.91	0.29
Economic Status Proxy				
Free Lunch	0.24	0.43	0.11	0.31
Reduced Lunch	0.07	0.25	0.05	0.21
Homeless	0.01	0.09	0.01	0.08
Special Populations				
Section 504	0.01	0.10	0.01	0.09
Gifted & Talented	0.10	0.29	0.21	0.41
IEP	0.15	0.36	0.02	0.13
Testing - 11th Grade ITED				
Reading - Standard Score	285.75	43.84	307.29	35.44
Reading - Percentile Rank	59.38	26.77	73.19	20.63
Mathematics - Standard Score	286.43	41.55	319.42	28.89
Mathematics - Percentile Rank	60.89	28.63	83.27	18.27
Science - National Standard Score	294.70	44.21	323.90	34.95
Science - Percentile Rank	64.55	27.04	81.74	18.82
Testing - 8th Grade ITBS				
Reading - Standard Score	254.56	41.13	274.68	29.33
Reading - Percentile Rank	57.63	27.57	72.23	20.42
Mathematics - Standard Score	257.45	37.47	282.95	25.41
Mathematics - Percentile Rank	59.57	27.70	79.52	18.18
Science - National Standard Score	263.88	36.22	285.17	27.92
Science - Percentile Rank	62.19	24.95	77.21	17.22

Note: Data represents pooled 2009 and 2010 cohorts. Testing data reflects national standard score and percentile ranks. Control group includes students from districts with PLTW, but who did not enroll in any of those courses.

Figure 2: Hypothetical Trade-off between Matching Algorithms



Note: The diagram is for illustrative purposes. The scale and precise location may differ by data set.

Table 2: Propensity Score Coefficients

	Coefficient	z-statistic
Intercept	-1654.0	-19.57
ITBS, 8th Grade		
Reading	0.0	-1.86
Mathematics	0.0	8.69
Science	0.0	4.90
Demographics		
Black	-0.6	-1.45
Asian	0.0	-0.11
Hispanic	-0.3	-0.71
Male	1.8	21.72
Economic Status Proxy		
Free Lunch	-0.2	-2.47
Reduced Lunch	0.0	-0.14
Homeless	-0.4	-0.69
Special Populations		
IEP	-1.0	-5.77
Section 504	-0.5	-1.42
Gifted & Talented	0.3	4.28

Note: Estimates are for enrollment into PLTW courses in either 9th or 10th grade. Data includes controls for students in either the 2009 or 2010 cohorts.

Table 3: Standardized differences between PLTW and non-PLTW students in 8th grade by matching algorithm

		Matching Methodology		
	Prior to Matching	Nearest Neighbor with Exact Matching on Districts	Genetic Algorithm One-to-One	Genetic Algorithm One-to-Two
Propensity Score Distance	0.953	0.036	-0.001	0.002
Cohort	-0.067	-0.075	0	0.001
ITBS, 8th Grade				
Reading	0.597	0.106	-0.002	-0.005
Mathematics	0.891	0.103	0.001	0.004
Science	0.719	0.148	0.002	0.007
Demographics				
White	0.353	-0.044	0	0
African American	-0.386	0.034	0	0
Asian/Pacific Islander	0.027	-0.024	0	0
Hispanic	-0.224	0.05	0	0
American Indian	-0.097	0.052	0	0
Male	0.972	-0.051	0	0.002
Economic Status Proxy				
Free Lunch	-0.405	-0.035	0	0
Reduced Lunch	-0.064	0.011	0	0.003
Homeless	-0.068	0.026	0	0
Special Populations				
IEP	-0.563	-0.014	0	0.004
Section 504 Plan	-0.086	0.052	0	0
Gifted & Talented	0.271	0.038	0.003	0
School District	-0.049	0	NA	NA
Sample Size (n_i)	15,660	1,500	1,477	2,111
Treated Observations ($n_{i\tau}$)	751	750	751	751
Control Observations ($n_{i\tau'}$)	14,909	750	726	1,360

Note: Figures above show $(\mu_\tau - \mu_{\tau'})/\sigma_{pooled}$. Positive sign denotes participants were above the population averages.

Table 4: Descriptive statistics of matched data set: Genetic matching one-to-two

Variable	Weighted Mean	Weighted Standard Deviation
PLTW Participant	0.38	0.49
Demographics		
White	0.90	0.30
Black	0.03	0.17
Asian	0.04	0.18
Hispanic	0.03	0.17
American Indian	0.00	0.06
Male	0.85	0.36
Economic Status Proxy		
Free Lunch	0.11	0.31
Reduced Lunch	0.04	0.21
Homeless Status	0.00	0.06
Special Populations		
IEP	0.02	0.14
Section 504	0.01	0.09
Gifted Talented	0.23	0.42
11th Grade scores		
Reading	306.28	37.31
Mathematics	315.06	31.14
Science	320.28	36.43
8th Grade Scores		
Reading	277.48	30.72
Mathematics	283.72	26.66
Science	288.92	29.88

Note: Descriptive data for the treatment and control group after matching using propensity score matching using genetic algorithms with one treatment matched up to two nontreatment cases. The weighted means and standard deviations are based on the weights assigned from the matching algorithm.

Table 5: Iowa Project Lead The Way's impact (fixed effects) on mathematics scores, 8th to 11th grade

	Genetic One-to-Two		Genetic One-to-One		Nearest Neighbor	
	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
Intercept (α)	272.5	150.87	272.74	139.6	125.4	9.3
PLTW Participant (τ)	2.8	1.51	1.97	1.1	-14.1	-1.2
ITED (Junior-year test) (T)	30.8	21.40	31.24	20.5	10.6	12.4
PLTW Participant x ITED (Δ)	5.2	2.40	3.95	1.7	1.1	1.4
Additional Testing Controls:						
Midyear	4.0	2.21	3.41	1.8	0.9	0.3
Spring	5.8	2.28	8.59	3.1	1.6	0.4
Demographics:						
African-American	-14.1	-5.58	-14.67	-4.9	-16.9	-4.2
Asian/Pacific Islander	0.0	-0.01	4.10	1.4	2.6	0.9
Hispanic	-14.9	-5.94	-14.28	-4.8	-12.9	-3.7
American Indian	15.1	1.94	13.18	1.1	5.8	0.9
Male	3.4	3.18	3.26	2.6	4.1	2.6
Economic Status Proxy						
Free Lunch Eligible	-8.0	-6.01	-8.68	-5.5	-2.0	-1.1
Reduced Lunch Eligible	-4.1	-2.38	-6.50	-3.2	-5.4	-2.4
Homeless	3.4	0.40	9.22	0.9	-3.3	-0.4
Special Populations						
IEP	-31.6	-13.73	-28.26	-10.0	-29.9	-8.5
Section 504 Plan	-6.6	-1.28	-0.38	-0.1	3.6	0.5
Gifted and Talented	26.2	26.08	26.17	22.0	26.7	21.5
Number of Observations	4075		2854		2867	
School Districts	76		71		45	
AIC	37739		26530		26581	
BIC	37941		26721		26772	
LogLikelihood	-18837		-13233		-13259	
McFadden's R-squared	6.1%		5.7%		6.1%	

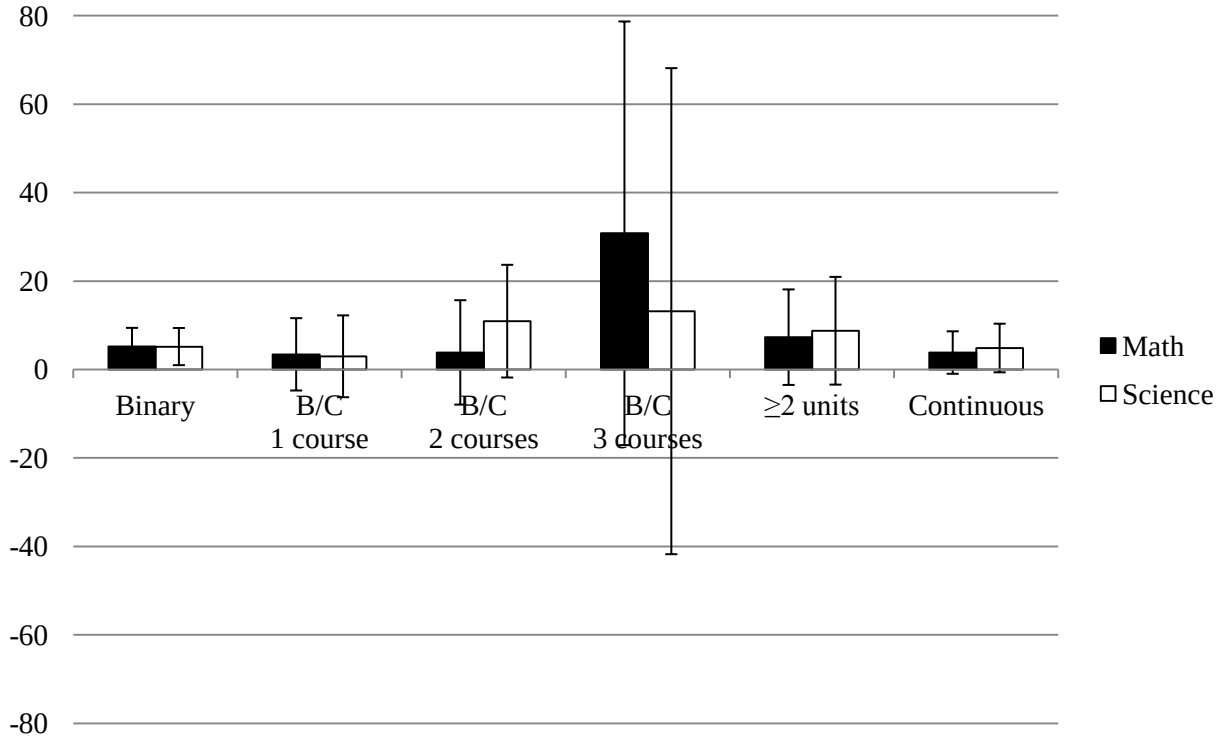
Note: Estimates are for the national standardized test score, mathematics, between 8th grade Iowa Test of Basic Skills (ITBS) and 11th grade Iowa Test of Educational Development (ITED). Estimates are shown as national scale scores. Estimates are for the class of 2009 and 2010 cohorts.

Table 6: Iowa Project Lead The Way's impact (fixed effects) on science scores, 8th to 11th grade

	Genetic One-to-Two		Genetic One-to-One		Nearest Neighbor	
	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
Intercept (α)	279.8	139.19	122.80	11.1	279.2	103.5
PLTW Participant (τ)	-0.8	-0.36	-22.05	-1.5	-0.4	-0.2
ITED (Junior-year test) (T)	32.1	20.67	11.01	15.1	32.7	17.4
PLTW Participant x ITED (Δ)	5.2	2.02	1.60	1.6	4.5	1.5
Additional Testing Controls:						
Midyear	2.3	1.16	-17.93	-4.7	-12.1	-3.1
Spring	3.5	1.24	2.54	0.7	3.0	0.9
Demographics:						
African-American	-13.4	-4.61	3.60	1.5	2.1	0.8
Asian/Pacific Islander	-0.4	-0.14	5.40	1.7	2.2	0.6
Hispanic	-17.1	-5.93	-11.90	-2.9	-10.7	-2.8
American Indian	-10.8	-0.83	-17.98	-1.6	-18.1	-1.1
Male	0.2	0.15	0.66	0.4	-1.3	-0.8
Economic Status Proxy						
Free Lunch Eligible	-11.7	-7.59	-7.73	-3.7	-10.1	-5.4
Reduced Lunch Eligible	-7.6	-3.74	-8.68	-3.2	-9.2	-3.7
Homeless	-15.6	-1.64	-2.62	-0.2	-8.4	-0.7
Special Populations						
IEP	-31.7	-12.00	-27.16	-6.8	-32.7	-10.4
Section 504 Plan	-8.6	-1.51	-5.95	-0.8	-3.5	-0.5
Gifted and Talented	30.2	25.56	28.66	18.7	30.1	21.2
Number of Observations	4071		2851		2862	
School Districts	76		71		45	
AIC	29283		27432		27568	
BIC	39485		27623		27759	
LogLikelihood	-19610		-13684		-13752	
McFadden's R-squared	4.7%		4.8%		4.8%	

Note: Estimates are for the national standardized test score, mathematics, between 8th grade Iowa Test of Basic Skills (ITBS) and 11th grade Iowa Test of Educational Development (ITED). Estimates are shown as national scale scores. Estimates are for the class of 2009 and 2010 cohorts.

Figure 3: Summary of estimated impact (fixed effects) of Project Lead The Way on mathematics and science scores with binary, continuous/binary, and continuous measures



Note: Estimated impacts shown for separate regressions using binary controls (PLTW participant or nonparticipants), a continuous/binary control (number of PLTW Carnegie units, each as a binary variable), and continuous control (number of PLTW Carnegie units). Above estimates are derived from a one-to-two match using genetic algorithms. Estimates are shown as national scale scores. Covariates include additional testing controls, demographics, economic status proxy, special populations, and clustering by school district. Error bars represent the 95% confidence interval.

Table 6: Average number of Carnegie units in mathematics and science in 9th and 10th grade by treatment and control group

Course	Treatment Group	Control Group	Difference in Means	t-statistic
Biology	0.87	0.61	0.26	56.80
	(0.43)	(0.56)		
Chemistry	0.33	0.12	0.21	57.58
	(0.47)	(0.32)		
Physics	0.12	0.01	0.11	49.02
	(0.33)	(0.11)		
Algebra 1	0.58	0.39	0.18	34.93
	(0.58)	(0.57)		
Geometry	0.80	0.33	0.46	110.84
	(0.42)	(0.49)		
Algebra 2	0.38	0.14	0.24	60.90
	(0.49)	(0.35)		
Algebra 3 or Trigonometry	0.09	0.01	0.08	40.85
	(0.29)	(0.11)		
Precalculus	0.03	0.01	0.02	13.73
	(0.18)	(0.12)		
Calculus	0.01	0.00	0.01	6.72
	(0.12)	(0.04)		
Probability & Statistics	0.01	0.00	0.00	4.73
	(0.07)	(0.05)		
International Baccalaureate Mathematics	0.00	0.00	0.00	-3.51
	0.00	(0.02)		
Business/Technical Mathematics	0.00	0.06	-0.06	-30.28
	(0.04)	(0.29)		
Other Mathematics	0.05	0.01	0.05	24.99
	(0.27)	(0.09)		

Note: Data shown for matched data set using genetic algorithms using a maximum of two control cases matched to each treatment case.

Table 8: Estimated impact (fixed effects) of Project Lead The Way on mathematics and science scores
with course covariates

	Mathematics		Science	
	Estimate	t-statistic	Estimate	t-statistic
PLTW Participant x ITED (Δ)	0.5	0.2	0.8	0.3
Cumulative courses (Carnegie units)				
Biology	0.1	0.1	2.3	1.4
Chemistry	8.5	6.0	10.1	6.0
Physics	5.6	2.3	5.8	2.0
Algebra 1	-6.5	-5.4	-4.2	-3.0
Algebra 2	12.2	8.3	15.6	8.9
Algebra 3 or Trigonometry	10.2	3.7	9.2	2.8
Precalculus	12.6	3.9	9.6	2.5
Calculus	-1.6	-0.3	-4.3	-0.6
Probability & Statistics	5.8	0.9	6.1	0.8
International Baccalaureate Mathematics	-12.2	-1.2	-15.6	-1.3
Business/Technical Mathematics	-19.2	-6.7	-19.7	-5.8
Other Mathematics	-6.4	-1.7	-3.1	-0.7
McFadden's R-squared	6.6%		5.7%	

Note: Above estimates are fixed effects derived from a one-to-two match using genetic algorithms. Estimates are shown as national scale scores. Covariates include additional testing controls, demographics, economic status proxy, special populations, and clustering by school district. Omitted course variables are physical science and geometry.