

Evaluating the Impact of the Petey
Greene Prisoner Assistance Program

August 2013

Evaluating the Impact of the Petey Greene Prisoner Assistance Program

August 2013

Daniel J. Kowalski

Program Analyst
Office of Educational Services
New Jersey Department of Corrections

Provided by the 2012 Charles W. Puttkammer '58 Prisoner Reentry Fellowship

*Completed in Partnership with:
The New Jersey Department of Corrections, and
The Pace Center for Civic Engagement at Princeton University*

Acknowledgements

This research study is heavily indebted to the guidance, mentorship, and supervision of Jecrois Jean-Baptiste, Director of Education for the New Jersey Department of Corrections. Director Jean-Baptiste oversaw tutoring program implementation and evaluation design within the Department in addition to advising me through the entirety of the research process. Volunteer Programs Coordinator Michael Ritter was critical in facilitating site visits, communicating between colleges and prisons, maintaining volunteer information across the State. In addition, this study could not be complete without the tireless dedication to education from the great supervisors and teachers at the NJ Department of Corrections, especially including Mark Yaros, Cynthia Troiani, Nicole Cruz, Cathy Morgan, Eugene Maggioncalda, Deniece Gray, Celeste Thatcher, and Michael Bianchini.

I also deeply appreciate the feedback from a host of associates from Princeton University, including Professors Nolan McCarty, Kosuke Imai, and Amy Lerman as well as Puttkammer Fellow Jarrah O'Neill and SINSI Scholar Muhsin Usman Hassan. Additionally, a number of college students helped tremendously in administering and collecting surveys and other research tasks, especially including David Hammer, Grace Li, and Ray Chao. Ultimate thanks to Elsie Sheidler and the Princeton University Pace Center, which toiled to coordinate the fellowship that produced this study.

Finally and most importantly, this project was sponsored by the generous Charles W. Puttkammer '58 Prisoner Reentry Fellowship, established by Charlie Puttkammer to provide recent graduates of Princeton University experience in prison education and reentry work and establish impact in the field of corrections one individual at a time. Jim Farrin, my long time mentor, role model, and friend has support me in this project every step of the way, and for that I am full of gratitude and respect.

Table of Contents

I.	Executive Summary	7
II.	Introduction	9
III.	Review of Literature	12
IV.	Program History	16
V.	Methods of Analysis	17
VI.	Facility Tutoring Models	23
VII.	Functional Level Gains: Math and Reading	27
VIII.	GED Achievement	36
IX.	Impact Beyond the Classroom	39
X.	Evaluation Conclusions	43
XI.	Recommendations	44
XII.	Appendix	47
	a. Sources Cited	47
	b. Individualized Inmate Learning Plan (IILP) Folder	51
	c. Test Score Matching and Regression Results	52
	d. Exit Survey Template	56
	e. Exit Survey Results	57

Table of Figures

<i>i. Petey Green Logic Model</i>	<i>11</i>
<i>ii. Metrics Dashboard</i>	<i>16</i>
<i>iii. Tutors by College</i>	<i>16</i>
<i>iv. Summary of Participating Prisons</i>	<i>23</i>
<i>v. Comparison of Prison Tutoring Models</i>	<i>24</i>
<i>vi. PRISON 1 Math-Tutoring Functional Level Gains</i>	<i>29</i>
<i>vii. PRISON 1 Reading-Tutoring Functional Level Gains</i>	<i>29</i>
<i>viii. PRISON 2 Math-Tutoring Functional Level Gains</i>	<i>30</i>
<i>ix. PRISON 3 Math-Tutoring Functional Level Gains</i>	<i>32</i>
<i>x. Learning Gains Findings and Methodologies Summary</i>	<i>35</i>
<i>xi. GED Achievement Comparison</i>	<i>36</i>
<i>xii. GED Passing Rates Comparison</i>	<i>38</i>
<i>xiii. Tutored Inmates Release Patterns</i>	<i>42</i>

Abbreviations and Acronyms

<i>(p<)</i>	<i>Statistical Significance Level</i>
<i>ABE</i>	<i>Adult Basic Education</i>
<i>ANCOVA</i>	<i>Analysis of Covariance Test</i>
<i>DOC</i>	<i>Department of Corrections (New Jersey)</i>
<i>ESL</i>	<i>English as a Second Language</i>
<i>GED</i>	<i>General Education Development test/certificate</i>
<i>HSD</i>	<i>High School Diploma</i>
<i>IILP</i>	<i>Individualized Inmate Learning Plan (AKA “Tutoring Folder”)</i>
<i>TABE</i>	<i>Test of Adult Basic Education</i>

Executive Summary

The Petey Greene Program, established in Princeton, NJ in 2008, provides volunteer tutoring services to incarcerated individuals with low skills—inmates who function between the 1-12th grade levels and lack a high school diploma or GED.

Correctional education is consistently one of the most effective interventions in reducing recidivism and increasing employment. Academic tutoring is an effective tool for improving educational levels. However, academic tutoring within correctional facilities is little understood and awaits rigorous quantitative evaluation. This study employs a semester-long analysis of the Petey Greene Program's impact on inmate academic gains in math and reading in five different facilities across the state of New Jersey. Our evaluation presents the program's impact on basic functional level gains as well as a qualitative analysis of tutoring models. It also includes important GED outcomes for tutored inmates and attempts to expand the scope of impact for future evaluators. Finally, the study presents a set of recommendations for program expansion, operations, and evaluation.

Functional Level Gains

In this study, we developed a rigorous matching design to measure test score improvements between inmates participating in tutoring and similar inmates not participating in tutoring in both basic math and basic reading across five participating facilities. Notably, we found the following results:

- ❖ Math tutoring at both Prison 1 and Prison 2 accelerated test score progress by more than one (~1.6 and ~1.1) full grade level during the term, with statistical significance.
- ❖ Reading tutoring at Prison 1 accelerated test score progress by more than one (2.3) full grade level during the term, with statistical significance.

Tutoring Model Observations

In addition, we analyzed qualitatively the tutoring models at each participating facility, noting major differences in tutor group size and style, primary subject tutored, and other tutoring model observations. We found the following:

- ❖ In Spring 2013, we identified nearly 200 volunteers and at least 211 inmates as participating in Petey Greene Program tutoring, primarily in basic Math and Reading.
- ❖ Facility tutoring models were highly varied, in terms of the size of tutoring sessions, staff supervision, classroom environment, and timing of tutoring.

GED Achievement

Tracking participating inmates throughout the semester, we observed a major outcome in correctional education: GED achievement. This aspect of our results provides an outcomes benchmark for the Petey Greene Program. In the Spring of 2013, we found:

- ❖ 17% of all tutored inmates achieved the GED within the period of study.
- ❖ Tutored inmates account for 25% of GED achievement for participating facilities, contributing 36 out of 144 total GEDs.
- ❖ Tutored inmates passed the GED with a 90% passing rate as compared to an average facility passing rate of 83%.

Beyond the Classroom

We gathered information from a number of different sources, including surveys, semi-structured interviews, and site visits to develop a wider perspective of program impact, including impact on inmates, volunteers, and prison staff. Our analysis indicates:

- ❖ The Petey Greene program helps to decrease facility violence levels, motivate students, communicate individual inmate needs to education staff, and hold students and education staff accountable.
- ❖ Volunteers gain pre-professional experience, particularly in the fields of criminal justice (~20% of volunteers), education (~10%), and public policy (~25%).
- ❖ 58% of tutored inmates will be released within two years of participating in tutoring. Over one-third (36%) will leave within the same year of program participation.

Recommendations

Finally, we propose a set of recommendations based on our Spring 2013 evaluation of the Petey Greene Program. An outline of our recommendations:

- ❖ Clarify program goals and align services and metrics to those goals.
- ❖ Establish more effective tutoring models and materials.
- ❖ Develop an improved tutor training or professional development program.
- ❖ Proceed with continuous monitoring and further evaluation of the Petey Greene program upon program maturity.

Introduction

The United States of America is one of the most literate nations in the world: Less than 1% of the total population is functionally illiterate (CIA World Factbook, 2013). Yet, nearly 60% of the prison population is functionally illiterate, and more than 80% of inmates do not hold a high school diploma or General Educational Development (GED) certificate (Wolford, 1989; Kozol, 1985). Nearly 95% of inmates will leave federal and state prisons each year, which adds to about 700,000 men and women (Guerino, Harrison, and Sabol 2012). Within three years of release, depending on the State of interest, 40%-60% of ex-offenders commit new crimes or violate the terms of their release and are incarcerated once again (The Pew Center on the States 2011). Such a high recidivism rates exhaust more than \$50 billion per year in state budgets (National Association of State Budget officers 2011).

The New Jersey Department of Corrections has ascertained that inmates enrolled in educational programs average between a 5th and 7th grade functional levels in math and reading, as determined by the federally-acknowledged Test of Adult Basic Education (TABE). Notwithstanding, the population serviced by adult basic education (ABE) in New Jersey correctional facilities has increased by 87% from 2009 to 2013 while the number serviced by GED-level instruction has increased by 56%, all during a recessionary period of fewer teaching staff and stagnant program funds. Furthermore, inmates with very low baseline literacy levels experience more difficulty increasing those levels in adult basic programming than do inmates who begin at higher literacy levels (NJDOC, 2013).

Enrollment in correctional academic programs is correlated with higher TABE scores in reading, math, and language arts (Messemer and Valentine, 2004). In-prison enrollment in education is also correlated with higher employment rates and lower recidivism rates (Cecil et al, 2000; Fabelo, 2002; Aos et al., 2006; Gaes, 2008; Lockwood et al., 2012; Piotrowski and Lathrop, 2012). In fact, meta-analysis of the literature demonstrates a 13% reduction in recidivism rate, on average (Davis et al., 2013). Hence, increases in TABE scores are also correlated with higher employment and lower recidivism. However, prisons and jails across the US have limited resources to help increase inmate reading and math TABE scores.

Volunteers have responded to the need to supplement existing correctional education programs en masse, notably through a project in New Jersey called the Petey Greene Prisoner Assistance Program, which provides one-on-one and small group tutoring to inmates using volunteer college students and college community members. As of yet, there is no literature on the impact of participation in external volunteer tutoring on inmate learning gains within the prison environment. This research is thus critical as such volunteer programs consume time and organizational energy.

The purpose of our study is to evaluate the impact of the Petey Greene Prisoner Assistance Program for four major reasons:

1. To describe to all relevant stakeholders precisely how the program functions within the participating prisons.
2. To provide a grounded framework for future research and evaluation of the program and clarify program standards, goals, and processes.
3. To analyze quantitatively inmate academic progress attributable to the program.
4. To aid stakeholders in program expansion using a clear set of information tools.

The first and second guiding purposes of this research are critical because the phenomenon of volunteer tutoring for adults within the prison context is not well studied. Any evaluation of such a program would thus require a deeper understanding of the mechanisms, processes, and extraneous factors by which the program functions on the facility level. The third and fourth purposes of this study are related: Current and future stakeholders attempting to expand the program need measurable outcomes and benchmarks, in line with the adage, "One cannot manage what one cannot measure." Because the program is in its very incipient stages of data collection and analysis, this study also proposes and utilizes a set of materials for data collection and programmatic purposes.

Program Logic Model

The Petey Greene Prisoner Assistance Program organizes volunteer college students and community members to tutor at participating correctional facilities across New Jersey. It provides transportation, orientation, and community for volunteers and pairs volunteer tutors with academically struggling inmates generally in one-on-one and small group, two-hour sessions. The program logic model in figure i illustrates the inputs, outputs, and intended outcomes of the Petey Greene Program. This evaluation is primarily focused on outcomes, although it also describes output in the form of tutoring models. The Petey Greene Program specifically works towards two main objectives:

1. Improving inmate learning gains in basic reading, math, and language arts (grade equivalencies of 0-12) through one-on-one and small group tutoring and aiding inmates in attaining their high school equivalency certificate.
2. Increasing awareness of prison life and education through community immersion in the prison environment.

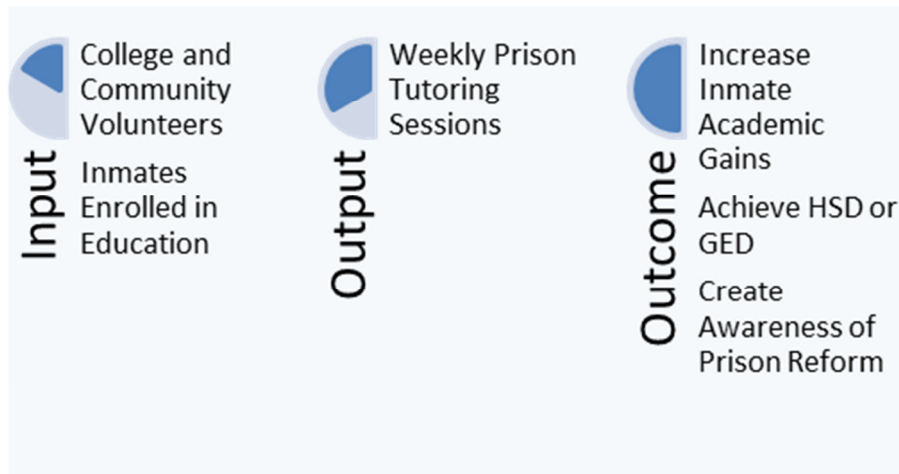


Figure i: Program Logic Model

This study tackles the first goal of the Petey Greene Program in improving inmate learning gains in basic reading, math, and language arts through a semester-long pretest-posttest design in five different facilities, utilizing a propensity score matching procedure to create statistically viable control groups and multivariate regressions involving pretest covariates. Additionally, we compare tutored inmate achievement on the GED to average facility achievement on the GED and analyze the differences in tutoring models at each facility. The findings of this evaluation should inform (1) the current impact of the program, (2) semester progress benchmarks, and (3) recommendations for program expansion and further program evaluation.

Structure of Study

The evaluation proceeds from a review of the relevant literature to a description of the methods employed for data collection and analysis. The evaluation then gives a short quantitative history of the Petey Greene Program to arrive at the facilities under study for the Spring 2013 term. The next section describes the demographics of the participating facilities as well as their unique tutoring models, which will frame our interpretations of the quantitative evaluation of inmate functional level gains. Section seven provides a justification of the methods employed for each case as well as the observed correlation of participation in tutoring and TABE test score increases. We include a discussion of GED achievement as well as some qualitative effects of the program to provide a broader sense of the program's impact. We conclude the evaluation with a discussion of the main findings of this study and recommendations for program operations, expansion, and evaluation.

Review of Literature

The following section provides a review of the quantitative and qualitative literature regarding correctional education, tutoring programs, and impact evaluations within the field of public safety.

Correctional Education as a Means

Education within the prison system has taken on the challenge of the undereducated caste of incarcerated Americans, meeting with great success. Nationally, correctional facilities provide adult literacy education for up to 25% of the incarcerated adults who have not graduated from high school (Harlow, 2003).

These adult education programs significantly improve inmate functional levels. Using TABE scores as indicators of educational achievement, Messemer and Valentine (2004) argued that an adult basic education program in a closed security prison significantly improved academic achievement in reading and math. Their study found that the number of classroom hours to gain 1.0 grade level in reading, math, and language were 118.4 hours, 54.2 hours, and 36.2 hours, respectively.

Such learning gains within the prison environment are correlated with lower recidivism rates and higher employment rates. A recent evaluation of correctional education programs in Indiana provides evidence that inmates receiving their high school diploma or its equivalency and inmates enrolled in postsecondary education while incarceration greatly reduce their chances of recidivating (Lockwood et al., 2012). In Texas, researchers found that educational programs significantly improved inmate reading levels and employment rates and significantly decreased recidivism rates, particularly for non-readers and the illiterate. Average wages, too, increased once an ex-offender reached a 9.0 reading level and especially increased upon achieving a high school equivalency diploma or GED (Fabelo, 2002). A recent meta-analysis of correctional education program evaluations by the RAND Corporation show that inmates participating in education had a 13% reduction in recidivism rate, 13% higher odds of obtaining employment post release and 28% higher odds for vocational programs (Davis et al., 2013). A wealth of additional work corroborates the powerful impact of correctional education on recidivism and employment (Piotrowski and Lathrop, 2012; Aos et al., 2006; Gaes, 2008).

While correctional education provides some of the most effective programs in terms of recidivism, employment, and further educational attainment while incarcerated, the available resources are limited. In New Jersey, for instance, fewer teachers and more inmates enrolled in education reduces individualized instruction and seat time in class. Hence, additional aid in correctional education programming is crucial.

Volunteer Academic Tutors in the Correctional Environment

Volunteer initiatives have responded to this need, supplementing existing academic programs within the prison system. While volunteerism is not new to corrections in America, volunteer tutor programs have largely grown out of the “reintegration” philosophy of corrections, engaging the community as a vehicle of change while bolstering individual agency.

The “Reintegration Model” of corrections, which is unique in its complimentary focus on communal and individual factors of change, largely began to gain momentum with the 1967 report commissioned by President Lyndon B. Johnson entitled “The Challenge of Crime in a Free Society.” Mapping out a set of priorities and recommendations for criminal justice reform, the report systematically examined the component systems of police, courts, parole boards, and correctional institutions, the technological tools and strategies for reform, and the factors that facilitate criminal behavior. While its main focus was the criminal justice system, the commission was not shy to introduce potential points of partnership between educational institutions and community-based organizations. Such partnerships are meant to ease offender transition into society by providing the public an understanding of prison life while assisting inmates with their particular rehabilitative needs.

The field of correctional education, like other correctional programs, often suffers from a scarcity of funds, which contributes to larger waiting lists and insufficient support (U.S. Department of Education, 2009; Tolbert, 2012). Hence, well-organized and effective volunteer work is crucial in such settings. Volunteer programs often serve a dual purpose: providing training and resources toward reintegrating inmates into their communities and providing non-incarcerated citizens the experience of a prison environment. Such experiences are vital to the reintegration model of corrections, which requires members of the communities themselves to understand inmate needs and experiences. The reintegration theory of correctional education hypothesizes that community and volunteer-based initiatives, which encourage both offender agency and community responsibility in the education process, should be successful in reducing recidivism, increasing post-incarceration employment, and achieving other major targets (Lutze et al., 2011; Guy, 2009; Bazemore et al., 2004).

As part of a larger reintegration tradition, volunteer programs have emerged to fill a void by providing community members, college students, and professionals the venue for supporting inmates in their academic pursuits. This support has taken a variety of forms, from mentoring to tutoring to classroom teaching, within a number of different states across the US and a diversity of service providers (Spencer, 2012; Sonnen and Kempf, 2008; North Carolina Department of Public Safety, 2012). Academic tutoring programs thus represent a unique case of the volunteer-based aspect of the “reintegration” movement. However, there is not much literature to support adult tutoring programs despite the abundant literature on best practices and volunteer training manuals (Sandlin and St. Clair, 2005). Yet, in other contexts, tutoring is largely beneficial.

Studies of literacy tutoring provided to at-risk youth in the school setting demonstrate that tutoring programs can have a significant effect on reading levels both in the short-term and long after the program terminates (Fitzgerald, 2001; Burns et al., 2008). Young students assessed as having learning disabilities likewise benefit from personalized,

one-to-one tutoring over a short period of time (Osborn et al., 2007). Additional studies suggest that noncertified reading tutors, for instance volunteers, can be effective with struggling readers, in large part due to the amount and quality of guidance from the supervising reading teacher (Morris, 2006; Wasik, 1997). Much of the research on tutoring programs arose out of the *America Reads Challenge*, a national commitment expressed by President Bill Clinton to the goal that every child will read independently and well by the end of the third grade (Wasik, 1997). Consequently, most of the existing literature on tutoring programs is literacy-based, as opposed to analyzing gains in math or other subjects. Some research, however, suggests that volunteer tutoring by college students also has a positive effect on child math gains (Baker et al., 2006). Furthermore, because of the context of these works, little attention has been given to adult learners, particularly in correctional facilities (Ziegler, McCallum, and Bell, 2009).

Volunteer tutoring programs also have their skeptics. These critics focus on volunteers as under-skilled and ill-suited to work with critically at risk learners (Perin, 1999). Some argue that at-risk learners should have more time with qualified teachers rather than being left with uncertified tutors. Others debate the effectiveness of tutor training, in particular (Sandlin and St. Clair, 2005). However, Ziegler, et al (2009) also argued that volunteer instructors have about the same level of knowledge as part-time and full-time teachers, in terms of content. Waite (1983) claims that the average individual is capable of assisting someone to read. Other studies report a strong relationship between training and instructional knowledge (Bell et al., 2004). The importance of instructional knowledge, i.e. knowing *how* to teach as opposed to *what* to teach, is thus at the crux of volunteer tutor professional development (Belzer, 2007). Belzer, (2006a, 2006b) also argues for improved and continual tutor training throughout the program. In terms of adult tutoring, which overwhelmingly involves reading-disabled adult literacy learners, tutors generally utilize what they learn in orientation quite effectively, despite feeling ill prepared to tutor (Sandman-Hurley, 2008).

Yet volunteer tutors are here to stay. Some researchers estimate that about 60% of adult literacy instructors are volunteers (Ziegler et al., 2007). Because of the immense interest in adult literacy tutoring, the lack of research on volunteer tutoring within the adult corrections environment, and the great political debate around the subject, a closer look at such programs is imperative.

Program Evaluations in Correctional Education

In the field of corrections, programs often live or die by the impact evaluation. At the time of this study, we have not found any rigorous analysis of volunteer tutoring programs within a correctional setting. The dearth of rigorous quantitative work in the field of correctional program evaluation likewise raises the stakes of such a study. Hence, our analysis of the Petey Greene Prisoner Assistance Program, a college-based volunteer tutoring program that partners college students with inmates for weekly, one-on-one intensive tutoring, seeks to fill the empirical void.

While past evaluations of correctional programs have certainly enriched the field, there is room for improvement. Few existing evaluations of education programs incorporate rigorous methods or randomized treatment and are thus open to criticisms of

bias and confounding factors (James, 2011). Additionally, the dominant outcome measures of such studies are often misleading. Wade (2007) writes:

“Recidivism rates cannot measure whether an educational program helped inmates obtain new knowledge. Future evaluations need to consider learning gains the inmates may achieve, because it is unethical to base the success of a program on recidivism rates when a lack of marketable skills might be leading inmates to commit crimes.”

Indeed, the Messemer and Valentine (2004) study is somewhat unique in its approach to evaluating a correctional education program, measuring TABE scale score improvement as opposed to recidivism or other long-term indicators. Such research in the area of correctional education on academic achievement is crucial as it requires less time and resources while providing meaningful information about a program’s efficacy. For this reason, we take inmate functional level gains as determined by TABE scores to define our primary quantitative outcome of study.

Program History

The Petey Greene Prisoner Assistance Program began its pilot in the spring of 2008 at Albert C. Wagner Youth Correctional Facility in Bordentown, NJ. In five years, the program has expanded to include seven of New Jersey’s twelve state prisons and five colleges, including Princeton, Rutgers, Seton Hall, The College of New Jersey, and Rowan.

Academic Year	2008-09	2009-10	2010-11	2011-12	2012-13
Fiscal Year	FY2009	FY2010	FY2011	FY2012	FY2013
Number of Colleges	1	1	1	2	5
Number of Prisons	1	1	2	3	7
Number of States	1	1	1	1	1
Number of Volunteers	38	34	52	130	198
Number of Volunteer Visits	-	-	-	1596	2466

Figure ii: Program Metrics Dashboard

The program has also grown to accompany nearly 200 volunteers, made of college undergraduates, graduate students, and community members who provide tutoring for roughly two hours on a weekly basis at one of the participating prisons. As presented in Figure iii, roughly one-third of volunteers were Princeton students and community members while another third hailed from TCNJ. An additional 20% of volunteers were Rutgers students, and the remaining was split between Seton Hall and Rowan.

Of the seven participating facilities, we chose the five facilities that have operated the tutoring program for at least one semester. Each facility operated a distinct tutoring model and each produced slightly different outcomes. Replicating and improving upon these outcomes and tutoring models is of central concern as the Petey Greene Program seeks expansion within New Jersey as well as to other States.

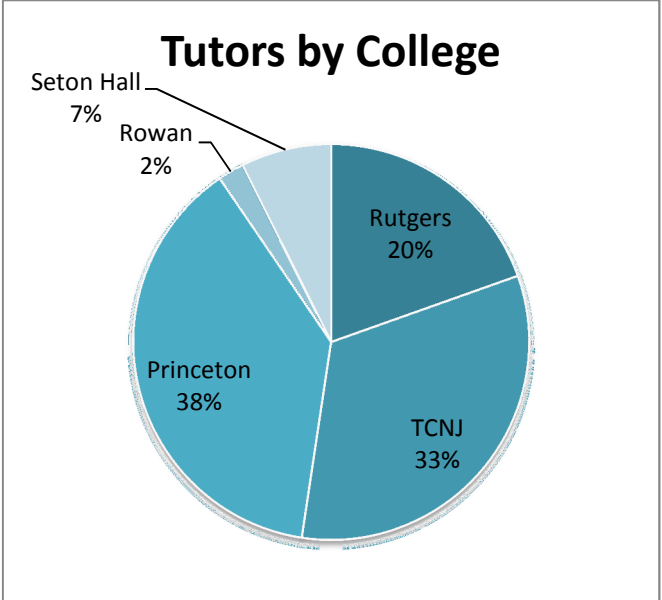


Figure iii: Tutors by College

Methods of Analysis

This study takes a primarily quantitative mixed-methods approach, combining case study techniques with a set of statistical procedures for each prison. While the quantitative aspect of the study is our focus, our results are contingent on the widely varying particulars of each program. Hence, we believe it appropriate to interpret our quantitative study through a thick context of *what* is happening in the tutoring session as well as *how* tutors themselves engage and react in the tutoring process from researcher site visits and tutor exit surveys. While these qualitative details may complicate our understanding of *why* different outcomes occur, they will nonetheless enrich our interpretation of those outcomes for added external validity. The qualitative aspects of this study primarily examine differences in facility tutoring models and unintended effects of the Petey Greene Program, as described by tutors and education staff. The quantitative aspects of this study deal with inmate academic learning gains, as determined by scale scores and grade equivalency scores on the Test of Adult Basic Education (TABE), as well as inmate GED achievement, in the form of count data and passing percentages. While the latter is not *rigorous* in that it does not utilize a control group with random assignment, it is nonetheless beneficial for establishing program benchmark metrics. The analysis of inmate academic learning gains via TABE scores, however, comprises the rigorous quantitative aspect of our present study, as described in this section.

Research Questions

The quantitative and qualitative methods employed attempt to answer the following four questions:

1. How do different prisons and tutoring models compare to each other? And might those comparisons enhance or inhibit program effectiveness? (*Qualitative*)
2. Does the volunteer tutoring program succeed in achieving its desired impact to increase inmate basic education levels in reading and math? (*Quantitative*)
3. Does tutoring contribute to inmate GED achievement? (*Quantitative*)
4. Are there additional, unintended aspects of Petey Greene's impact? (*Qualitative*)

For each question, we used a slightly different methodology, as described below. However, we often shared the same sources of data between methodologies, including the Department case load database, inmate tutoring folders, site visits, unstructured interviews with volunteers and education staff, and tutor exit surveys.

Question 1: Identifying Tutoring Models

We studied five prisons with regular tutoring program operations: Prison 1, Prison 2, Prison 3, Prison 4, and Prison 5. For each prison, we conducted a qualitative case study with each facility program as our unit of observation to understand how tutoring models compare to each other and how those differences might alter service delivery. The case studies included site visits, session observations, and volunteer tutor surveys to better understand the mechanisms that might drive inmate academic achievement within the tutoring session and to provide a thick context for the program evaluation. This analysis includes an overview of the structure of tutoring at the facility and descriptive statistics regarding the particular prison. The structure of the tutoring session may be of prime importance for future researchers, as each prison organized the tutoring in their facility differently. As we will see, this may have led to the difference in inmate academic achievement. We will also discuss how the different tutoring models affected our data collection.

Site visits primarily consisted of an analysis of tutoring models, which included (a) tutor supervision, (b) the size of the tutoring group, (c) the environment for the tutoring session, (d) inmate engagement in tutoring, and (e) the use of academic materials throughout the tutoring session. Each facility had relatively different standards and practices for organizing the tutoring despite the supply of tutors coming from the same organizational source, i.e. the Petey Greene Prisoner Assistance Program. The facilities designed their own tutoring models as per the facility restrictions due to architecture, custody, time, or other program constraints. Facilities with similar constraints may be able to generalize the findings from these case studies to their own needs as well as adopt or amend program designs as needed.

Question 2: Effect of Tutoring Participation on Functional Level Gains

Using the individual inmate as our unit of analysis, we must understand the impact of participation in the program on learning progress. We define functional level progress in terms of test score progress on the Test of Adult Basic Education (TABE). In order to do so, we must utilize a pretest-posttest design and consider participation in the tutoring program our treatment variable. For a strong research design, this treatment should be randomly assigned to inmates. However, as is the case with much research in both education and corrections, there are ethics issues with such random assignment. Instead, we will take a different approach to establish near-causal inference.

For each prison, we quantitatively measure inmate academic gains without disrupting the “open-door” policy of the tutoring program or significantly altering the existing tutoring models in each of the five facilities under study, some of which have functioned for five years at the time of this study. The “open-door” policy is common in a variety of correctional programs, allowing inmates to enter and leave programs based on self-selection. The self-selection policy takes control of the assignment to treatment (i.e. tutoring) out of the hands of the researchers in a non-random way, hence the need for an observational study. While some education researchers argue that observational studies can increase the external validity of a program evaluation, others focus on the limitations in

drawing accurate causal inference (internal validity) from such studies (Dimitrov and Rumrill, 2003).

However, in order to best control for the non-random, non-controlled aspect of the study, we have employed a one-to-one propensity score matching design for each facility, allowing us to artificially create control and treatment groups “as-if” randomly. The propensity score is derived from participation in tutoring regressed against predictors of motivation and tutoring program selection in order to satisfy the “ignorability assumption,” required for drawing causal inference in matching designs (Gelman and Hill, 2007). These predictors include: custody level, age, race, prior offenses, baseline education level, number of years until release, time in prison to-date, and time between TABE pretest and posttest. All inmates in this study, whether tutored or not, were enrolled in educational programming during the time period studied. Such a design will match inmates who received tutoring with those who have not based on the propensity score, which produces *on average* a sample of as-if randomly assigned inmates.

After matching, we apply a multivariate regression analysis to each facility and each subject tutored, using the TABE posttest scale scores as our dependent variable, the number of tutoring sessions as our independent variable, the TABE pretest scale score as the covariate, and a set of demographic information as control variables. The coefficients from the multivariate regression will provide us with a sense of how relatively effective tutoring is for academic achievement while a simple analysis of covariance (ANCOVA) test will provide significance levels of the tutoring program treatment. We simulate a number of models for each facility to find the best fit, which we will describe in each facility’s data analysis section. Of note, we have decided to separate our analysis into five different facilities as opposed to aggregating our data into one hierarchical model because the assignment mechanisms and facility-level factors are themselves so varied, including the types of tutoring models and unique prison cultures and program interactions. Hence, we rely on the case study approach to describe such facility level factors instead of including them in our statistical models.

As a policy in the New Jersey Department of Corrections, every inmate must be administered a TABE test at least every six months. Hence, Department policy has ensured a pretest and posttest for almost every inmate. There are two exceptions to this rule. First, for inmates who are in isolation or “Administrative Segregation,” standardized testing is a security issue. However, these inmates did not receive tutoring during the time period studied. Second, inmates who are released for home or to a half-way house within that six month period may not be tested again. However, this is becoming rarer as teachers and supervisors of education have a strong incentive to conduct a posttest on inmates who are soon to leave, as per funding guidelines for the Department of Corrections.

The data for our analysis is derived from two sources: (1) the IILP tracking form and (2) the internal Department case load database. The IILP provided for each inmate contains a tracking form, which documents the date, subject area, and assignments for each tutoring session as well as the inmate’s name and identification number. The identification number connects this information with the information in the Department of Correction’s internal database, called ITAG. All testing information, demographic data and custodial history are contained within this internal database. The IILPs date back to the beginning of the tutoring period and end at the conclusion of the period: January 1, 2013 to May 1, 2013. Hence, inmates were provided a 3-4 month window for tutoring, depending on the facility.

The ITAG database dates back to 2000 and thus contains all relevant and recent information on every inmate in New Jersey adult custody. We use testing data from the beginning and end of the IILP time period for our analysis.

IILP. The Individualized Inmate Learning Plan (IILP) is a folder specific to each inmate, containing a tracking form, a comments sheet, and worksheets for that session. The folder mirrors a similar instrument developed by the Corporation for National Community Service (CNCS) called the “Student Monthly Contact Log” (Sample Instrument ED2, 2011). Both instruments contain identifying information like name and number, a place for tutor/instructor comments, the date and contact hours for the lesson, the subject of the lesson, and the assignment(s) from the lesson. The tutor documents this information, and the folder is kept with the teacher so as not to become a potential security issue. The IILP is also important for the program, as it allows a continuum of services despite tutor shifts. At the end of the period under study, we collect all IILPs and combine the information with demographic and testing data from the primary Department database, TAG.

TAG Database. The TAG database is the New Jersey Department of Correction’s internal warehouse of caseload information. It contains pretest-posttest data, inmate demographic and historical information, and sentencing data for every inmate incarcerated in a state facility. This database is the basis of most of the information for our propensity score models and learning gains models.

Methodology Ratings

The standard for rating research methodologies within the fields of education and corrections have been set by the What Works Clearinghouse (WWC) and the Maryland Scientific Methods Scale (MSMS), respectively. In its 2013 meta-analysis of correctional education programs, the RAND Corporation used both scales to rate the methodologies of the studies under review (Davis, et al., 2013). The meta-analysis utilized the WWC’s scale of “Meets standards,” “Meets standards with reservations,” and “Does not meet standards” and the MSMS scale of one through five, with five as the most rigorous study of crime prevention programs (Farrington, et al., 2002; What Works Clearinghouse, 2011).

This meta-analysis considered the “Meets standards” and MSMS of 5 to describe a study with randomized control and treatment groups with low participant attrition. “Meets standards with reservations” and MSMS of 4 is reserved for studies that use a quasi-experimental design (or high-attrition randomized design) in which the treatment and control groups are matched at baseline on at least age, prior offenses, and baseline education levels.

Finally, the “Does not meet standards” for education research aligns generally with all 1, 2, and 3 on the MSMS ratings scale. A 3 on the MSMS signifies a study that controls for at least some baseline differences between groups, often through matching. A score of 2 signifies no random assignment and no statistical controls for baseline differences, while a score of 1 indicates the study had no separate comparison group.

We are not concerned with attrition for this study as none of the facility-level research designs involved randomization, and facilities involved in the study were very evidently at either end of the high-low attrition extreme, primarily because of facility-level data collection practices. In other words, a facility either had the requisite data for rigorous

study or it did not; there was not much attrition due to inmates being released or failing to take a posttest.

We explain the facility-level methodology ratings in the quantitatively-focused “Functional Level Gains” section of this study in order to assess our research findings alongside the rigor of their respective research methodologies. This strategy is imperative for evaluating a program with such a diversity of operational designs.

Question 3: Tutoring on GED Achievement

To what extent does participation in tutoring contribute to GED achievement, one of the major outcomes of correctional education programming? Our answer to this research question will provide descriptive statistics of GED achievement for inmates receiving tutoring compared to the general population in education for the Spring of 2013. The individual inmate is our unit of analysis for this section. Using inmate tutoring folders, we compile a list of inmates identified as participating in one-on-one or small group tutoring and compare this list to the list of inmates who either passed or failed the GED during the same term. We then record the breakdown of GED count data (i.e. GED achievement) and passing rates for each facility and for both the tutored and non-tutored sub-groups. We will not be able to infer direct causal impact, but we will gain a sense of program contribution to GED achievement, which is particularly important for facilities that focused tutoring resources on higher skilled (GED-level) inmates. These figures can also act as a set of benchmarks for certificate achievement of tutored inmates.

Question 4: Developing a Grounded Theory for Program Impact

Fourthly, we attempt a grounded theory of program impact based primarily on qualitative data, using all individual stakeholders as our unit of analysis. Individual stakeholders include teachers, inmates, tutors, prison staff, and program organizers. Grounded theory research occurs when the researcher leaves off *a priori* judgment of program expectations and instead develops a theory of the phenomenon as he or she is experiencing that phenomenon. For this aspect of the study, the only *a priori* data we sought out was sentencing figures. Based on those inmates identified as participating in tutoring, we wanted to know how long they would stay in corrections custody before release. These figures are critical as they can lead to future research regarding inmate recidivism, employment, and continuing education, indicators that are in high demand within the field of correctional program evaluation. Unfortunately, given the short time window of this study, such an evaluation is not feasible.

In addition, we utilized a combination of volunteer and prison staff unstructured interviews regarding the program and tutor exit surveys to better understand how various participants in the program articulate the impact of the Petey Greene Program. We did not survey or interview inmates extensively or formally as this was not feasible given time constraints and Department ethics constraints.

Tutor Exit Survey. Each tutor at the four participating colleges was provided an exit survey, either at the facility of tutoring, through the involved student organization, or

through the respective campus civic center. The survey template is attached to the appendix of this study. The survey was anonymous and asked personal demographic questions about the tutor, binary questions about comfortability in and organization of the program, and open-ended questions regarding positive and negative feedback. The survey was also helpful to program organizers as they have taken into consideration the feedback in their plans for expansion.

Limitations of Study

While our methods employed have strong external validity, or generalizability, some lack in internal validity, or independent causality. Outcomes based on an analysis of tutoring models, for instance, can be widely generalizable, yet we cannot be sure exactly what particular aspects of the tutoring models or prison demographics caused the discrepancies in outcomes across facilities. Our GED achievement analysis also lacks internal validity, as we cannot determine that tutoring alone caused GED achievement or high passing rates.

Our analysis of test score correlations, however, has both strong internal and external validity, particularly for facilities with larger inmate populations and more tutoring sessions per inmate. These factors allow our cases to have large sample sizes with which we can match a control group and approximate effect sizes, after controlling for confounding variables. The drawback of this approach is that it underestimates facility-level effects and spill-over effects (indirect aspects of program impact) and is also dependent on accurate data collection and transmission.

Finally, a thorough grounded theory of volunteer tutoring within correctional systems would have to be more robust than our attempt, which has the humble purpose of opening future evaluators to a broader conception of impact and communicating those facility-level and spill-over effects not captured in our statistical analyses.

Facility Tutoring Models

One of the most important aspects of our study is an explanation of what actually happens in the prison and during the tutoring sessions. An understanding of outcomes is incomplete without a thorough knowledge of means and mechanisms. This section describes the different tutoring models utilized in each facility under study: Prison 1, Prison 2, Prison 3, Prison 4, and Prison 5 Minimum Security Unit.

The differences in tutoring models and thus the particular mechanisms of impact could significantly alter the ways in which educational outcomes are achieved. Our analysis of tutoring models, based on facility site visits and unstructured interviews with supervisors of education and teachers, demonstrates a close approximation to the regular functioning of the tutoring program at a facility level.

This section first provides a snapshot of relevant facility demographics, then a comparative overview of tutoring models, and finally an analysis of how the different tutoring models slightly change our analysis and interpretation of outcomes.

Overview of Participating Facilities

Of the five facilities under study during the spring 2013 semester, one (Prison 1) was an adult mid-to-high security facility, another (Prison 5) was an adult low security unit, and the remaining three (Prison 4, Prison 3, and Prison 2) were all youth facilities containing varying security levels.¹ All tutoring sessions under study took place within the medium security education wings of their respective facilities. Notably, all facilities under study were male state prisons. Figure iv presents a breakdown of relevant demographic data, including age, average sentence, and average math and reading levels, per participating facility.

Each of the three youth facilities had an average age of 23-24 while the two adult facilities maintained populations with an average age in the 30s. Sentence lengths were much more varied, however. Prison 5, the minimum unit, had the shortest average sentence, with the average inmate serving only 1.4 years before his parole eligibility date. Prison 2 also hosted a relatively short-termed population with only

Facility	Avg. Age	Avg. Sentence Length	Avg. Reading Level	Avg. Math Level
PRISON 1	35.3	3.9 years	6.8	6.5
PRISON 2	23.2	1.8 years	7.4	5.7
PRISON 3	24.1	2.9 years	7.4	5.8
PRISON 4	23.5	2.6 years	6.6	5.9
PRISON 5	39.1	1.4 years	5.3	5.4

Figure iv: Summary of Participating Prisons. Average levels in math and reading correspond to TABE grade equivalencies.

¹ In New Jersey, a youth facility is not the same as a juvenile facility. Youth facilities house inmates adjudicated as adults who are generally younger than the overall incarcerated population and usually between the ages of 18-26.

1.8 years average sentence. Prison 1 had the longest average sentence of nearly four years, on average.

The reading and math levels (as determined by the TABE) of inmates in educational programming were, on average, between a 5th-7th grade level, with some variation between facilities. In addition, math levels were generally lower than reading levels. While inmates at Prison 1 have a higher average math score than the other facilities, Prison 3 and Prison 2 have higher reading levels. Prison 5 had the lowest of both math and reading levels with a 5th grade equivalent in both.

Overview of Facility Tutoring Models

The discrepancies in average test scores within facilities closely mirror the number of sessions tutored in those subjects. As Prison 3 and Prison 2 have higher average reading scores than math scores, these two facilities focused almost entirely on math tutoring. Similarly, as Prison 1, Prison 5, and Prison 4 tend to experience roughly equivalent average math and reading scores, these facilities have split tutoring fairly evenly between math and reading, as illustrated in Figure v. Of note, Prison 4’s reading tutoring was primarily geared toward students learning English as a Second Language, in contrast with Prison 5 and Prison 1, which provided reading tutoring for native English speakers.

Figure v: Prison Tutoring Models

Facility	College	n of Volunteers	n of Inmates	Avg. Sessions / Inmate	Space and Time	Session Sizes	Supervision	Primary Subject Tutored
PRISON 1	Seton Hall	14	12	10.1	Pull-Out: Supplementary	1-on-1 Inmate TAs	Law Librarian-Direct	Math (55%)
PRISON 2	Rutgers	34	61	2.0	Push-In Pull-Out: Substitutionary/Supplementary	1-on-1 Small Group	Teacher-Direct, Indirect	Math (82%)
PRISON 3	Princeton	49	91	2.8	Push-In	1-on-1 Small Group Classroom Rotation	Teacher-Direct	Math (88%)
PRISON 4	Princeton	23	15	3.6	Pull-Out: Substitutionary	1-on-1 Small Group	Teacher-Indirect	Math (51%)
PRISON 5	TCNJ	62	33	7.8	Push-In	1-on-1 Small Group	Teacher-Direct	Reading (54%)

Furthermore, Prison 5 and Prison 1 used the same group of inmates for tutoring over the entirety of the semester, averaging 7.8 and 10.1 tutoring sessions per participating inmate, respectively. The youth facilities, on the other hand, tended to spread services to a larger number of inmates for fewer sessions per inmate, with Prison 4 receiving slightly more resources per inmate. Each tutoring session was roughly 1.5-2 hours long. This was consistent across facilities and generally only varied based on particular daily constraints like movements, lock-down, or transportation problems.

The models for tutoring sessions were very different across facilities. The two “push-in” facilities, Prison 3 and Prison 5, hosted students within the classroom during class time due to space and security supervision constraints. Both facilities utilized direct teacher supervision of tutors and inmates, with teachers often helping tutors with instruction or directing the rest of the class during the tutoring session. Utilizing push-in models, both Prison 3 and Prison 5 were thus bound to substitute regular class time for tutoring time. Prison 5 used only individualized and small group tutoring with inmates preselected by the teacher whereas Prison 3’s tutors additionally taught classes and rotated around the room to help inmates as needed. Additionally, while all other facilities maintained tutoring folders for inmates, Prison 3 experimented with tutoring folders designated for tutors rather than inmates. Hence the materials for Prison 3’s model were different from the others in that they traced tutor activity rather than inmate activity. We later reflect upon these materials in light of the findings from our study.

Prison 1 and Prison 4, on the other hand, primarily operated individualized and small group pull-out sessions in which the teacher was not directly supervising. We identified pull-out sessions as those in which both tutor and inmate would be outside of the classroom, for instance in the hallway—like Prison 4—or in the law library—like Prison 1. Because teachers were not directly present, these facilities used other staff (e.g. the law librarian) or indirect teacher supervision, in which the teacher would occasionally check-in during the length of the tutoring session. Notably, Prison 4’s tutoring sessions occurred during class time while Prison 1’s tutoring sessions were entirely supplementary afternoon sessions. The supplementary sessions took place in a quiet environment with a limited number of inmates, and each tutored inmate was also provided support from an inmate teacher’s assistant, who knows the teachers and material, in addition to a volunteer tutor from the participating college. As the only prison that exclusively operated a supplementary tutoring program, Prison 1 will be important for us to estimate the *added value* of participation in tutoring as opposed to the value of substituting individualized tutoring for normal classroom instruction.

Finally, Prison 2’s tutoring model was a hybrid of the other four models, as this facility operated two different tutoring models simultaneously. During the day, the model was very similar to that of Prison 3, Prison 4, and Prison 5. Some inmates were pulled out of class and tutored in auxiliary rooms with indirect teacher supervision while others were tutored in the classroom, push-in style, with direct teacher supervision. During the day, tutoring was a substitute for regular classroom instruction. However, in the evening, the program was much more similar to Prison 1’s program, utilizing pull-out supplemental sessions with indirect supervision from teachers or education supervision staff. Inmate tutoring sessions were limited to one-on-one and small groups, and when the facility experienced influxes of tutors, it utilized them in other realms of the prison, for instance in administrative work, during the tutoring times.

What do the tutoring models mean for our study?

The nuances of the different tutoring models across facilities enhance our perspective of how the tutoring program functions and thus of how we are to interpret and generalize our findings. For instance, we may not be able to generalize our findings from a reading-

based, push-in session to a math-based supplementary session. Hence, we must treat each facility and each model as its own case. Aggregating the data from these cases will introduce confounding variables for which we cannot fully control. A thicker, case-based analysis, then, allows us to consider our outcomes analysis in relation to the precise operations of the programs.

First, whereas pull-out supplementary sessions help us to determine the added value of tutoring on an inmate's academic growth, an analysis of push-in and pull-out substitutionary sessions instead compares individualized and small group tutoring to participation in regular classroom work.

Furthermore, push-in sessions involve an entirely different learning environment from pull-out sessions. The former maintains engagement in the classroom community while the latter is more intimate and personalized. As previously discussed, commentators have largely advocated for supervision of tutoring programs by certified teachers and ongoing professional development support for tutors; the difference in supervisory models is similarly critical. Also, for those facilities that held primarily math-based tutoring sessions, we must only analyze tutoring program outcomes in terms of math gains. The small number of tutoring sessions in subjects other than math, for these facilities, will not meet the threshold of power tests of significance.

Finally, the session size is also deeply meaningful for how we interpret our data analysis, particularly for facilities with tutoring models more complex than the generic one-on-one and small group design. Prison 1 organized inmate teachers' assistants who also helped in tutoring and preparing tutoring materials. These TA's were crucial for communication between tutors and teachers as they rarely crossed paths in this model. The TA's may have also added to inmate learning gains in ways that other models could not, and we will not be able to precisely measure the particular effect of TA's. At Prison 3, the complexity of the tutoring model implies that our analysis will deeply underestimate the impact of the Petey Greene Program as our analysis excludes considerations of rotational tutoring and classroom instruction provided by tutors.

Our qualitative analysis of the disparate tutoring models at Prison 1, Prison 2, Prison 3, Prison 4, and Prison 5 help us to accurately interpret and generalize our findings of functional level correlations and GED achievement patterns.

Functional Level Gains: Math and Reading

Functional levels have become one of the primary metrics for education reform initiatives. Federal adult educational grants are generally awarded based on the number of adults serviced and the performance of those adults vis-à-vis achievement of functional level standards. The Department of Corrections regularly assesses these functional levels through the Test of Adult Basic Education (TABE). We made use of the regular testing to observe differences in inmate learning gains due to tutoring through a matching process and pretest-posttest multivariate regression model (explained in the Methods section).

Of the five facilities under study, three maintained testing data viable for the learning gains aspect of our analysis. One facility demonstrated a statistically significant effect of tutoring on both math and reading levels while another facility showed a significant effect in math tutoring, its primary subject tutored. Finally, the third facility with current and recorded scores did not experience any significant difference between tutored inmates' scores and similar, non-tutored inmates' scores. We speculate as to why we did not find a significant difference in test scores in more depth in this section.

We believe that some combination of program organization, facility demographics, and tutoring models account for the discrepancy in outcome between facilities, however more study is needed to parse out the precise reasons. Nonetheless, our positive results provide strong evidence that the Petey Greene Program has a significant impact on the academic learning gains of inmates enrolled in correctional education programming.

This section presents each facility as its own quantitative, self-enclosed case due to major differences in selection procedures, tutoring models, and facility demographics. Each of these facility-level factors could have a major effect on program efficacy, which we would not be able to measure with only five cases. However, taken as independent cases, we have more ability to generalize these findings to prisons and tutoring models of similar design and operation. Starting with Prison 1 in Newark, NJ, we work through two youth facilities—Prison 2 and Prison 3—and describe data limitations at Prison 4 and Prison 5. Finally, we compare findings at the end of this section.

Prison 1: Math and Reading

Our analysis provides support for a strong positive correlation between participation in the tutoring program and increased inmate math and reading levels for inmates at Prison 1.

At Prison 1, supervisory staff selected twelve inmates to participate in weekly tutoring for the entirety of the semester based on three factors: (1) academic need – inmates with lower TABE levels were selected over those with higher TABE levels; (2) level of engagement in education – inmates selected for tutoring had to display a strong work ethic; and (3) behavior – selected inmates were assumed to work more congenially with outside volunteers.

For our statistical models, we matched inmates in the tutoring program with inmates not in the tutoring program based on pretest math and reading scores and expected time in prison before release, and we controlled for covariate pretest scores and age. Matching on pretest scores addressed the selection processes based on academic need while our time until release variable addressed behavior-based selection: inmates with less expected time until release have a strong incentive to behave properly in order to earn or maintain an upcoming parole date. Level of engagement in education was the only selection variable not directly controlled for in our matching procedures. However, we controlled for age in our analyses of posttest scores in lieu of direct matching on engagement, since older inmates experienced greater learning gains than relatively younger inmates. Furthermore, the qualitative feedback of Prison 1’s supervisor of education is that older inmates are more engaged in education, on average, than younger inmates. Hence, controlling age may be a strong substitute for controlling for levels of engagement. Such matching procedures and controls allow us to infer “as-if” randomness in assignment to tutoring and thus mitigate much of the assignment biases, improving the generalizability (external validity) of our findings. Finally, no other variables (including prior offenses or recidivism) were significant predictors of inmate involvement in tutoring or posttest scores.

In math, we found a statistically significant, positive correlation between engagement in the tutoring program and math scale score increases. Our predictive model of posttest scores, after matching, accounts for 72% of the variation in posttest scores with a high degree of statistical significance. Furthermore, our model predicts an added value of roughly thirty (30) scale score points, or about one full grade level equivalent, for inmates receiving tutoring compared to what they would have scored without tutoring.

The average tutored inmate received roughly 55% or 5-6 sessions of math tutoring throughout the semester and scored roughly a 6th grade level on the math pretest. Figure vi, below, demonstrates the average math level increase in scale scores and rough grade equivalencies for our tutored group and control group with quintile scores for a sense of range.

From a sixth grade math level on average, tutored inmates reached an eighth grade math level on average within the semester. Non-tutored inmates went from a sixth grade to nearly a seventh grade equivalent, on average. One will notice the similarities in 25-75% quintile scores for tutored and non-tutored inmates. This may indicate that inmates within the middle range of scores (around sixth grade) may have benefited more from tutoring than those at the lower or higher ends of the spectrum. Our analysis of covariance test also validates our findings. In sum, individualized, one-on-one weekly supplemental tutoring with the help of volunteers and on-site inmate teacher’s assistants propelled inmates from reaching one semester’s worth of progress (or .5 grade levels) to just under four semester’s worth of progress (or nearly 2.0 grade levels) in basic math, with a high level of statistical significance (~93% confidence).

	Tutored	Non-Tutored
Avg. Math	514	519

	Tutored	Non-Tutored
Avg. Math	6.2	6.4

Pretest		
Avg. Math Posttest	550	525

Pretest		
Avg. Math Posttest	8.1	6.7

$n = 22$

Adjusted $r^2 = 0.724$

Treatment p-value : 0.07

Figure vi: Prison 1 Math Gains. The left-hand table represents scale scores while the right-hand table approximates grade level equivalencies. Note: Average posttest scores are regression estimates based on average baseline information of each treatment and control group.

In reading, we similarly found a strong, positive correlation between inmate participation in tutoring and reading level progress. The average inmate received 40%, or about 4-5 sessions, of supplemental tutoring in reading and language arts, and scored roughly a sixth grade reading level on the pretest. Our predictive model for reading posttest scores is not as strong as for math, accounting for roughly half (56%) of the variation in reading posttest scores. However, participation in the tutoring program observed a very high coefficient with statistical significance ($p < .07$) for reading scores, with tutoring adding roughly 50 scale score points, or about 1.5 grade levels, to inmate learning gains. Analysis of covariance tests similarly validated this correlation.

	Tutored	Non-Tutored
Avg. Reading Pretest	522	515
Avg. Reading Posttest	556	510

	Tutored	Non-Tutored
Avg. Reading Pretest	6.3	6.1
Avg. Reading Posttest	8.2	5.7

$n = 20$

Adjusted $r^2 = 0.560$

Treatment p-value : 0.07

Figure vii: Prison 1 Reading Gains. Note: Average posttest scores are regression estimates based on average baseline information of each treatment and control group.

In summary, the case of Prison 1's participation in the Petey Greene tutoring program demonstrated a high added value of one-on-one, sustained tutoring to inmate academic learning gains in both math and reading. Indeed, such increases in basic math and reading may not be entirely independent of each other, as many educational supervisory staff believe basic math and reading gains to go hand-in-hand, as improved reading skills reinforce mid-to-high level math in the form of word problems, one of the major curricular challenges facing correctional education programming in New Jersey.

Finally, based on our findings, after matching and controlling for the primary assignment and confounding variables, we conclude that Prison 1's supplementary tutoring model contributes to math and reading gains 3-4 times more efficiently than regular

academic programming alone. The limitation of this case study, however, was sample size. With only 12 inmates receiving tutoring, and thus 12 inmates comprising our one-to-one matched control group, statistical generalizability should be approached cautiously.

Prison 2: Math

Prison 2 also demonstrated significant improvement in tutored inmates’ basic math levels. Because 82% of tutoring sessions comprised solely of basic math tutoring, we conducted an evaluation on math levels alone for Prison 2.

In this facility, teachers chose inmates for tutoring as part of their lesson plans, before class, based primarily on inmate need. Inmates who were at the lowest levels and inmates at the highest levels relative to others in their class were selected for individualized, one-on-one tutoring during class time and occasionally during supplemental night classes.

Our matching procedure included math pretest scores, age, and time in prison. Prison 2’s data collection protocol uses TABE grade equivalencies rather than scale scores almost exclusively. Hence, we relied on grade equivalencies in this case as opposed to raw scale scores. Our matching model described that inmates were more likely to be selected for tutoring if they had low math scores, were older, and/or were more recently admitted to prison. Criminal history was not a significant factor. Our final regression over math posttest scores included the pretest covariate in addition to the number of sessions of math tutoring in which an inmate had participated. Because Prison 2 had fewer sessions of tutoring per student and more subject-focused approach to the tutoring program, the number of sessions is a better predictor for program impact than simple binary participation (as used in the case of Prison 1).

After matching and controlling for relevant variables, we found a high, positive, statistically significant correlation of the number of tutoring sessions on inmate posttest scores in basic math. Our final predictive model, which included a pretest covariate and the number of math sessions, explains 55% of the variation in posttest scores

	Tutored	Non-Tutored
Avg. Math Pretest	5.4	5.9
Avg. Math Posttest	7.3	6.7

n = 36

Adjusted r² = 0.548

Treatment p-value : 0.08

Figure viii: Prison 2 Math Gains in Grade Equivalencies. Note: Average posttest scores are regression estimates based on average baseline information of each treatment and control group.

While the average inmate under study scored between a 5th-6th grade level in math, tutored inmates achieved just over a 7th grade math level while non-tutored inmates achieved just under a 7th grade reading level for an average difference of 1.9 grade levels and .8 grade levels, respectively. Hence, in the case study of Prison 2, tutoring added a full

(1.1) grade level per semester for low scoring inmates, accelerating test score gains by a factor of two in basic math.

With over 60 inmates serviced by the tutoring program in one semester, exogenous assignment of the treatment (i.e. tutoring services), and a proper matching procedure, Prison 2's outcomes have strong internal and external validity and can be applied widely to similarly situated prisons and tutoring models. A longitudinal study of 1-3 years, however, is critical for Prison 2 as the tutoring program was still in its infancy during the period of this study, and the average number of tutoring sessions per inmate was heavily truncated thusly. Nonetheless, these findings complement our findings from Prison 1 nicely, as the tutoring models and prison demographics were widely different.

Prison 3: Math

While Prison 3 utilized a model fairly similar to Prison 2 with similar facility demographics and a well-organized data collection and transmission processes, we found no significant correlation between inmate test score gains and inmate participation in tutoring.

Like Prison 2, Prison 3's 90 inmates recorded as participating in Petey Greene tutoring largely engaged in basic math instruction, with 88% of tutoring sessions comprised of basic math. Hence, we focused only on test score gains in math for our analysis of this facility's tutoring program.

As described earlier, the Prison 3 tutoring model provided a combination of assignment mechanisms: Some inmates self-selected into tutoring while others were assigned by teachers. In some classrooms, tutoring was individual and one-on-one, others utilized small groups, and still others included whole-class instruction by the tutor or used tutors as teacher's assistants, helping all inmates in the classroom on a short-term, as-needed basis. In our study, we only looked at inmates identified as participating in one-on-one and small group tutoring. Generally, participating inmates were those with lower math skills relative to the rest of their class. Hence, both our matching process and covariate control included math pretest scores. However, we did not find any other variables that confidently predicted either inmate participation in tutoring or inmate posttest scores (including criminal history or age).

The average inmate under study scored at roughly a sixth (~6.3) grade math level on the pretest, and both the average tutored inmate and average non-tutored inmate scored at about a 7th grade level on the posttest (~7.3), on average. While scale score gains were slightly higher for tutored inmates (35 scale score points) compared to similarly matched, non-tutored inmates (24 scale score points), the difference was not statistically significant. In other words, we cannot be confident that the difference was not due simply to chance. Our regression model over posttest scores also shows a slight but positive coefficient for participation in tutoring, yet the coefficient is similarly not statistically significant ($p = 0.53$). Likewise, no combination of control variables produced a significance value for our statistical models for Prison 3. Hence, we must conclude that we cannot reject the null hypothesis that participation in the Petey Greene Program had no effect on inmate learning gains in math during the period studied. Our outcomes analysis of test score correlations in the case of Prison 3 is inconclusive.

	Tutored	Non-Tutored
Avg. Math Pretest	511	515
Avg. Math Posttest	546	539

$n = 106$

Adjusted $r^2 = .121$

	Tutored	Non-Tutored
Avg. Math Pretest	6.2	6.3
Avg. Math Posttest	7.7	7.3

Treatment p-value : 0.53

Figure ix: Prison 3 Math Gains in Grade Equivalencies. Note: Average posttest scores are regression estimates based on average baseline information of each treatment and control group.

However, there may be a good reason for the inconclusive analysis of this facility. After conducting the quantitative evaluation, we learned that tutor folders were often incomplete, underestimating the number of sessions tutored per inmate and often misidentifying or failing to identify inmates receiving tutoring. Perhaps the primary reason for these data collection problems is that Prison 3 used different tutor folders than the other participating facilities. Prison 3 utilized folders that tracked tutors as opposed to inmates. Combined with the fact that tutors often did not have a distinct, predetermined tutoring model to follow and frequently engaged in tutoring groups of scale, both identification of individual inmates and recording the entirety of the tutoring session could have been cumbersome. Hence, while test data was complete, current, and generally accurate, tutoring data was not available.

Another reason for the lack of statistical significance could be that the variety of tutoring models employed at Prison 3 caused our analysis to deeply underestimate the impact of tutoring at the facility. When inmates move through individual, group and classroom tutoring with limited accuracy in recording practices, some inmates will receive much unrecorded tutoring and little recorded tutoring while others will receive much recorded tutoring without a great deal of tutoring overall. Similarly, the spillover effects of small group tutoring could also allow inmates not engaging in tutoring to benefit indirectly from the tutoring program, either through increased levels of attendance, motivation, or even interaction with tutored inmates.

Finally, the third potential explanation of no significant correlation could be that tutoring during classroom time simply stands as a substitute to participation in regular class instruction. In other words, tutoring may be just as effective as classroom instruction in improving inmate basic math levels. The only reason this justification might be viable is that tutoring occurred solely during class hours as opposed to supplemental to normal school hours (i.e. the Prison 1 and aspects of the Prison 2 models).

In the end, the Prison 3 model is very similar to the Prison 2 model, with the exception of a looser, more decentralized structure, lack of supplementary tutoring, and tutor-centric (as opposed to inmate-centric) materials. It is probable that our outcomes analysis of inmate test score gains would be similar if Prison 3 had a clearer tutoring model and easier-to-use materials. There may, however, have been limitations in the models and materials at Prison 3. Future evaluators should study the Petey Greene Program's

operations in Prison 3 more closely once the program has reached a more mature stage and use methods that can more accurately capture the entirety of impact, particularly including its spillover effects.

Prison 4 (Bordentown, NJ)

We did not include Prison 4 in our analysis of test score correlations because of a lack of useable testing data. Our current statistical procedures would be invalid as so few inmates within the program and within our matched control group had recent, relevant TABE scores. This is due in part because most of the inmates were preparing for the GED exam and were thus left out of TABE testing after reaching the required GED-prep score threshold. Also, Prison 4 tutors serviced a large number of ESL students, whose scores would not be reflected in our database of TABE scores (as ESL students generally use a different assessment with scores that cannot load onto the DOC's primary database).

Prison 5 (Trenton, NJ)

While Prison 5 provided a wealth of Pre-GED and in-house TABE data, particularly on reading levels, it was not possible to create a reliable control group or pretest-posttest design for this facility without significantly interfering with the program design and operations. As all inmates occupied the same classroom and as Prison 5 received a constantly high volume of tutors throughout the semester, nearly all inmates received tutoring at some point within their educational program. Furthermore, because Prison 5 classrooms primarily service higher skilled adult inmates serving shorter sentences, the facility only updated TABE scores when the inmate received high enough scores to qualify to sit for the GED exam. Hence, of the available TABE data from Prison 5, most learning gains exceed the 11th grade equivalency level, regardless of time in the classroom or number of tutoring sessions. The uniformity of the dependent variable (i.e. TABE posttest score) would invalidate our current statistical procedure. However, the Petey Greene Program contributed markedly to GED achievement at Prison 5. The following section on GED achievement describes such gains.

Findings and Methodologies Summary

In summary, two of the five facilities under study produced sufficient data and were designed in such a way to meet the rigorous standards set by the Maryland Scientific Methods Scale (MSMS) in criminal justice research and the What Works Clearinghouse (WWC) scale in education research. The tutoring program at both of these facilities demonstrated large and statistically significant improvements to inmate functional levels in math, and one facility displayed large and statistically significant improvements in reading as well.

The other facilities under study did not meet the rigorous research methodology standards of MSMS or WWC due to data inaccuracy at the collection stage, a lack of discrete treatment and control groups, missing pretest-posttest data, and non-variable outcome (posttest) data for participating inmates. Figure x provides a table summary of findings and methodologies from the learning gains study of all five participating facilities.

Discussion

In conclusion of this section, participation in the volunteer tutoring program had a significant impact on inmate test score gains, which is a standard indicator of adult learning gains, in two different subject areas: Math and Reading. These findings are corroborated by strong research methodologies and linear program designs. Facilities that could not undergo rigorous evaluation, as determined by the MSMS and WWC scales, used tutoring more for GED-level inmates than low skilled inmates, who are more likely to have reliable basic education test data. The following section thus provides an analysis of tutoring and GED achievement, which will be illuminating for the two facilities lacking in TABE test scores data.

Figure x: Learning Gains Findings and Methodology Ratings Summary

Facility	Findings	MSMS Rating	WWC Rating	Rating Rationale
Prison 1	Math: Tutored inmates gained 1.6 avg. g.e. levels over non-tutored Reading: Tutored inmates gained 2.3 avg. g.e. levels over non-tutored	4	Meets standards with reservations	Rigorous matching design, discrete treatment and comparison groups, low attrition; No randomization, small sample size
Prison 2	Math: Tutored inmates gained 1.1 avg. g.e. levels over non-tutored	4	Meets standards with reservations	Rigorous matching design, high sample size, low attrition; No randomization
Prison 3	Math: Participation in tutoring is not statistically significant	3	Does not meet standards	Matching procedures, controls for baseline differences, large sample size; No randomization, non-discrete treatment and comparison groups, high error term for tutor data accuracy
Prison 5	N/A; Outcome measure is non-variable	2	Does not meet standards	Identified treatment and comparison groups; No randomization, posttest data is non-variable
Prison 4	N/A; Pretest-posttest data not available	1	Does not meet standards	Lack of pretest-posttest data, small sample size

GED Achievement

The second largest population serviced by the Petey Greene Program included those inmates preparing to take the high school equivalency (e.g. the “GED”) exam. The learning gains of these higher skilled inmates were often not captured in regular administrations of the TABE, as these inmates were deemed ready to take the GED only after meeting a certain threshold on the TABE. This threshold is determined by the facility but generally floats around a 10th grade equivalent in all subjects. After this point, inmates take practice exams (“Pre-GEDs”) in lieu of TABE tests. Hence, one way to capture the achievements of these inmates is through a comparison of GED achievement of tutored inmates in relation to the general population (including tutored inmates). While this approach lacks the rigor of test score correlations, it nonetheless provides critical information on one major impact target for the Petey Greene Program.

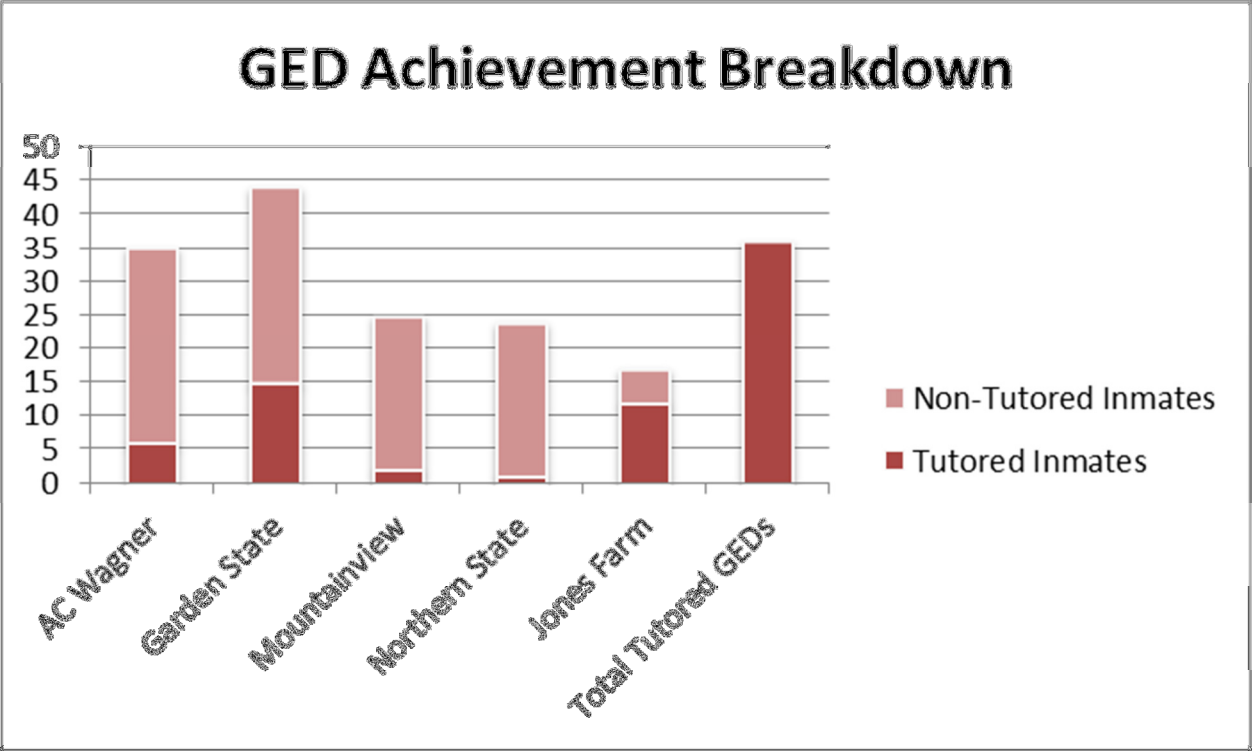


Figure xi: GED Achievement Breakdown

Contribution of Tutoring to GED Achievement

Of the just over 200 inmates serviced in one-on-one and small group tutoring during the spring 2013 semester, 36 (or approx. 17%) received their GED within the same semester. Figure xi, pictured above, illustrates the achievement of tutored inmates.

Tutored inmates at Prison 5 contributed a substantial portion to the facility's GED achievement with over 70% of inmates achieving the GED that semester also receiving tutoring. Prison 3 similarly showed a large share of inmates passing the GED also engaged in tutoring with nearly one-third (32%) of GED recipients engaging in tutoring services.

Of the tutored inmates at Prison 1, only one tested for the GED exam during the period under study. This is because the bulk of the inmates at this facility either scored very low on the TABE test or were close to passing the GED but had enough time on their sentences to wait for the next GED administration.

Prison 2 utilized its tutors primarily for low skilled adults, who would not have been prepared for GED-level instruction, and youth inmates on track for their high school diplomas. Hence, learning gains at Prison 2 would not register on GED achievement measures.

While it does not appear that the tutored inmates contributed a large share of the overall GED achievement for spring 2013, they were nonetheless a large portion of the tutored inmates at-large. More precisely, six (50%) of the twelve inmates receiving tutoring in this period passed the exam within the same period. Such success can boost morale for inmates in the tutoring program who are not yet ready for or at the cusp of GED achievement.

At the end of one semester, of all inmates passing their GED at the five facilities under study, one-quarter (25%) passed with the aid of one-on-one and small group volunteer tutoring through the Petey Greene Program.

GED Passing Efficiency

Tutored inmates not only made up a sizeable portion of the GED-achieving population for the period under study; they were also able to pass the GED at a slightly higher percentage rate than average for their respective facilities. While the average passing rate on the GED from January 2013 to July 2013 for these five facilities was 83%, tutored inmates passed at a rate of 90%. As GED-ready inmates generally only received tutoring in areas challenging to them, it seems unlikely that the +7% difference in passing rates is due solely to biased selection into the program.

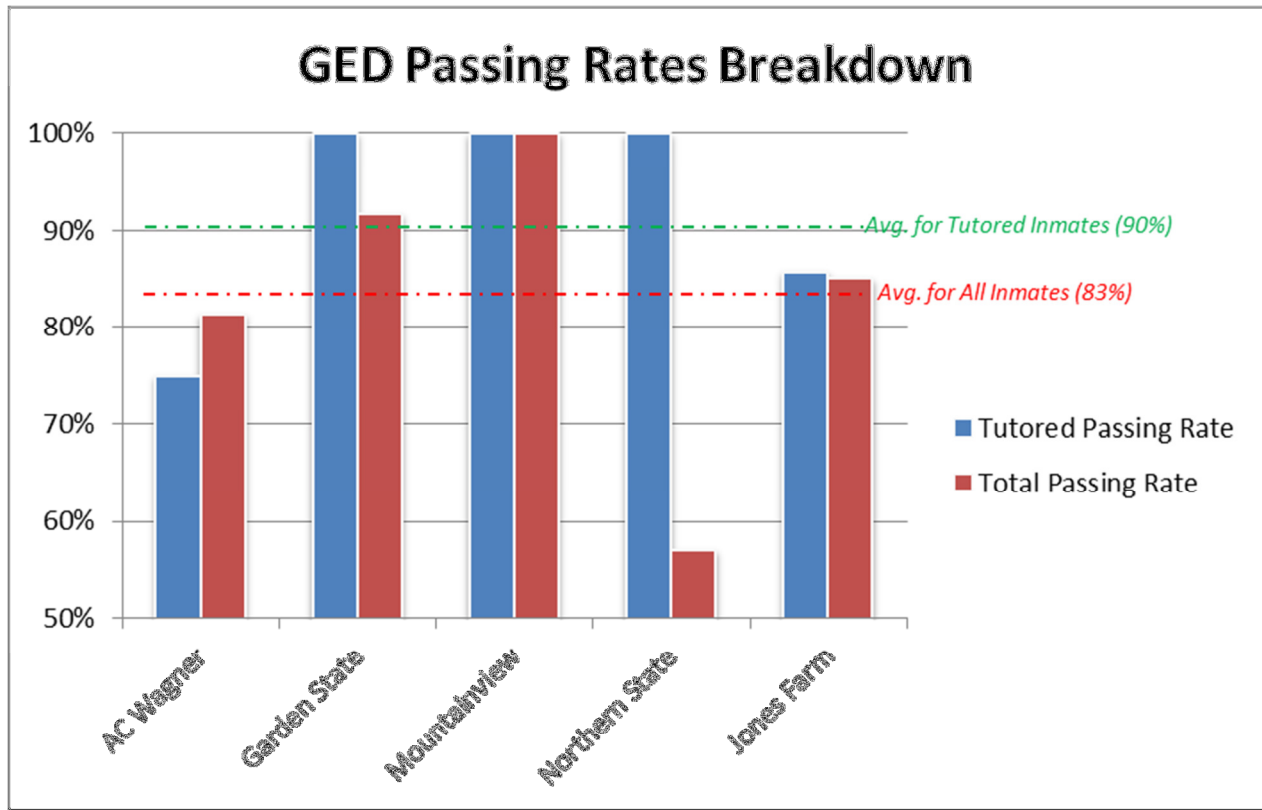


Figure xii: GED Passing Rates Breakdown

Prison 3 and Prison 5, the facilities that led the pack in its contribution of tutored inmates to the GED-achieving pool, observed increases in passing rates for tutored inmates (+8% for Prison 3 and +1% for Prison 5). All Prison 2 inmates sitting for the GED during this semester passed with a 100% pass rate, and the only tutored inmate from Prison 1 to take the test also passed on his first attempt within the time period under study. Prison 4 was the only facility with a tutored inmate pass percentage less than the facility average (-6%).

One must consider, however, that comparisons of GED achievement to facility averages are meant not as a measure of causal impact but rather as a measure of contribution. While we could not develop a control group against which to compare tutored inmates to identify any causal relationship with GED achievement, it is certainly noteworthy that in a single semester, the program helped at least 36 inmates gain their GED with a 90% pass rate. Such figures may serve as benchmarks, as opposed to impact metrics, for program expansion.

Impact Beyond the Classroom

Not all educational impact can be quantified and not all educational impact manifests itself within the confines of the classroom or the tutoring session. In our site visits, surveys, and interviews with tutors we found four distinct patterns of unintended ancillary learning experiences and facility-level effects: inmate and volunteer civic education, pre-professional training, increased educational staff accountability, and decreased levels of violence within the prison. Furthermore, upon reviewing the sentencing status of all inmates receiving tutoring, we found that over one third of all inmates in the tutoring program are due for release within the same year of tutoring. These findings are important for future evaluators in considering the full effects of the Petey Greene Program as well as for stakeholders to have a better consideration of the reach of their program.

Civic Education

In our analysis of the Petey Greene Program, we observed much educational impact that could not be quantified in this study. Educational progress not in the form of credential achievement or measureable academic progress in math and reading often came in the form of a thick civic education, for both inmates and volunteers.

Inmates asked volunteers on numerous occasions about their motivations for tutoring. Most were surprised to learn that volunteers provided their services for free with no college credit or other incentive. In an environment dominated by *quid pro quo*, altruistic volunteerism came as a great surprise to many inmates. The Petey Greene Program, then, provides inmates the experience of a civically engaged, educational community.

Volunteers also received a thorough civic education through the program. For one tutor, the program “systematically broke down the stereotypes I had about prisons and prisoners.” Immersion in the prison environment allowed volunteers to observe the implications of criminal behavior up-close. “I feel I learned many lessons beyond that of a classroom.... the improvements I made in myself and others, the ability to meet people I never would have before, the opportunity for a new shift in perspective.” Volunteers later established popular campus organizations and campaigns, including Students for Prison Education and Reform (SPEAR) and the Ban the Box- New Jersey campaign. Such engagement is indicative of college students and community members gaining a broader perspective on the types of people who constitute their political and educational communities to include inmates, probationers, and the recently incarcerated.

Work Experience

While many tutors volunteered their time to “give back” and perform “service to the community,” a significant number of volunteers joined the Petey Greene Program to

advance distinct professional goals. Such an orientation was particularly acute in volunteers from Rutgers University and The College of New Jersey. Volunteers seeking professional experience were primarily interested in the areas of policy, criminal justice and criminology, psychology, and education. Based on tutor exit surveys, we estimate roughly 20% of volunteers studying criminal justice and psychology, 10% studying education, and 25% studying the social sciences.

One volunteer joined the program “to get to interact with inmates” because she was “interested in corrections and criminal justice.” Another explained, “[The Program] gave me the opportunity to interact with inmates. It was a great experience because I would like to work with inmates in my future career.”

Other volunteers with less career certainty used the program to explore potential fields of study and work: “I am interested in majoring in secondary education. I signed up because it would be a good way for me to decide whether or not teaching is for me.” Another volunteered his services “to help towards being the solution before I, one day, am able to combat the crime problem from a more effective occupational position.” Finally, many tutors decided to teach in small groups or classroom settings in order to improve public speaking skills.

Some facilities even used extraneous volunteers to aid in administrative work. Others developed intern programs fed by the volunteer program. Hence, the Petey Greene Program provided both a civic and pre-professional education to volunteers.

Accountability

Numerous supervisors and education staff have also described the importance of education volunteers in holding education staff accountable. This principle applies particularly for larger facilities that utilize tutors during class time, as administrative supervision becomes more difficult as educational population increases. Similarly, tutors interact with teachers to keep their students accountable and supported. Volunteers at each facility recommended to teachers advancement of materials and assessments for inmates. For instance, teacher regularly responded to tutor requests for new assignments and additional resources like multiplication tables. Such individualized attention allows teachers and supervisory staff a more expanded entry point into the needs of each inmate.

“A Calming Effect”

One prison administrator has recurrently argued that the presence of volunteers has a calming effect on the inmate population, even citing a 50% reduction in violence. As the Department of Corrections’ primary mission is security, the potential, unintended consequence of violence reduction is critical. Furthermore, teachers and supervisors have observed inmate attentiveness and motivation increase when a tutor is in the classroom. Future evaluators should look more closely at the relationship between volunteerism and facility violence and educational motivation.

After Prison

Of all inmates tutored in the spring semester alone, over one-third will be released by the end of the calendar year (2013) and another 22% will be released within the following year (2014). Hence, inmates receiving tutoring will soon be looking for work, enjoying their families, and possibly even continuing their education; the contact point between tutors and inmates is crucial. On a site visit to Prison 1, one inmate who was soon to be released was working on basic language arts and math work. He said that he chose to engage in education and supplementary tutoring “to help my kids with homework and help them read.” Other inmates, particularly those from the low security unit Prison 5, needed intense, short-term, personal tutoring to achieve their GEDs before their short prison bids ended. Figure xiii shows a breakdown of tutored inmates by their expected release dates.

Roughly half of the inmates tutored at Prison 3 and Prison 2 would be released by the end of 2014 while all tutored inmates at Prison 5 are expected to leave by that time as well. Inmates tutored at Prison 4 and Prison 1 generally have more time in correctional education before release. This is most often related to the higher security levels of their inmates.

Longitudinal studies of inmates who receive tutoring are much needed. Given that over one-third of tutored inmates leave prison within the same year they receive academic tutoring, the Petey Greene Program becomes a key transitional point into the broader community for many inmates. Having a well-structured and motivational environment of personalized education immediately before release might increase the likelihood of inmates pursuing education upon release.

Expected Release Dates for Inmates Receiving Tutoring

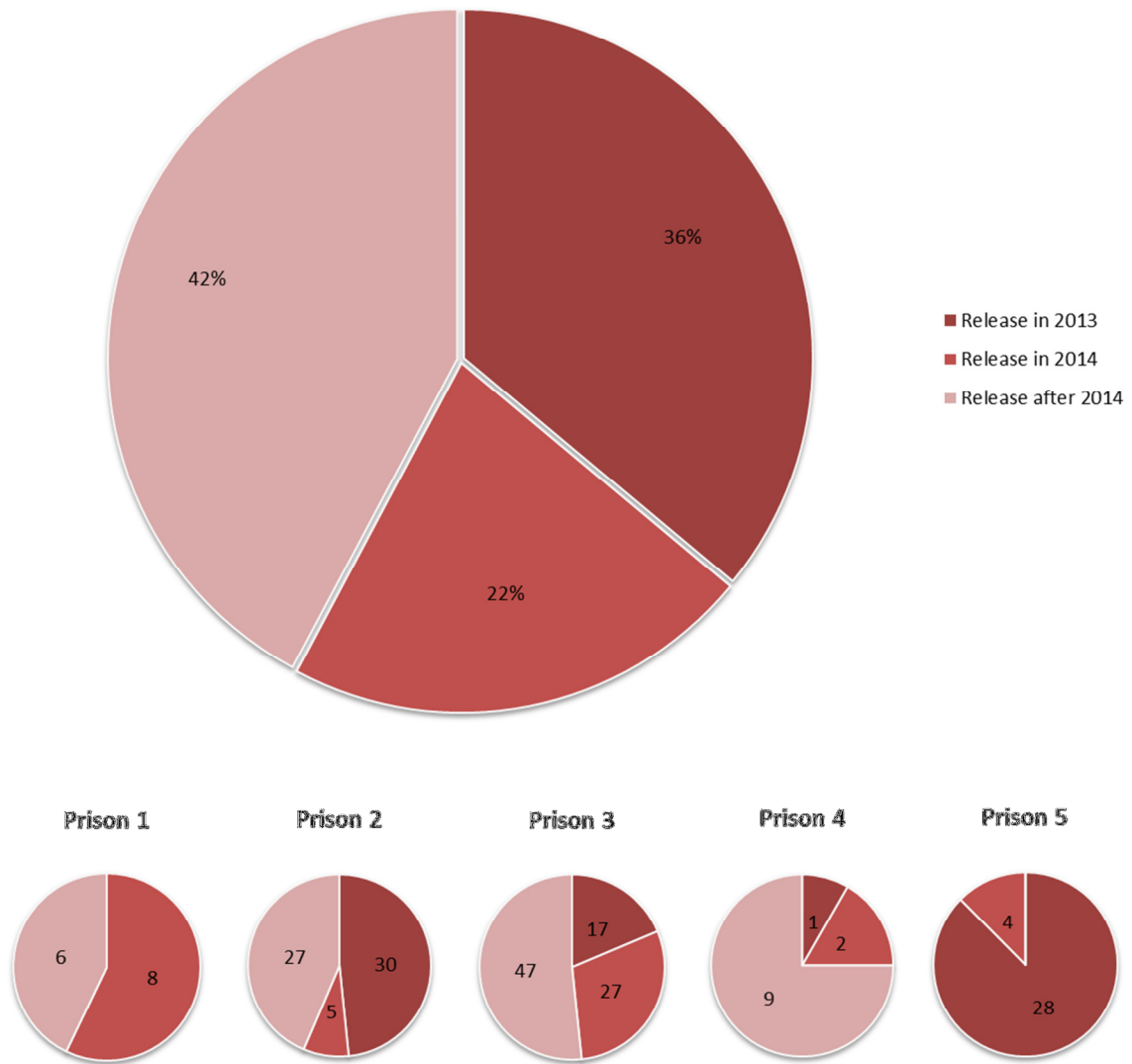


Figure xiii: Breakdown of Expected Release Time for Tutored Inmates

Evaluation Conclusions

In conclusion of our study, we believe the Petey Greene Prisoner Assistance Program is very effective in increasing both basic math and reading levels. We found that participation in the program for one semester improved inmate functional levels by at least 1.0 grade equivalencies greater than non-participation.

We assert that our analysis of the only prison for which we did not observe impact—Prison 3—used inaccurate data and underestimated the true effect of the program at the prison. While we cannot reject the inconclusiveness of our findings from Prison 3, we believe that the results from the Prison 1 and Prison 2 cases are rigorous enough to stand alone.

We have also identified other possible areas of impact, including increasing facility accountability and motivation, providing tutors with pre-professional experience and a hands-on civic education, as well as contributing to inmate GED achievement.

Furthermore, we believe that proper program organization, tutoring models, and materials were critical to program impact and accurate analysis of that impact.

Recommendations

After a close analysis of both tutoring models and tutoring program outcomes, we offer three major recommendations for program expansion and operations as well as a short note to future program evaluators:

- ❖ Clarify program goals and align services accordingly
- ❖ Use more effective tutoring models and materials based on program goals
- ❖ Develop an improved training program for volunteer tutors
- ❖ Proceed with continuous monitoring and evaluation upon program maturity.

Clarify Program Goals and Align Services

In our analysis of the tutoring program's operations within the prison, we observed that academic programs within the correctional system are highly diversified. In one prison alone, academically enrolled inmates could have been working toward their high school diploma via credit-based courses or studying for the GED exam through test-based practices and heavy self-study or even catching up with basic math and reading skills. Many facilities in New Jersey also operate postsecondary programs and supplementary academic programs.

Hence, the goals of the Petey Greene Program must align with the existing academic programs in order to gain a greater degree of focus. Otherwise, with goals that stretch the width of all academic programs, tutoring services can become disorganized. These goals should be presented to prospective volunteer tutors during orientation and continually reinforced through training.

Each of the different goals and academic tracks comes with a different set of outcomes: GEDs, high school diplomas or credits, and TABE score progress. The findings of this study should help inform the program's intended measurable outcomes and benchmarks. For instance, adult basic education students accelerated nearly two (~2.0) grade levels in math and reading as opposed to roughly one-half to one (.5-1.0) grade levels without tutoring. The Petey Greene Program should monitor, even by proxy, inmate test scores to ensure this benchmark is met each semester. Similarly, for GED-ready students, the Petey Greene Program contributed to a 90% passing rate for 36 inmates. Stakeholders should decide on short-term benchmarks and longer term goals for inmates based, at least in part, on these figures. Proper assessments and data communication *must* also follow. Without regular testing data, a rigorous quantitative evaluation is nearly impossible. Participating facilities should invest the time and resources to update test scores when college students and community members are investing their time to support educational programming.

Effective Tutoring Models and Materials

Tutoring models should be clearly delineated and should use volunteers effectively and efficiently, given facility constraints like space, resources, and security concerns. When possible, tutoring programs should prioritize pull-out and supplementary models as they add seat time rather than substituting the type of seat time (e.g. classroom instruction vs. tutoring session). Teacher or education staff supervision is critical, as established by the literature of Belzer, Borris, and Wasik.

Additionally, one-on-one or small group designs are preferable both for evaluation purposes and tutoring program design. Classroom instruction involves an entirely different set of skills, preparation, and accreditation. Tutoring should develop individuals at the fringes of the class rather than act as a substitute teacher for the entire class. Additionally, if possible, the cohort design employed by Prison 1, Prison 5, and to a lesser extent Prison 4, may produce more significant outcomes. Selecting inmates for the entire semester of tutoring helps to organize the program and all participants learn the collective goals very quickly.

Furthermore, folders should track inmates rather than tutors. Tutors who were provided both sets of folders often complained about the burden of the paperwork and frequently failed to record accurate data in the only tool that we currently have for monitoring the tutoring program. The discrepancies in materials may even account for the lack of significant results in one of our functional level cases.

Program coordinators may want to consider other means of recording the name of the inmate tutored, date, subject tutored, and worksheets completed, perhaps even using off-site technology during trips between prison and the college campus. Digital recording tools would improve the accuracy of the program monitoring metrics as well as remove the burden of paperwork from the tutor. In particular, a centralized system of data collection, input, management, and manipulation is a must for accurate program operation and evaluation.

Develop Training for Volunteers

Volunteers often requested additional training, and the literature bears the importance of ongoing professional development for volunteer tutors. As math was the subject most often tutored, followed by basic reading and language skills, tutor training should be oriented around these subjects accordingly. On the GED, math and writing tend to be subjects of chronic need as well.

The majority of volunteer tutors have no background in education, so a series of crash courses in tutoring “best practices” is paramount. Additionally, tutors should maintain high attendance. While our evaluation did not track volunteer attendance, specifically, we observed much higher attendance rates from certain colleges over others, pointing to possible cultural or policy-based reasons for absenteeism.

A Note to Future Evaluators

Throughout our evaluation, we have mentioned areas for further research. We believe more research is needed to understand “why” certain models work over others, to devise best practices for academic tutoring within the prisons, and to better evaluate program outcomes vis-à-vis program goals. In this study, we found that the program had a different model in each facility, often with slightly different goals. We also found that data collection and transmission was did not meet the standards for rigorous evaluation in some of the facilities. Hence, the program should plan another rigorous impact evaluation once the programs at each facility reach sufficient maturity for study.

To summarize a few important questions, future evaluators should consider why different models or facilities produce different outcomes, how to measure “spill-over” effects and impact on the facility at-large (rather than solely impact on the individual), how best to rigorously measure the program’s impact on GED achievement, and what effect the program has on college students and community members.

In addition, longitudinal research, including tracking inmates who have participated in tutoring services and later been released, is also important. Recidivism and employment are still the metrics of choice in the field of corrections research. The snapshot of one semester’s worth of research does not do justice to the gains of those inmates and tutors who participate in the program for years.

Upon program expansion, evaluators should incorporate evaluation principles into program design, including digital data collection tools and “random encouragement” assignment mechanisms, by which inmates are exogenously randomly encouraged to participate in tutoring while others are not. These types of evaluation designs can produce stronger causal arguments than correlations studies or matching procedures. Future evaluators will also find it critical to assess whether or not individual test scores were valid, as facilities have mentioned both lack of motivation at times and cheating at other times, both of which might alter our analysis. Data security is also a primary concern, as correctional facilities generally do not allow external access to case load data. In these cases, evaluators should partner with data personnel within the participating correctional facility to complete some of their analysis.

Appendix

Sources Cited

- Aos, S., Miller, M., and Drake, E. (2006). *Evidence-Based Adult Corrections Programs: What Works and What Does Not*. Olympia: Washington State Institute for Public Policy.
- Baker, J.D., Rieg, S.A., and Clendaniel, T. (2006). An Investigation of an after School Math Tutoring Program: University Tutors + Elementary Students = A Successful Partnership. *Education* 126:3.
- Bazemore, G., Stinchcomb, J. A Civic Engagement Model of Reentry: Involving Community Through Service and Restorative Justice. *Federal Probation* 68(2): 14-24.
- Bell, S.M., Ziegler, M., and McCallum, R.S. (2004). What adult educators know compared to what they say they know about providing research-based reading instruction. *Journal of Adolescent and Adult Literacy*, 47(7), 542-563.
- Belzer, A. (2007). Volunteer one-to-one tutoring: Critical factors in providing quality instruction. In A. Belzer (Ed.), *Toward defining and improving quality in adult basic education: Issues and challenges* (pp. 107-122), Mahwah, NJ: Lawrence Erlbaum.
- Belzer, A. (2006a). Less May Be More: Rethinking Adult Literacy Volunteer Tutor Training. *Journal of Literacy Research*. 38(2):111-140.
- Belzer, A. (2006b). What are they doing in there? Case studies of volunteer tutors and adult literacy learners. *Journal of Adolescent and Adult Literacy*, 49(7), 560-572.
- Burns, M.K., Senesac, B.J., Silberglitt, B. (2008). Longitudinal Effect of a Volunteer Tutoring Program on Reading Skills of Students Identified as At-Risk for Reading Failure: A Two-Year Follow-Up Study. *Literacy Research and Instruction*. Volume 47:27-37.
- Cecil, D., Daniella, K., Drapkin, D., MacKenzie, D.L., and Hickman, L.J. (2000). The Effectiveness of Adult Basic Education and Life-Skills Programs in Reducing Recidivism: A Review and Assessment of the Research." *Journal of Correctional Education* 51(2): 207-226.
- The Challenge of Crime in a Free Society: A Report by the President's Commission on Law Enforcement and Administration of Justice* (1967). United States Government Printing Offices: Washington, D.C.
- CIA World Factbook (2013). *United States*. www.cia.gov/library/publications/the-world-factbook/geos/us.html

- Davis, LM., Bozick, R., Steele, J.L., Saunders, J., and Miles, J.N.V. (2013). *Evaluating the Effectiveness of Correctional Education: A Meta-Analysis of Programs that Provide Education to Incarcerated Adults*. Rand Corporation.
- Fabelo, Tony (2002). The Impact of Prison Education on Community Reintegration of Inmates: The Texas Case. *The Journal of Correctional Education* 53(3):106-110.
- Farrington, D.P., Gottfredson, D.C., Sherman, L.W., and Welsh, B.C. (2002). The Maryland Scientific Methods Scale," in L.W. Sherman, D.P. Farrington, B.C. Welsh, and D.L. MacKenzie, eds., *Evidence-Based Crime Prevention*. London: Routledge, pp. 3-21.
- Fitzgerald, J. (2001). Can minimally trained college student volunteers help young at-risk children to read better? *Reading Research Quarterly*, 36:28-47.
- Gaes, Gerald G. (2008). The Impact of Prison Education Programs on Post-Release Outcomes. *From the Reentry Roundtable on Education*, March 31 and April 1, 2008.
- Gelman, Andrew and Jennifer Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Guerino, Paul, Paige M. Harrison, and William J. Sabol (2011). *Prisoners in 2010*. NCJ 236096. Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.
- Guy, Roger (2009). Cleaning up after the Mess: The Individual-Community Nexus in Correctional Rehabilitation. *Journal of Applied Social Science*, 3, 10-23.
- Harlow, C.W. (2003). *Education and correctional populations*. Washington, DC: US Department of Justice, Office of Justice Programs.
- James, Nathan (2011). Offender Reentry: Correctional Statistics, Reintegration into the Community, and Recidivism. Congressional Research Service.
- Kozol, J. (1985). *Illiterate America*. New York: New American Library.
- Lockwood, S., Nally, J.M., Ho, Taiping, Knutson, Katie (2012). The Effect of Correctional Education on Postrelease Employment and Recidivism: A 5-Year Follow-Up Study on the State of Indiana. *Crime & Delinquency*. 58(3): 380-396.
- Lutze, Faith E., Johnson, W. Wesley, Clear, Todd R., Latessa, Edward J. and Slate, Risdon N. (2011). The Future of Community Corrections is Now: Stop Dreaming and Take Action. *Journal of Contemporary Criminal Justice*. 28, 42-59.
- Messemer, Jonathan E., and Thomas Valentine (2004). "The Learning Gains of Male Inmates Participating in a Basic Skills Program." *Adult Basic Education*. 14(2):67-89.

- Morris, Darrell (2006). Using Noncertified Tutors to work with At-Risk Readers: An Evidence-Based Model. *Elementary School Journal* 106(4), 351-362.
- New Jersey Department of Corrections (2013). Progress in Mandatory Education: A4202 Stages II and III. *Part of an unpublished report presented at the State Assembly Budget Hearings for Fiscal Year 2014*.
- National Association of State Budget Officers (2011). *State Expenditure Report: Examining Fiscal 2009-2011 State Spending*. Washington DC: Author.
- Osborn, Jan; Freeman, Amy; Burley, Margaret; Wilson, Rich; Jones, Eric; et al. (2007). Effect of Tutoring on Reading Achievement for Students with Cognitive Disabilities.
- Perin, Dolores (1999). Professionalizing adult literacy: Would a credential help? *Journal of Adolescent and Adult Literacy*. Volume 42(8):610-619.
- The Pew Center on the States (2011). *State of Recidivism: The Revolving Door of America's Prisons*. Washington, DC: The Pew Charitable Trusts.
- Piotrowski, C., Lathrop, P.J. (2012). *Education*, 132(3), 683-688.
- Sample Instrument: Student Participation in Education Program, ED2, Student Monthly Contact Log (2011). Corporation for National and Community Service.
- Sandlin, J., and St. Clair, V. (2005). "Volunteers in Adult Literacy Education." In J. Coming, B. Garner, and C. Smith (Eds.), *Review of adult learning and literacy*. 5:125-154. Mahway, NJ: Lawrence Erlbaum Associates, Inc.
- Sandman-Hurley, Kelli (2008). Volunteers Tutoring Reading-Disabled Adult Literacy Learners: a Case Study. *Adult Basic Education and Literacy Journal* 2(2), 94-103.
- Sonnen, P., and Kempf, K (2008). Idaho Department of Correction Volunteer Handbook. Spencer, L.S. (2012). Program Description Booklet. Massachusetts Department of Correction.
- Tolbert, Michelle (2012). *A Reentry Education Model: Supporting Education and Career Advancement for Low-Skill Individuals in Corrections*. Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education.
- U.S. Department of Education, Office of Vocational and Adult Education (2009). *Partnerships Between Community Colleges and Prisons: Providing Workforce education and Training to Reduce Recidivism*. Washington, DC: Author.
- Volunteers in Prisons (2012). North Carolina Department of Public Safety. www.doc.state.nc.us/dop/volunteer.htm.

- Wade, Barbara (2007). "Studies of Correctional Education Programs." *Adult Basic Education and Literacy Journal* 1(1):27-31.
- Waite, P. (1983). *The role of volunteers in adult literacy programs*. Syracuse, NJ: Laubach Literacy Action.
- Wasik, Barbara A. (1997). *Volunteer Tutoring Programs: A Review of Research of Achievement Outcomes*. Center for Research on the Education of Students Placed at Risk (CRESPAR).
- What Works Clearinghouse. (2011). *Procedures an Standards Handbook (Version 2.1)*. Washington, D.C.: Institut of Education Sciences, U.S. Department of Education.
- Wolford, B.I. (1989). Correctional facilities. In S.B. Merriam and P.M. Cunningham (Eds.), *Handbook of adult and continuing education* (pp. 356-368). San Francisco: Jossey-Bass.
- Ziegler, Mary, R. Steve McCallum, and Sherry Mee Bell (2009). Volunteer Instructors in Adult Literacy: Who Are They and What Do They Know About Reading Instruction? *Adult Basic Education and Literacy Journal*. Volume 3(3):131-139.
- Ziegler, M., McCallum, R.S., and Bell, S.M. (2007). Adult educators in the United States: Who are they and what do they know about teaching reading? In J.R. Kruidenier and S.M. Bell (Eds.), *Perspectives on language and literacy* (Vol. 33, pp. 50-56). Baltimore, MD: International Dyslexia Association.

Test Score Matching and Regression Results, by Facility—Prison 2 (Math)

Matching Reg			
		Reg Type:	Logistic
VAR		Coeff.	p-value
(Intercept)		-0.1642	0.94775
Math Prettest GE Score		-0.4154	0.000216
Time in Prison		-1.0269	0.0607
Age		0.1307	0.285

Matching Balance			
VAR: Unmatched	Treatment	Control	M Difference
Math Prettest GE Score	4.64	7.47	-2.83
Time in Prison	0.23	0.55	-0.32
Age	21.61	21.79	-0.18
VAR: Matched	Treatment	Control	M Difference
Math Prettest GE Score	5.39	5.95	-0.56
Time in Prison	0.29	0.28	0.01
Age	22.27	22.07	0.2

Final Reg	Reg Type:	Linear	ANCOVA				
VAR	Coeff.	p-value	VAR				
(Intercept)	1.0925	0.2539		Df	Sum Sq	Mean Sq	F value
Math Prettest GE Score	0.9462	2.10E-07	m.ge1	1	213.611	213.611	42.2187
Number of Math Sessions	0.6339	0.0752	tut.m	1	17.045	17.045	3.3689
			Residuals	34	172.028	5.060	
Adjusted R Squared	0.548						
P-Value	5.26E-07						

Estimated Grade Equivalent		
	Tutored	Non-Tutored
Math Prettest	5.4	5.9
Math Posttest	7.6	6.7
	Tutored	Non-Tutored
25%	3.1	4.1
Math Prettest	5.4	5.9
75%	7.2	6.9
25%	4.3	5
Math Posttest	7.3	6.7
75%	12	7.9

Test Score Matching and Regression Results —Prison 3 (Math)

Matching Reg			
VAR	Coeff.	p-value	
(Intercept)	-1.492	0.0461	
Math Pret	0.00219	0.1289	

Matching Balance			
VAR: Unmatched	Treatment	Control	M Difference
Math Prettest Scale Score	516.82	490.78	26.04
VAR: Matched	Treatment	Control	M Difference
Math Prettest Scale Score	511.35	515.43	-4.08

Final Reg			
VAR	Reg Type:	Linear	
	Coeff.	p-value	
(Intercept)		370.558	1.87E-13
Math Prettest Scale Score		0.32766	1.51E-04
Tutoring Treatment		8.56968	0.52644
Adjusted R Squared		0.1205	
P-Value		6.45E-04	

Tutor Exit Survey

Office of Educational Services
NJ Department of Corrections

College/University:

Facility:

Date:

Background Questionnaire

Q1: In which city and state did you grow up?

Q2: How long have you participated in tutoring with the Department of Corrections?

Q3: Have you tutored before in any other environment? If so, for how long?

Q4: What year in college are you?

Freshman Sophomore Junior Senior Graduate Student

Q5: What is your current major/concentration and minors/certificates?

Q6: Do you receive financial aid from your school or from the government? If so, about what percentage of that aid covers your overall cost of school?

None Less than 25% 25-50% 50-75% 75% or More

Q8: What race or ethnicity do you primarily identify with?

White Black Asian Hispanic Native American Other

Q9: Do you have a family member or close friend who has ever been incarcerated?

Yes No

Q10: Can you name one teacher at the correctional facility in which you tutor?

Yes No

Q11: Can you name one inmate at the correctional facility in which you tutor?

Yes No

Q12: How many tutoring sessions have you missed this semester? What were the main reasons for missing those sessions?

Tutor Exit Survey Results.

Facility	WYCF	GYCF	MYCF	NSP	JF	Summary
College	Princeton	Princeton	Rutgers	Seton Hall	TCNJ	
Number of Participants	23	49	34	14	62	182
Number of Survey Responses	11	18	8	13	14	64
Sex	Males (5) Females (4) No Response (2)	Females (9) Males (4) No Response (5)	Females (6) Males (2)	Females (9) Males (4)	Females (8) Males (2) No Response (4)	Females (36) Males (17) No Response (11)
Race/ Ethnicity	Asian (5)	White (10)	White (5)	White (7)	White (11)	White (38)
	White (5)	Asian (5)	Asian (1)	Black (4)	Hispanic (1)	Asian (11)
	Hispanic (1)	Black (1) Other (2)	Other (1) No Response (1)	Hispanic (2)	No Response (2)	Black (5) Hispanic (4) Other (3) No Response (3)
College Year	Fresh (3)	Fresh (6)		Fresh (7)	Fresh (4)	Fresh (20)
	Soph (2)	Soph (7)	Soph (1)	Soph (1)	Soph (3)	Soph (13)
	Junior (3)	Junior (1)		Junior (2)	Junior (3)	Junior (9)
	GS (1) CM (2)	Senior (3) GS (1)	Senior (6) No Response (1)	Senior (3)	Senior (2) No Response (2)	Senior (14) GS (2) CM (2) No Response (3)
Home State	PA (2), CA, CT, IN, MA, TX, VA, WA, VT, Argentina	CA (3), PA (3), NY (2), AK, IA, IN, MA, MO, NC, WA, WV, N. Ireland, New Zealand	NJ (7), CT	NJ (8), CA(2), IL, MO, PA	NJ (13), No Response (1)	NJ (28) PA (6) CA (6) Other (21) No Response (1)
Field of Study	WWS (3), PSY (1), PHY (1), PHI (1), Math (1), EEB (1), Undecided (1), NA (2)	WWS (4), POL (2), SOC (2), ANT (1), BIO (1), CHEM (1), CBE (2), COS (1), COM (1), MAE (1), SLAV (1)	CRIM and PSY (5), PSY (2), POL (1)	Nursing (4), Languages (2), PSY (1), Marketing (1), BIO (1), ANT (1), Education (1), Sports Manag. (1), Undecided(1)	CRIM and PSY (6), Education (5), HIS (1), Undecided(1), No Response(1)	Social Sciences (17) Criminology/Psych. (14) STEM (9) Education (6) Humanities (6) Business (2) Undecided (3) No Response (3)
Financial Aid	None (2)	None (6)	None (1)	None (2)	None (7)	None (18)
	50-75% (4)	Less than 25% (2)	Less than 25% (1)	Less than 25% (1)	Less than 25% (1)	Less than 25% (5)
	75%+ (3)	25-50% (1)	25-50% (1)	25-50% (3)	25-50% (2)	25-50% (7)
	NR (2)	50-75% (4)	50-75% (1)	50-75% (1)	50-75% (2)	50-75% (12)
Time in Program	1.41 years (.5-4 years)	1.11 years (.5-2 years)	.50 years (.5 years)	.58 years (.5-1 years)	.68 years (.5-1 years)	.87 years (.5-4 years)
Experience Tutoring	73% some experience	87.5% some experience	42.9% some experience	53.9% some experience	42.9% some experience	61.9% some experience
Friend/Family in Prison?	Yes (1), No (10)	Yes (0), No (18)	Yes (1), No (7)	Yes (4), No (9)	Yes (0), No (14)	Yes (6), No (58)
Identify a Teacher?	90.9%	100.0%	85.7%	61.5%	100.0%	85.9%
Identify a Student-Inmate?	90.9%	94.4%	71.4%	100.0%	100.0%	93.7%
Number of Sessions Missed	1.6 Per Tutor Average	2.21 Per Tutor Average	1.0 Per Tutor Average	1.23 Per Tutor Average	1.14 Per Tutor Average	1.44 Per Tutor Average