Analysis of Associations between Contemporaneous Job Corps Performance Measures and Impact Estimates from the National Job Corps Study

Final Report

January 31, 2011

Jane Fortson
Peter Z. Schochet

MATHEMATICA
Policy Research, Inc.

Contract Number:
DOLQ091A20941/DOLU101A21559

Mathematica Reference Number:
06808-300

Submitted to:
U.S. Department of Labor
200 Constitution Ave., NW
Washington, DC 20210
Project Officer: Bogdan Tereshchenko

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Peter Schochet

Analysis of Associations between Contemporaneous Job Corps Performance Measures and Impact Estimates from the National Job Corps Study

Final Report

January 31, 2011

Jane Fortson
Peter Z. Schochet

ACKNOWLEDGEMENTS

We are grateful to Mark Flick and Davin Reed for their excellent research assistance. John Burghardt provided helpful comments on both the design memorandum and this report. Terry Johnson at Battelle Memorial Institute provided performance measure data for program years 1994–1996, and also provided useful background on the performance measurement system and its changes over time. Finally, we are grateful to Celina Kamler for help preparing this report, to Walt Brower for editorial assistance, and to Jill Miller, Sheena Flowers, Linda Heath, and Dawn Patterson for secretarial support.

Jane Fortson
Peter Z. Schochet

CONTENTS

TABLES

FIGURES

EXECUTIVE SUMMARY

Job Corps is the nation's largest vocationally focused education and training program for disadvantaged young people. It serves young men and women between the ages of 16 and 24 at 124 center campuses nationwide, primarily in residential settings. The program's goal is to prepare young people for successful careers. Each year, Job Corps serves more than 60,000 students at a cost of about $1.5 billion, which is more than 60 percent of all funds spent by the U.S. Department of Labor on youth training and employment services. To examine the effectiveness of Job Corps, the Department's Employment and Training Administration sponsored the National Job Corps Study (NJCS) in 1993.

The NJCS used survey and administrative earnings records data to estimate the Job Corps program's average impacts on students' employment and related outcomes. From late 1994 to early 1996, nearly 81,000 young people nationwide were randomly assigned to either a treatment group, who were allowed to enroll in Job Corps, or a control group, who were not allowed to enroll for a period of three years. NJCS findings are based on comparisons of the outcomes of about 9,500 treatment group members in the research sample and 6,000 control group members. The main impact analysis found that Job Corps improved education and training outcomes (such as the receipt of General Educational Development and vocational certificates and time spent in school), significantly reduced criminal activity, and improved earnings and employment outcomes in the two years after program exit, although the longer-term analysis did not demonstrate that impacts were sustained beyond the two-year period (Schochet et al. 2008).

The NJCS also examined the extent to which impacts (average treatment-control differences) on key outcomes were associated with the aggregate overall center performance measure used by Job Corps. The NJCS found that impacts on key outcomes were *not* associated with the overall aggregate measure of center performance (Schochet and Burghardt 2008). Students in higher-performing centers had better outcomes; however, the same pattern was observed for the *control* group members who would have been assigned to those centers.

This study extends the previous work analyzing the relationship between Job Corps performance measures and center-level impact estimates from the NJCS. We examine whether adjusting performance measures for student characteristics results in positive statistical associations between performance measures and impacts. Using linear regression, we constructed new measures of center performance that adjust for differences in individual and local area characteristics of center participants to measure the component of center performance that is not explained by these characteristics. We ran separate regressions on data from ETA-652 intake forms and from the NJCS baseline survey, adjusted not only the aggregate measure but also component performance measures in the Job Corps categories of program achievement, placement, and quality/compliance, and considered three years' worth of performance measure data. Our goal was to test for any associations between center-level impacts and unadjusted or adjusted performance measures.

We address the following research questions:

- Are center performance rankings changed by regression adjustment?
- To what extent are regression-adjusted performance measures better able to distinguish between centers with larger impacts and those with smaller impacts?

- Are there specific performance measures (either unadjusted or adjusted) that are more associated with impacts than others or the summary (overall rating) measure that was used for the NJCS?

Figure 1 illustrates the steps in our analysis. First, we gathered performance measure data for all three years of the NJCS. After linking NJCS participants to Job Corps centers using intake counselors' predictions of center assignment (available for both the treatment and control groups), we aggregated student characteristics to the center level. Then, we used regression models to adjust center performance measures for average student characteristics. We calculated NJCS impacts at the center level for seven different outcome measures, including educational services, educational attainment, arrests, and earnings. Finally, we compared adjusted and unadjusted performance measures (including different components and different program years) to center-level impact estimates.

Our overall finding is that although regression adjustment changes the performance rankings of centers, *the adjusted performance ratings remain uncorrelated with center-level impacts*. In particular, we find that:

- The mix of students in high-performing centers is modestly different from students in low-performing centers (Table 1).
- Regression-adjusting for characteristics accounts for some of the variance in the performance measures and changes center rankings, albeit not dramatically. Regression-adjusted and unadjusted performance measures are positively correlated (Table 2).
- The correlations between impacts and performance measures are generally not significantly different from zero (Table 3). It is noteworthy that regression-adjusted performance measures are no better than unadjusted performance measures in this regard. We find similar results whether we use the ETA-652 data or the more detailed NJCS baseline survey data.
- Our findings hold for overall measures of performance as well as components of center performance and different program years; that is, the relationship between impacts and different performance measure components is also generally weak and shows no consistent pattern.
- Similarly, among the subgroups we analyzed, there are no particular groups of centers for which relationships between performance measures and impacts are significant.
- In contrast, when we create center-level "performance measure" constructs using treatment group outcomes from the NJCS follow-up survey data, we find them to be positively correlated with the NJCS impact estimates. Exploring these findings may be an avenue for future research.

In conclusion, although regression adjustment had some effect on the performance rankings, it did not make performance measures any more predictive of impacts. While the baseline covariates explain some portion of the variance in the performance measures, important unobserved differences among centers, possibly related to their propensity to produce impacts, appear to remain.

Figure 1. Design of Analysis of Job Corps Performance Measures and Impact Estimates

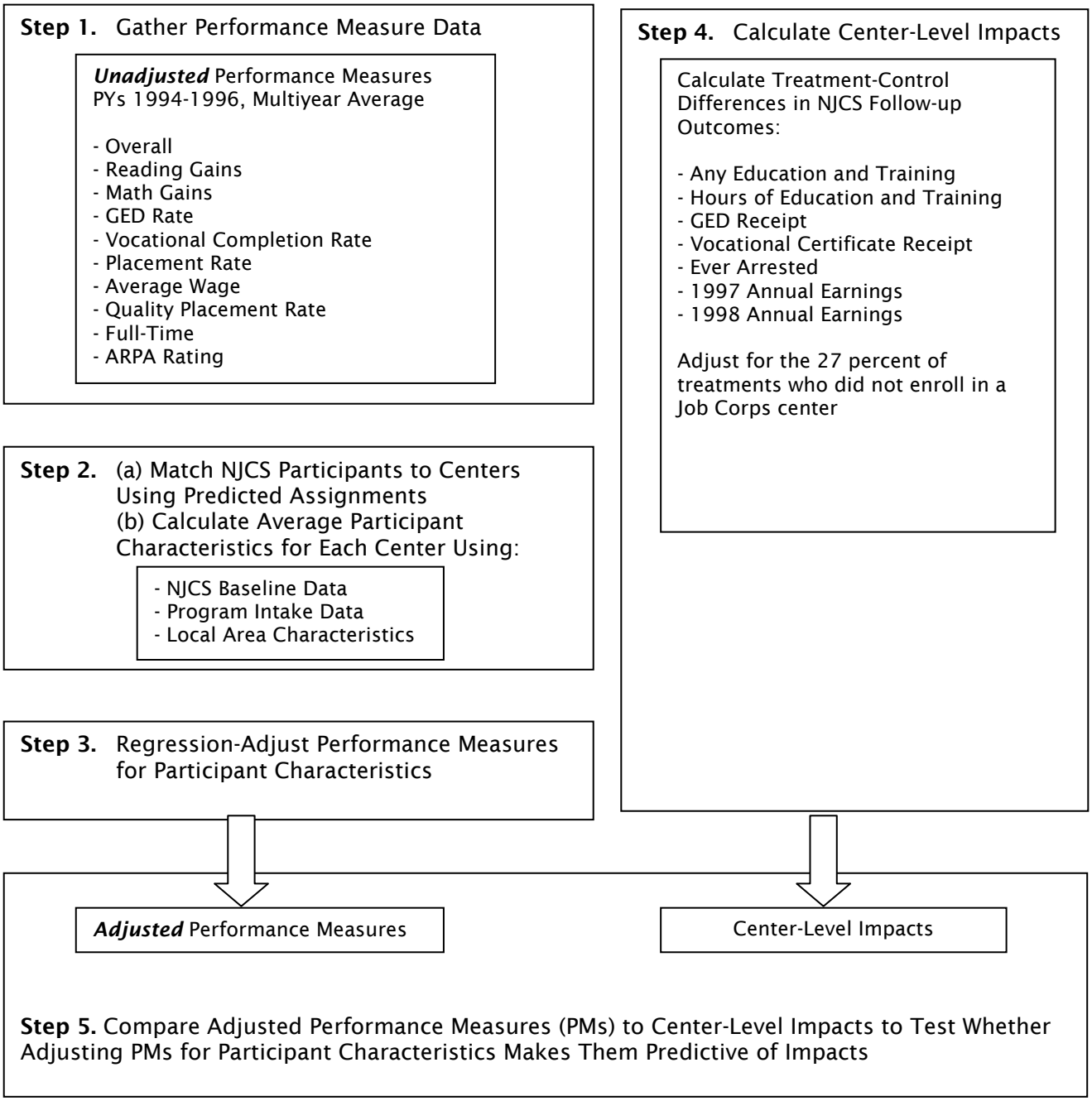


Table 1. Average Baseline Characteristics, by Overall Center Performance Tercile

	Unadjusted Overall Center Performance Tercile			
	Low	Medium	High	*p-value*
Selected NJCS Baseline Characteristics				
Demographic Characteristics				
Non-Hispanic White	0.245	0.344	0.391	0.023**
Non-Hispanic Black	0.558	0.479	0.249	0.000***
Hispanic	0.139	0.106	0.227	0.017**
Other race	0.058	0.071	0.133	0.040**
Female	0.388	0.361	0.399	0.663
Native language English	0.897	0.916	0.799	0.003***
Native language Spanish	0.061	0.048	0.124	0.026**
Native language other	0.042	0.035	0.077	0.066*
Age 15-17	0.443	0.447	0.427	0.708
Age 18-20	0.395	0.401	0.415	0.498
Age >20	0.162	0.153	0.158	0.817
Education and Skills				
High school degree	0.151	0.152	0.191	0.034**
GED	0.042	0.047	0.057	0.147
Vocational degree	0.025	0.024	0.029	0.465
Highest grade completed 0-8	0.167	0.161	0.125	0.008***
Highest grade completed 9-11	0.659	0.657	0.649	0.803
Highest grade completed >11	0.174	0.181	0.226	0.009***
Employment History				
Currently working	0.202	0.205	0.228	0.203
Earnings in past year <$1,000	0.507	0.483	0.497	0.590
Earnings in past year $1,000-$4,999	0.285	0.296	0.292	0.771
Earnings in past year $5,000-9,999	0.142	0.146	0.139	0.821
Earnings in past year >$10,000	0.066	0.075	0.071	0.596
Physical or emotional problem that limited work	0.054	0.050	0.058	0.517
Family Status				
Has child	0.193	0.168	0.148	0.142
Socioeconomic Status				
Did not receive food stamps over past year	0.574	0.602	0.630	0.088*
Received food stamps some of past year	0.058	0.074	0.065	0.092*
Received food stamps all of past year	0.368	0.325	0.306	0.035**
Criminal History				
Ever arrested	0.267	0.298	0.301	0.231
Drug Use				
Used no drugs over past year	0.711	0.663	0.622	0.001***
Used hard drugs occasionally over past year	0.040	0.058	0.101	0.000***
Used hard drugs frequently over past year	0.005	0.014	0.026	0.000***
Selected ETA-652 Baseline Characteristics				
Demographic Characteristics				
Male	0.612	0.639	0.601	0.663
Non-Hispanic White	0.266	0.368	0.429	0.020**
Non-Hispanic Black	0.587	0.511	0.269	0.000***
Hispanic	0.109	0.079	0.197	0.020**
American Indian	0.028	0.032	0.075	0.249
Asian	0.010	0.010	0.031	0.005***
Age 14-17	0.446	0.447	0.429	0.712
Age 18-20	0.393	0.401	0.413	0.493
Age >20	0.161	0.153	0.158	0.857
Education and Skills				
Highest grade completed 0-8	0.178	0.175	0.124	0.002***
Highest grade completed 9-11	0.638	0.634	0.643	0.873

	Unadjusted Overall Center Performance Tercile			
	Low	Medium	High	*p-value*
Highest grade completed >11	0.184	0.191	0.233	0.020**
Employment History				
Estimated annual income $0-$400	0.010	0.014	0.022	0.042**
Estimated annual income $401-$6,528	0.236	0.265	0.283	0.210
Estimated annual income >$6,529	0.256	0.266	0.245	0.677
Estimated annual income missing	0.497	0.455	0.451	0.377
Socioeconomic Status				
Receiving public assistance	0.458	0.403	0.389	0.034**
Health and Health Care				
Covered by health insurance or Medicaid	0.328	0.363	0.432	0.013**
Local Area Characteristics				
Demographic Characteristics				
Percentage white	0.708	0.740	0.790	0.014**
Percentage black	0.221	0.195	0.093	0.000***
Average household size	2.723	2.683	2.789	0.037**
Percentage urban	0.750	0.699	0.772	0.127
Percentage of families with a female head	0.199	0.192	0.171	0.020**
Percentage foreign-born	0.621	0.563	0.835	0.226
Total births	10512	7805	18741	0.037**
Percentage of births to teens <18 years	0.057	0.051	0.049	0.032**
Crime				
Deaths by homicide and legal intervention (rate)	0.000	0.000	0.000	0.012**
Percentage of population in juvenile institutions	0.000	0.000	0.001	0.023**
Economic Characteristics				
Percentage of families in poverty	0.140	0.121	0.121	0.129
Median household income	31726	33230	34064	0.049**
Percent households with income:				
<$5,000	0.089	0.076	0.066	0.000***
$5,000-$9,999	0.107	0.104	0.101	0.339
$10,000-$14,999	0.099	0.095	0.097	0.590
$15,000-$24,999	0.186	0.184	0.187	0.810
$25,000-$49,999	0.320	0.330	0.338	0.003***
$50,000-$99,999	0.166	0.174	0.176	0.593
>$100,000	0.032	0.036	0.034	0.707
Unemployment rate, 16+	0.061	0.060	0.067	0.117
Number of Centers	**33**	**33**	**34**	

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF.

Notes: All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Terciles are based on the three-year average overall rating. The reported p-value refers to an F-test which tests whether the three groups are jointly significant.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

Table 2. Correlations between Unadjusted Center Performance and Adjusted Center Performance, All Components in All Years

	Correlation Between Unadjusted and Adjusted Performance Measures							
	NJCS-Adjusted				ETA-652-Adjusted			
Performance Measure	PY94	PY95	PY96	Multiyear Average	PY94	PY95	PY96	Multiyear Average
Overall	0.54	0.69	0.65	0.58	0.48	0.58	0.59	0.53
Reading Gains	0.54	0.74	--	0.74	0.61	0.67	--	0.64
Math Gains	0.66	0.82	--	0.75	0.55	0.64	--	0.56
GED Rate	0.53	0.64	0.64	0.49	0.39	0.54	0.57	0.45
Vocational Completion Rate	0.63	0.71	0.71	0.77	0.63	0.65	0.67	0.65
Placement Rate	0.55	0.56	0.51	0.55	0.48	0.51	0.52	0.49
Average Wage	0.59	0.40	0.48	0.46	0.45	0.41	0.48	0.40
Quality Placement	0.74	0.66	0.65	0.63	0.55	0.53	0.54	0.49
Full-Time	--	0.55	0.65	0.56	--	0.50	0.54	0.52
ARPA Rating	0.72	0.62	--	0.66	0.60	0.62	--	0.58

Sample size = 100 centers.
Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF.

Notes: All correlations are statistically significant at the 1 percent level. NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. NJCS adjustment controls for participant characteristics from the NJCS baseline survey using a forward-selection stepwise regression with inclusion and exclusion p-value thresholds of 0.20. ETA-652 adjustment controls for participant characteristics from the ETA-652 intake form, with all variables included in the model. Both adjustments use center-level averages of participant characteristics for the NJCS baseline sample. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight.

Table 3. Correlations between Center-Level Impacts and Multiyear Average Performance Ratings (Unadjusted, NJCS-Adjusted, and ETA-652-Adjusted)

	Overall Rating			GED Rating			Vocational Completion Rating			Average Wage Rating			Placement Rating		
Outcome for Impact Estimate	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj
Any Educational Services	-0.02	-0.06	-0.01	-0.08	0.08	-0.02	0.08	0.06	0.07	0.09	0.13	0.15	0.05	-0.75	0.04
Hours of Educational Services	0.17*	-0.03	0.08	0.02	-0.03	-0.03	0.19*	0.13	0.06	-0.01	-0.04	-0.04	0.19*	0.16	0.12
GED Receipt	0.15	-0.08	-0.10	0.12	-0.10	-0.11	0.13	0.02	-0.12	0.05	0.10	-0.08	0.13	0.06	-0.10
Vocational Certificate Receipt	0.13	-0.04	-0.01	0.00	-0.06	-0.14	0.14	0.04	0.03	-0.11	0.05	-0.07	0.23**	0.12	0.08
Ever Arrested	-0.02	-0.06	-0.09	0.02	0.07	-0.03	-0.06	-0.10	-0.12	-0.04	-0.08	-0.01	-0.04	-0.04	0.03
1997 Annual Earnings	-0.14	-0.19*	-0.22**	-0.22**	-0.32***	-0.25**	-0.08	-0.05	-0.13	0.07	0.03	-0.18*	0.03	0.04	-0.01
1998 Annual Earnings	-0.09	-0.11	-0.11	-0.28***	-0.32***	-0.23**	-0.02	-0.01	-0.01	0.05	0.02	-0.18*	0.08	0.11	0.03

Sample size = 100 centers.
Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: Table shows the correlation based on a multiyear average of the center's performance rating and the center-level impact estimate. NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. NJCS adjustment controls for participant characteristics from the NJCS baseline survey using a forward-selection stepwise regression with inclusion and exclusion p-value thresholds of 0.20. ETA-652 adjustment controls for participant characteristics from the ETA-652 intake form, with all variables included in the model. Both adjustments use center-level averages of participant characteristics for the NJCS baseline sample. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

I. INTRODUCTION

This study analyzes the relationship between Job Corps performance measures and center-level impact estimates from the National Job Corps Study (NJCS). Our study focuses on whether adjusting performance measures for the characteristics of students served changes the association between performance measures and impacts. In this Chapter, we briefly describe the Job Corps program and the NJCS before discussing our key research questions and main findings.

Job Corps is the nation's largest vocationally focused education and training program for disadvantaged young people. It serves young men and women between the ages of 16 and 24, primarily in a residential setting. The program's goal is to prepare young people for successful careers. Each year, Job Corps serves more than 60,000 students at a cost of about $1.5 billion, which is more than 60 percent of all funds spent by the U.S. Department of Labor on youth training and employment services. Apart from Job Corps, the Department's Employment and Training Administration (ETA) supports the youth formula program under the Workforce Investment Act of 1998 and an array of specialized discretionary youth development grants. To examine the effectiveness of Job Corps, ETA sponsored the NJCS in 1993.

The NJCS used survey and administrative earnings records data to examine the Job Corps program's average impacts on students' employment and related outcomes. From late 1994 to early 1996, nearly 81,000 young people nationwide were randomly assigned to either a treatment group, who were allowed to enroll in Job Corps, or a control group, who were not allowed to enroll for a period of three years. NJCS findings are based on comparisons of the outcomes of about 9,500 treatment group members in the research sample and 6,000 control group members. The remaining treatment youth were allowed to enroll in Job Corps, but were not in the research sample (for which data were collected). The main impact analysis found that Job Corps improved education and training outcomes (such as the receipt of General Educational Development [GED] and vocational certificates and time spent in school), significantly reduced criminal activity, and improved earnings and employment outcomes in the two years after program exit, although the longer-term analysis did not demonstrate that impacts were sustained beyond the two-year period (Schochet et al. 2008).

The NJCS also examined the extent to which impacts (average treatment-control differences) on key outcomes were associated with the aggregate overall center performance measure used by Job Corps. The Job Corps performance measurement system gathers data that are used to rate Job Corps centers on the outcomes of their participants. During PYs 1994-1996, the system included eight or nine different measures in three areas: (1) *program achievement* measures, including reading gains, math gains, the rate of attainment of a GED certificate, and the vocational completion rate; (2) *placement* measures, including the placement rate, the average wage at placement, the percentage of full-time placements, and the percentage of quality placements (defined as the percentage of placements in jobs that matched the area of training); and (3) *quality/compliance* measures, including one measure developed from observations made by regional office monitors during center reviews.[1] The key research question for this analysis was, Did higher-performing centers produce larger

[1] The current Job Corps performance measurement system uses performance measures that are similar to but different than those used during PYs 1994-1996. The current performance measurement system is described in more detail on page 7.

impacts than lower-performing centers? This question is policy relevant because Job Corps is a performance-driven program; key programmatic decisions are made based on how well centers perform.

The NJCS found that impacts on key outcomes were *not* associated with a summary measure of center performance (Schochet and Burghardt 2008). Students in higher-performing centers had better outcomes; however, the same pattern was observed for comparable control group members in these centers. Thus, at the time of the NJCS, the performance measurement system was not achieving the goal of ranking and rewarding centers on the basis of their ability to improve participant outcomes relative to what these outcomes would have been otherwise.

One potential reason for the finding of a lack of association between center impacts and performance measures is that most components of the performance measurement system were not adjusted for the characteristics of students that each center served, which may not be under a center's control. This study generates regression-adjusted performance measures that account for the characteristics of participants that centers serve. The study's goal is to test whether center-level impacts are associated with unadjusted and adjusted performance measures, both overall and disaggregated. In particular, we address the following research questions:

Q1. Are center performance rankings changed by regression adjustment?

Q2. To what extent are these regression-adjusted performance measures better able to distinguish between centers with larger impacts and those with smaller impacts?

Q3. Are there specific components of the performance measures (either unadjusted or adjusted) that are more associated with impacts than other component measures or the summary (overall rating) measure that was used for the NJCS?

To address these questions, we used NJCS survey and intake data and contemporaneous center performance measure data. Figure I.1 illustrates the steps in our analysis. First, we gathered performance measure data for all three years of the NJCS, including different components in each year. Then, we linked NJCS participants to Job Corps centers using intake counselors' predictions of center assignment (which are available for both research groups) and aggregated participant characteristics to the center level. Our third step was to regression-adjust center performance measures for average participant characteristics; this step resulted in the creation of several sets of adjusted performance measures. Next, we calculated NJCS impacts at the center level for seven different outcome measures; in the final step, we compared these impacts to the adjusted performance measures.

Our overall finding is that although regression adjustment changes the performance rankings of centers, the adjusted performance ratings remain uncorrelated with center-level impacts. Our key findings related to each research question are as follows:

A1. Students in high-performing centers are significantly different from students in low-performing centers. For most characteristics on which they differ, students in high-performing centers are more likely to have characteristics associated with better outcomes; however, there are a handful of disadvantageous characteristics that are more common among students in high-performing centers. Furthermore, differences across centers in the characteristics of students they serve are relatively small in magnitude. Regression-adjusting for characteristics changes center performance rankings to some

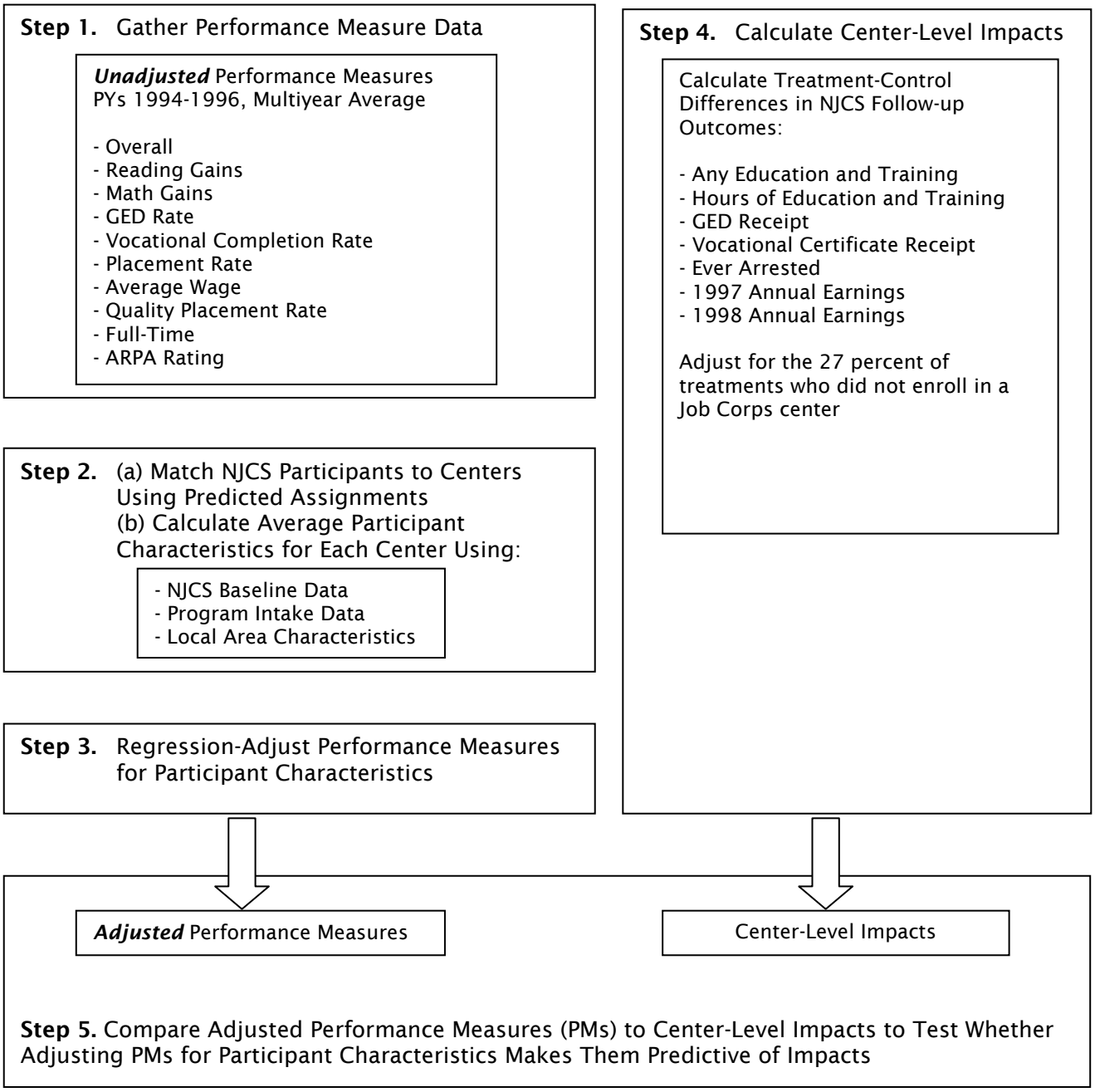
extent, but regression-adjusted and unadjusted performance measures are nevertheless positively correlated.

A2. The regression-adjusted performance measures are no better than unadjusted performance measures at distinguishing between centers with larger impacts and those with smaller impacts. The correlations between NJCS center-level impact estimates and contemporaneous performance measures are generally weak and insignificant.

A3. The relationship between impacts and different performance measure components is also generally weak and insignificant.

In Chapter II, we describe the data sources used in this analysis. We explain our analysis plan in Chapter III and present results in Chapter IV. Chapter V presents conclusions.

Figure I.1. Design of Analysis of Job Corps Performance Measures and Impact Estimates



II. DATA SOURCES

Our study draws on data from seven sources: (1) Job Corps performance measurement system data, (2) baseline data on NJCS participants from the NJCS baseline survey, (3) baseline data from program intake forms, (4) data on local area characteristics, (5) data on Job Corps center characteristics, (6) Job Corps intake counselors' predictions of center assignment for NJCS participants, and (7) follow-up data from the NJCS. These data are described in sections II.A to II.F.

A. Performance Measurement System Data

The Job Corps performance measurement system gathers data that are used to rate Job Corps centers on the outcomes of their participants. At the time of the previous study—Program Years (PYs) 1994 to 1996—the Job Corps performance measurement system included eight or nine measures in three areas: (1) *program achievement* measures, including reading gains, math gains, the rate of attainment of a GED certificate, and the vocational completion rate; (2) *placement* measures, including the placement rate, the average wage at placement, the percentage of full-time placements, and the percentage of quality placements (defined as the percentage of placements in jobs that matched the area of training); and (3) *quality/compliance* measures, including one measure developed from observations made by regional office monitors during center reviews.

With the exception of the GED completion rate and the average wage rate, these measures were not adjusted for the characteristics of students or their local areas. Instead, a center's score on each measure was compared to a fixed standard; the performance measure is the percentage of the goal that was met.[2] These percentages were then weighted to yield a summary measure of how well a center performed relative to its standards. Table II.1 summarizes the different performance measures and indicates the program years in which each component is available.

Appendix Table A.1 provides more details on the measures, the samples used to calculate each component, and the weights used to calculate the summary performance measure. For example, for the reading gains measure in PY 1994, the pool of students included those who scored less than 8.5 on the TABE 5/6 total reading test at program entry (or who did not take the test), the performance measure was the percentage of students in the pool who gained two grades or scored 8.5 on the follow-up TABE reading test, and the performance standard was 30 percent. The PY 1994 final reading gains measure was then calculated as the percentage of the 30 percent standard that the center met, and this measure was assigned a weight of 5.6 percent in calculating the summary measure.

As discussed in more detail below, the Job Corps performance measures are conceptually aligned to the key NJCS outcomes that are used for the analysis. However, there are important differences, such as the pool of students used for the constructs and data sources.

[2] For the GED completion rate and average wage rate, Job Corps used a regression model to establish performance standards. The GED model accounted for state differences in GED requirements, and the wage model controlled for differences in prevailing wages across geographic areas.

Table II.1. Job Corps Center Performance Measures, PYs 1994-1996

	Years Available		
Measure	PY 1994	PY 1995	PY 1996
Reading Gains. Percentage of students who gain two grades or score above a threshold on follow-up TABE reading test (among those who did not take or scored less than the threshold on TABE 5/6 total reading test at program entry).	X	X	
Math Gains. Percentage of students who gain two grades or score above a threshold on follow-up TABE math test (among those who did not take or scored less than the threshold on TABE 5/6 total math test at program entry).	X	X	
GED Rate. Percentage of students who obtain GED/high school degree, including bonus for students who initially score low on test (among those without high school diploma and who either did not take or scored more than a threshold on TABE 5/6 total reading test at program entry).	X	X	X
Vocational Completion Rate. Percentage of students who complete vocation at completer or advanced-completer level (depending on the year, either among all terminees or among those who stayed at least 60 days and participated in a vocational program with an approved training achievement record).	X	X	X
Placement Rate. Percentage of students placed in job/military or school, with bonus for advanced training (AT) or advanced career training (ACT) transfers (among terminees and Job Corps AT or ACT transfers).	X	X	X
Average Wage. Average wage of students placed in a job/military.	X	X	X
Quality Placement Rate. Percentage placed in a job training match, with or without a bonus for students placed in college or AT/ACT transfers (depending on the year). Measured among either all job/military completers or vocational completers with a placement record and those with a record that was due but not received (depending on the year).	X	X	X
Full Time. Percentage of students placed who are full-time among students placed in a job/military.		X	X
ARPA Rating. Regional office rating of center quality/compliance.	X	X	
Overall Rating. Weighted average of individual ratings (measure/standard).	X	X	X

Notes: "PYs 1994-1996" refers to PY 1994 (July 1, 1994 to June 30, 1995), PY 1995 (July 1, 1995 to June 30, 1996), and PY 1996 (July 1, 1996 to June 30, 1997). Schochet and Burghardt (2008) provide additional detail on performance measures and the weights used in constructing the overall rating. In Appendix Table A.1, we reproduce their Table 1.

Schochet and Burghardt (2008) used this summary measure (the overall rating score) to examine the association between center performance levels and center impacts. The summary measure covered the PY 1994 to 1996 periods, because that was when NJCS treatment group members were enrolled in Job Corps centers.

This study required full performance data covering the program year (PY) 1994 to 1996 periods, including performance components from each year. We obtained these data from Battelle Memorial

Institute, which calculates Job Corps performance measures for DOL. Because random assignment occurred between late 1994 and early 1996, the treatment group sample participated in Job Corps over a relatively long period. Thus, the main performance measures for our analysis were constructed using three-year (and two-year, where applicable) averages of the performance ratio (i.e., the center's score relative to its standard) for the overall rating and for each component. This approach allowed us to examine the association between center impacts and average (typical) center performance to which the treatment group sample was exposed. However, to examine the sensitivity of our results to changes in performance measures across years, we examined the association between performance measures for each PY and center impacts (that did not vary across the analyses). We restricted our analysis to the 102 Job Corps centers with performance measure data in all three years.

Today's Job Corps performance measurement system is similar to the system at the time of the NJCS. The structure is the same, and centers are evaluated along many of the same dimensions. Thus, our results are likely to be relevant to the system today. However, a handful of details are different. First, the current system rates centers along 14 different dimensions, rather than 8 or 9, as during PYs 1994-1996. Unlike in 1994-1996, current performance measures include the industry-recognized credential attainment rate, the former-enrollee initial placement rate, the graduate initial placement rate, the graduate six-month follow-up placement rate, graduate six-month average weekly earnings, and the graduate 12-month follow-up placement rate, as well as a measure that combines attainment of a high school diploma or GED and completion of a career technical training program. The current system excludes the ARPA rating used in PYs 1994 and 1995.

Second, as some of these measures illustrate, the current system includes additional measures of postplacement outcomes: measures of employment and earnings at 12 months after placement. Finally, the current system uses model-based standards—rather than national standards—for a larger share of performance measures. For PY 2010, several different performance measures used model-based goals: the high school diploma/GED attainment rate, the combined high school diploma/GED/career technical training attainment rate, the average literacy gain, the average numeracy gain, the graduate average wage at placement, and the graduate six-month average weekly earnings.[3]

B. Baseline Data from the National Job Corps Study and Program Intake Forms

To regression-adjust the center-level performance measures, we used data on the baseline characteristics of NJCS sample members. These data come from two sources: the ETA-652 program application form and the NJCS baseline survey. The former contains baseline data collected on all program participants at intake; the latter is more extensive, covering a wider range of characteristics. The ETA-652 data contain only limited participant information, and some fields have significant numbers of missing values. The NJCS baseline survey, which had a nearly 95 percent response rate, was conducted soon after random assignment and is more detailed and

[3] In addition to these broad changes, there have also been changes to the ways specific measures are calculated, including changes to the reference population definition, to the unit of measurement (e.g., grade-level equivalents instead of the educational functioning level scale), to the time frame of measurement, and to the weights of different components in calculating the overall rating.

complete than the ETA-652 (Table II.2). By using both sets of characteristics, we were able to see what regression adjustment achieves under two different scenarios. Using the NJCS baseline data, we were able to see what regression adjustment could achieve if a wide range of characteristics were measured, and measured well, at baseline (i.e., the best-case scenario). Using the ETA-652 data, we were able to assess the influence of regression adjustment under current conditions (i.e., the likely scenario).

The NJCS baseline data contain detailed information on family background characteristics, work and criminal histories, and individual demographic characteristics. Table II.2 lists the NJCS baseline measures that were used in this study; of the NJCS baseline measures, we selected an array of characteristics that were most correlated with our key outcomes. There are 14,653 NJCS study participants with NJCS baseline data.

Table II.2. Measures of Baseline Characteristics from the National Job Corps Study

Measure
Demographic Characteristics. Race, gender, age, native language*, geographic region, local area population density*.
Education and Skills. High school degree*, GED*, vocational degree*, highest grade completed, months in school in past year*.
Employment History. Ever worked*, job in past year*, currently working, months worked in past year*, occupational category of most recent job*, earnings in past year*, physical or emotional problem that limited work*.
Family Status. Marital status*, has child, pregnant*.
Socioeconomic Status. Receipt of welfare in childhood*, receipt of AFDC in past year, receipt of food stamps in past year, currently in public housing*.
Criminal History. Ever arrested*.
Drug Use. Frequency of marijuana use in past year*, frequency of use of hard drugs in past year*, ever in drug treatment*.

* Indicates measures that are not included in the ETA-652 program intake form.

The ETA-652 program intake form includes some of the same information as the NJCS baseline survey, but covers a narrower range of characteristics and has less detail. Table II.3 lists the ETA-652 baseline measures that were used in our analysis. Though ETA-652 data are available for a larger sample of NJCS participants (including some who did not participate in the baseline survey), we restricted our analysis to participants with full NJCS baseline data.

Table II.3. Measures of Baseline Characteristics from the ETA-652 Form

Measure
Demographic Characteristics. Race, gender, age, city size, prior military service, legal US resident.
Education and Skills. Months out of school, highest grade completed.
Employment History. Weeks since employed full-time, earnings per hour, annual income.
Family Status. Family status (head, related, etc.), number of dependents, needs child care.
Socioeconomic Status. Family in receipt of public assistance.
Criminal History. Convicted or adjudged delinquent.
Health and Health Care. Serious illness, under doctor's care, being treated, health insurance coverage.

C. Local Area Characteristics

Because local area characteristics may influence outcomes, we also accounted for differences in local area characteristics across students served by different Job Corps centers. To identify these characteristics for each NJCS participant, we matched pre-program participant zip codes to area characteristics from the 2008 Area Resource File (ARF), a compilation of data from numerous data sources.[4,5] The ARF includes health care data as well as economic and demographic characteristics. This study used the ARF local area characteristics listed in Table II.4.

We linked participants to local area characteristics using their zip code, which they reported at intake on the ETA-652 form. Although participant zip code at program exit may have a greater impact on post-program performance measures, the NJCS found that many participants return to their home communities after leaving Job Corps; therefore, zip code at application is likely to be a good reflection of zip code at program exit.[6] Of the 14,653 NJCS participants with baseline data, we were able to identify local area characteristics for 14,542 participants.

[4] The ARF was created by the Health Resources and Services Administration (a part of the U.S. Department of Health and Human Services) as a "National County-Level Health Resource Information Database." U.S. Department of Health and Human Services. October 5, 2010. [http://arf.hrsa.gov/index.htm].

[5] The ARF provides local area characteristics for each Federal Information Processing Standards (FIPS) county. Zip codes can be matched to county FIPS codes, which can then be matched to ARF data. For zip codes that span multiple counties, the data from each county are weighted by the percentage of addresses from the given zip code that fall in the county.

[6] We do not have measures of zip code at program exit, though the NJCS follow-up data do contain zip code at the 48-month interview. However, even if we did have data on zip code at exit, we would not want to use it to match participants to local area characteristics, because zip code at program exit could be influenced by the Job Corps program, and thus, could be thought of as a program outcome.

Table II.4. Measures of Local Area Characteristics from the Area Resource File

Measure of Population	Year	Source
Demographic Characteristics		
Percentage white	1990	1990 Census STF1A
Percentage black	1990	1990 Census STF1A
Average household size	1990	1990 Census STF1A
Percentage urban	1990	1990 Census STF3A
Percentage of families with a female head	1990	1990 Census STF1A
Percentage foreign-born	1990	County and City Data Book, 1994
Total births	1995	1995 Census Estimates of Population
Percentage of births to teens <18 years	1998-2000 1995	1998-2000 NCHS Natality Tape, 1995 Census Estimates of Population
Crime		
Deaths by homicide and legal intervention (rate)	1998-2000 1995	1998-2000 NCHS Mortality Tape, 1995 Census State and County Population Estimates Components of Change
Percentage in juvenile institutions	1990	1990 Census STF1A
Economic Characteristics		
Percentage of families in poverty	1989	1990 Census STF3A
Median household income	1995	Census Small Area Income and Poverty Estimates
Percentage of households in different income categories (<$5,000, $5,000-$9,999, $10,000-$14,999, $15,000-$24,999, $25,000-$49,999, $50,000-$99,999, >$100,000)	1989	1990 Census STF3A
Unemployment rate, 16+	1995	Bureau of Labor Statistics

Notes: STF1A = Census of Population and Household Summary Tape File 1A; STF3A = Census of Population and Housing Summary Tape File 3A; Census Estimates of Population = Population of Counties and Demographic Components of Population Change Time Series, U.S. Bureau of the Census; NCHS = National Center for Health Statistics (Centers for Disease Control and Prevention). Percentage of births to teens <18 years, deaths by homicide and legal intervention (rate), percentage of population in juvenile institutions, and average household size are calculated using multiple statistics.

D. Job Corps Center Characteristics

Our analysis also used data on Job Corps center characteristics, including type of operator (private or Federal agency), size (small, medium, or large), and region. For our main analyses, we did not adjust center performance measures for center characteristics; instead, in our subgroup analysis, we estimated differences in the results across different types of centers. However, as discussed further below, we also tested the robustness of our results to an alternative specification that controlled for center characteristics in adjusting performance measures.

E. Intake Counselors' Predictions of Center Assignment

Our analysis required data on predictions by intake counselors regarding likely center assignments of NJCS participants. We used these predictions to match NJCS study participants to

Job Corps centers, both to measure impacts at the center level and to measure participant characteristics. These predictions were required because nearly all control group members and about 27 percent of treatment group members did not attend Job Corps centers; without counselors' predictions of the center of assignment, we would not be able to identify treatment non-participants and control group members for each center. The predictions were collected at intake (which was prior to random assignment) and are thus available for both treatments and controls. They were found to be about 95 percent accurate (as determined by comparing predicted to actual center assignments for the treatment group). Of the 14,653 NJCS participants with full baseline data, we have intake counselor predictions of center assignment for 13,596. Of these, 13,454 were predicted to attend the 102 centers for which we have performance data for all three program years.

F. Follow-up Data from the National Job Corps Study

Center-level impacts were calculated using outcome data from the NJCS 12-, 30-, and 48-month follow-up interviews. Table II.5 shows the outcome measures used in this study.

The education-related and arrest outcomes pertain to the four years after random assignment. To match the presentation of earnings impact findings in key NJCS reports and journal articles, the earnings outcomes for this study pertain to calendar years 1997 and 1998 (roughly three and four years after random assignment, respectively). All outcome measures pertain to the full sample, except the GED attainment rate, which pertains to the 80 percent of the sample who did not have a high school credential at baseline.

Table II.5. Outcome Measures from the National Job Corps Study

Measures	Time Frame
Educational Services. Percentage of youth who participated in education and training. Total hours of participation in education and training.	During 48 months after random assignment
Educational Attainment. Percentage of youth who received a GED among those without a high school credential at baseline. Percentage of youth who received a vocational certificate.	During 48 months after random assignment
Arrests. Percentage of youth who were ever arrested.	During 48 months after random assignment
Earnings. Annual earnings based on the survey data.	1997, 1998

These outcome measures conceptually align with many of the Job Corps performance measures discussed above. For example, the performance measures pertaining to the in-program receipt of a GED or vocational certificate may be correlated with the outcomes on the total hours that treatment group members participated in education and training programs during the 48-month follow-up period. Similarly, the performance measures on short-term post-program employment experiences may be somewhat aligned to longer-term earnings and employment outcomes that are used in the impact analysis. This alignment suggests that higher-performing centers could have larger impacts than lower-performing centers after controlling for differences in the types of students centers serve.

Importantly, however, there are important differences between the performance and outcome measures that may weaken their association. These include (1) the mode of data collection; (2) the pools used to construct the measures (for example, the Job Corps vocational completion rate is measured using only those who remained on a Job Corps center for at least 60 days, whereas the

NJCS vocational completion rate pertains to the full sample); and (3) the fact that Job Corps performance measures include all center enrollees, compared to a random sample of enrollees for the NJCS outcome measures. These issues are discussed in more detail in Chapter IV.

It is important to note that the NJCS also calculated impacts using administrative earnings data from Social Security earnings records (SER). Because the Social Security Administration (SSA) did not release to Mathematica individual-level SER data for study sample members, this study used only survey data.[7] However, in general, the NJCS found that the pattern of earnings impacts was similar using the SSA and survey data, even though earnings levels were somewhat higher using the survey data. As described in detail elsewhere, it is unclear which data source provides a more accurate assessment of sample members' earnings (Schochet, Burghardt, and McConnell 2008). Note, however, that our findings here apply only to the survey data; it is impossible to know whether these findings would also apply to the SSA data.

The follow-up analysis sample includes NJCS participants who completed the 48-month interview, a total of 11,313 NJCS participants (6,288 treatments and 4,485 controls). About 81 percent of the treatment sample and 78 percent of the control sample responded to the 48-month interview. Of the follow-up participants, 10,409 were predicted to attend the 102 centers in our performance measure data. However, in two centers, there are ten or fewer NJCS participants in the followup sample. Impacts estimated using such small samples are likely to be very noisy estimates of the true program impact; because of this concern, we focused our analysis on the 100 centers (and the associated NJCS participants) with more than ten predicted participants in the follow-up sample. However, our results are robust to the inclusion of the two very small centers.

[7] In the previous study, SSA ran provided computer programs and provided the output. This process was not possible here, because agreements to access these data have expired, and reestablishing them would not have been feasible within the time constraints of the study.

III. STUDY METHODS

In Section III.A, we describe our initial descriptive analyses and methods for developing adjusted performance measures. In Section III.B, we describe our analytic approach to assessing the relationship between center impacts and adjusted and unadjusted performance measures. We link the discussion to the steps shown in Figure I.1.

A. Developing Adjusted Performance Measures

After gathering data on performance measures [**Step 1 of Figure I.1**], we linked data on the characteristics of participants to Job Corps centers using predicted center assignments. We then calculated center-level averages of participant characteristics from the NJCS baseline data and the ETA-652 program intake form, and local area characteristics from the ARF (linked to participants by their zip codes at intake) [**Step 2 of Figure I.1**]. We used these center-level averages in regression-adjusting performance measures to capture differences in the baseline characteristics of participants who attended different centers.

We calculated center-level averages using all baseline sample members (i.e., treatment and control group members); relative to alternative samples that could be used, this is the largest possible sample.[8] These larger sample sizes were important to reduce the noise in the center-level averages due to small samples.[9]

When calculating center-level average characteristics, we transformed categorical variables into indicators or groups of indicators. For each characteristic, we calculated the center average among predicted center participants with nonmissing data for that characteristic. For items with large numbers of missing values, we constructed center-level variables signifying the proportion of sample members with missing values.

Center-level averages were weighted by the NJCS baseline weight, which accounts for differences in sampling and survey response probabilities.[10] Center-level averages for both the NJCS data and the ETA-652 data were calculated among the 13,143 center respondents with complete NJCS baseline data who were predicted to attend the 100 centers in our primary sample. We calculated center-level averages of local area characteristics among the 13,039 baseline participants for whom we have local area characteristics data from the ARF.

After generating estimates of participant characteristics for each center, we linked these characteristics to the Job Corps performance measurement data. We obtained regression-adjusted

[8] This is one of several possible samples that could be used to calculate center-level averages. In addition to the larger sample size, another reason we chose to calculate center-level averages using the baseline sample (rather than the follow-up sample) is that this sample better approximates the baseline sample that would be available for the regression adjustment process in the absence of the NJCS. However, our results are robust to using the (smaller) follow-up sample.

[9] There are 102 Job Corps centers with performance measure data in all three program years; at baseline, 46 centers had fewer than 100 treatment and control group center designees, 42 centers had between 100 and 200, and 14 centers had more than 200.

[10] For a description of how the NJCS baseline weights were constructed, see the NJCS methodological appendix (Schochet 2001).

performance measures by regressing center-level performance measures on center-level average participant characteristics [**Step 3 of Figure I.1**]. The basic estimation equation was of the form:

(1) $PM_c = X_c\beta + \varepsilon_c,$

where PM_c is the center's performance measure, X_c is a row vector of center-level baseline participant characteristics (including the intercept), and β is the parameter vector to be estimated. The mean-zero error, ε_c, is the component of the center's performance level that cannot be explained by X_c and is assumed to be uncorrelated with X_c (i.e., that there is no omitted variable bias).

Equation (1) was estimated using ordinary least squares (OLS), where each center was weighted equally. The estimated residual,

(2) $\hat{\varepsilon}_c = PM_c - X_c\hat{\beta}$

is the regression-adjusted performance measure, where $\hat{\beta}$ is the estimated parameter vector. Under the OLS assumptions, this residual represents the part of the center's performance level that is not due to the types of students served.

We used various specifications for PM_c and X_c in equation (1). For PM_c, we used the center's overall performance rating, as well as components of that rating. In addition to using a two- or three-year average of performance measures, we conducted the analysis separately for each PY between 1994 and 1996. As discussed in Section II.B, we also estimated separate models using X_c variables from the ETA-652 forms and the more comprehensive array of participant characteristics available through the NJCS baseline survey. Because the NJCS baseline data are so extensive—with information on more characteristics than we have Job Corps centers—we relied on stepwise regression procedures (using a 0.20 p-value criterion for variable inclusion) to identify the set of baseline characteristics that have the most explanatory power in the model.

Finally, we also estimated regression-adjusted performance measures with and without controls for key center characteristics: size, operator, and region. Controlling for center characteristics accounts for systematic differences in performance across centers of different types. It holds center managers harmless for these differences, which means that centers with different characteristics—for instance, small versus large centers or centers run by private operators rather than Federal agencies—are evaluated according to different standards. Because DOL may want to evaluate all centers by the same standards (controlling only for the characteristics of the students they serve), our main approach was to exclude center characteristics at this stage, and to look at related subgroups in the second stage. However, as a sensitivity test, we calculated regression-adjusted performance measures that account for center characteristics.

B. Exploring the Relationships between Performance Measures and Impact Estimates at the Center Level

To assess the degree to which regression adjustment improves the ability of performance measures to distinguish between centers with high and low impacts, we looked at the association between Job Corps center performance measures (unadjusted and adjusted) and center-level impact estimates.

Center-level impacts could be measured in two different ways. First, we could estimate the intent-to-treat (ITT) effect—the effect of being offered the opportunity to join a given Job Corps center—which is the difference in weighted mean outcomes for the treatment and control groups. However, only 73 percent of treatment group members—i.e., those *offered* the opportunity to go to Job Corps—took up that offered and enrolled. We may be more interested in the impact of treatment on the treated (the TOT effect), namely an estimate of the effect of actually participating in Job Corps, rather than the effect of the offer to participate. If treatment group members who were offered program services but opted out of participating are unaffected by the program—a fairly weak assumption—a simple adjustment can be made to calculate the effect of the program on participants (the TOT effect). This "Bloom" adjustment involves the calculation of the effect of the program on participants by dividing the estimated impact (the ITT) by the participation rate, i.e., the rate of take-up of Job Corps assignments (Bloom 1984). The intuition for this result is that if the effect of the program on nonparticipants is known to be zero, the estimated impact can be attributed entirely to the proportion of the treatment group that actually participated in the program.

We generated center-level impact estimates using the predictions of intake counselors, discussed above in Section II.E. Because control group members were not actually assigned to centers, we relied on these predictions to match them to centers they would likely have attended. Center-level impacts were estimated as the difference in weighted mean outcomes between treatment and control group members, divided by the difference in the weighted participation rates [**Step 4 of Figure I.1**]. In particular, the TOT impact was estimated as:

$$(3) \quad impact_c = \frac{\bar{y}_{treatment,c} - \bar{y}_{control,c}}{\hat{p}_{treatment,c} - \hat{p}_{control,c}}$$

where $impact_c$ is the impact estimate for center c, $\bar{y}_{treatment,c}$ is the mean outcome among treatment group members predicted to attend center c, $\bar{y}_{control,c}$ is the mean outcome among control group members predicted to attend center c, $\hat{p}_{treatment,c}$ is the Job Corps participation rate among treatment group members predicted to attend center c, and $\hat{p}_{control,c}$ is the Job Corps participation rate among control group members predicted to attend center c.[11] The means and participation rates were weighted to account for sampling probabilities and survey nonresponse.

We calculated center-level impacts for all seven outcome measures shown in Table II.5 (receipt of educational services, hours of educational services, GED receipt, vocational certificate receipt, ever arrested, 1997 annual earnings, and 1998 annual earnings).

In the final step of our analysis, we compared center-level impact estimates (calculated using equation [3]) to adjusted and unadjusted performance measures [**Step 5 of Figure I.1**]. In particular, we calculated the correlations between different adjusted and unadjusted performance measures—including different components and different program years—and center-level impacts. Because some participants were enrolled in Job Corps in multiple program years, we calculated the impact

[11] Between random assignment and the 48-month follow-up interview, about 73 percent of the treatment group, and 1 percent of the control group, had enrolled in Job Corps. (Control group participants who enrolled in Job Corps after a three-year embargo period are considered nonparticipants here, since these are not cases of crossover.)

estimate for all NJCS participants predicted to attend the center, regardless of the program year. When comparing impact estimates to performance measures, we compared these pooled estimates to yearly and multiyear average performance measures.

Finally, we conducted several robustness checks, including analyses of the degree to which measurement error influenced our findings. We also present results from analytical extensions designed to explore and augment our main findings.

IV. RESULTS

A. Initial Descriptive Analysis

Before constructing regression-adjusted performance measures, we first conducted several descriptive analyses to help develop our analytic plan and to assess whether regression adjustment had the potential to change the unadjusted performance rankings.

1. Relationship between Unadjusted Performance Measures and Participant Characteristics

Understanding the relationship between unadjusted performance measures and center-level baseline participant characteristics is critical for determining (1) the extent to which centers with different ratings served systematically different participants, and hence (2) the scope for regression adjustment to influence performance measures. We categorized centers into three groups (with low, medium, and high ratings) and then tabulated, for each group, average baseline participant characteristics from the NJCS baseline data, the ETA-652 baseline data, and the ARF.

Table IV.1 shows average center characteristics for centers with low, medium, and high three-year average unadjusted overall performance ratings. Low, medium, and high are terciles based on the three-year average of the center's overall rating ratio. The characteristics presented in Table IV.1 are a subset of the characteristics used in our analysis. Characteristics were averaged first at the center level and then across centers in the performance tercile (weighting all centers equally). We used F-tests to gauge whether differences in the distribution of participant characteristics across the center performance terciles were statistically significant.

There are statistically significant differences in the baseline characteristics of participants across centers with low, medium, and high overall performance (Table IV.1). Students in high-performing centers largely had characteristics that are favorable for outcomes, with a handful of exceptions. Relative to low-performing centers, centers with high three-year average unadjusted overall performance had a smaller share of students who were black and a larger share who were white, Hispanic, or of other racial or ethnic groups. High-performing centers were also more likely to have participants with a high school degree, more likely to have health insurance or Medicaid coverage, and less likely to receive food stamps or public assistance. High-performing centers drew students from areas that had a larger share of white residents, a smaller share of female-headed families, and higher incomes. However, students at high-performing centers were more likely to have used hard drugs and less likely to be native English-speakers. Furthermore, the differences across groups are relatively small in magnitude, and there are several student characteristics on which centers do not differ, including gender, age, and arrests. The NJCS baseline and ETA-652 data exhibit similar patterns in terms of the characteristics that do and do not differ across groups.

These results suggest that there are differences in the characteristics of students attending centers with low, medium, and high ratings. However, the differences are small in magnitude and do not necessarily suggest that better-performing centers serve students that are systematically better off. Based on these findings, it is unclear how the relative rankings of different centers will change due to the regression adjustment of the performance measures to account for differences in student characteristics.

Table IV.1. Average Baseline Characteristics, by Overall Center Performance Tercile

	Unadjusted Overall Center Performance Tercile			
	Low	Medium	High	*p-value*
Selected NJCS Baseline Characteristics				
Demographic Characteristics				
Non-Hispanic White	0.245	0.344	0.391	0.023**
Non-Hispanic Black	0.558	0.479	0.249	0.000***
Hispanic	0.139	0.106	0.227	0.017**
Other race	0.058	0.071	0.133	0.040**
Female	0.388	0.361	0.399	0.663
Native language English	0.897	0.916	0.799	0.003***
Native language Spanish	0.061	0.048	0.124	0.026**
Native language other	0.042	0.035	0.077	0.066*
Age 15-17	0.443	0.447	0.427	0.708
Age 18-20	0.395	0.401	0.415	0.498
Age >20	0.162	0.153	0.158	0.817
Education and Skills				
High school degree	0.151	0.152	0.191	0.034**
GED	0.042	0.047	0.057	0.147
Vocational degree	0.025	0.024	0.029	0.465
Highest grade completed 0-8	0.167	0.161	0.125	0.008***
Highest grade completed 9-11	0.659	0.657	0.649	0.803
Highest grade completed >11	0.174	0.181	0.226	0.009***
Employment History				
Currently working	0.202	0.205	0.228	0.203
Earnings in past year <$1,000	0.507	0.483	0.497	0.590
Earnings in past year $1,000-$4,999	0.285	0.296	0.292	0.771
Earnings in past year $5,000-9,999	0.142	0.146	0.139	0.821
Earnings in past year >$10,000	0.066	0.075	0.071	0.596
Physical or emotional problem that limited work	0.054	0.050	0.058	0.517
Family Status				
Has child	0.193	0.168	0.148	0.142
Socioeconomic Status				
Did not receive food stamps over past year	0.574	0.602	0.630	0.088*
Received food stamps some of past year	0.058	0.074	0.065	0.092*
Received food stamps all of past year	0.368	0.325	0.306	0.035**
Criminal History				
Ever arrested	0.267	0.298	0.301	0.231
Drug Use				
Used no drugs over past year	0.711	0.663	0.622	0.001***
Used hard drugs occasionally over past year	0.040	0.058	0.101	0.000***
Used hard drugs frequently over past year	0.005	0.014	0.026	0.000***
Selected ETA-652 Baseline Characteristics				
Demographic Characteristics				
Male	0.612	0.639	0.601	0.663
Non-Hispanic White	0.266	0.368	0.429	0.020**
Non-Hispanic Black	0.587	0.511	0.269	0.000***
Hispanic	0.109	0.079	0.197	0.020**
American Indian	0.028	0.032	0.075	0.249
Asian	0.010	0.010	0.031	0.005***
Age 14-17	0.446	0.447	0.429	0.712
Age 18-20	0.393	0.401	0.413	0.493
Age >20	0.161	0.153	0.158	0.857
Education and Skills				
Highest grade completed 0-8	0.178	0.175	0.124	0.002***
Highest grade completed 9-11	0.638	0.634	0.643	0.873

	Unadjusted Overall Center Performance Tercile			
	Low	Medium	High	*p-value*
Highest grade completed >11	0.184	0.191	0.233	0.020**
Employment History				
Estimated annual income $0-$400	0.010	0.014	0.022	0.042**
Estimated annual income $401-$6,528	0.236	0.265	0.283	0.210
Estimated annual income >$6,529	0.256	0.266	0.245	0.677
Estimated annual income missing	0.497	0.455	0.451	0.377
Socioeconomic Status				
Receiving public assistance	0.458	0.403	0.389	0.034**
Health and Health Care				
Covered by health insurance or Medicaid	0.328	0.363	0.432	0.013**
Local Area Characteristics				
Demographic Characteristics				
Percentage white	0.708	0.740	0.790	0.014**
Percentage black	0.221	0.195	0.093	0.000***
Average household size	2.723	2.683	2.789	0.037**
Percentage urban	0.750	0.699	0.772	0.127
Percentage of families with a female head	0.199	0.192	0.171	0.020**
Percentage foreign-born	0.621	0.563	0.835	0.226
Total births	10512	7805	18741	0.037**
Percentage of births to teens <18 years	0.057	0.051	0.049	0.032**
Crime				
Deaths by homicide and legal intervention (rate)	0.000	0.000	0.000	0.012**
Percentage of population in juvenile institutions	0.000	0.000	0.001	0.023**
Economic Characteristics				
Percentage of families in poverty	0.140	0.121	0.121	0.129
Median household income	31726	33230	34064	0.049**
Percent households with income:				
<$5,000	0.089	0.076	0.066	0.000***
$5,000-$9,999	0.107	0.104	0.101	0.339
$10,000-$14,999	0.099	0.095	0.097	0.590
$15,000-$24,999	0.186	0.184	0.187	0.810
$25,000-$49,999	0.320	0.330	0.338	0.003***
$50,000-$99,999	0.166	0.174	0.176	0.593
>$100,000	0.032	0.036	0.034	0.707
Unemployment rate, 16+	0.061	0.060	0.067	0.117
Number of Centers	**33**	**33**	**34**	

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF.

Notes: All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Terciles are based on the three-year average overall rating. The reported p-value refers to an F-test which tests whether the three groups are jointly significant.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

2. Correlations between Different Performance Measures

The original NJCS analysis of performance measures focused on the overall rating, and not individual components of performance. To understand the relationship between the overall rating and the individual components, we estimated the pairwise correlations between center rankings based on different performance measure components.

We estimated the Spearman rank correlations between multiyear average center performance measures (Table IV.2). The rank correlation is a nonparametric approach to analyzing the correspondence between variables and is less sensitive to outliers. The different rankings of the components of performance measures are, in general, positively correlated (except for the full-time placement measure). In addition, the correlations tend to be higher for measures within the two main measure groups (the in-program and post-program groups) than across groups. There is, however, considerable heterogeneity in the rank correlations. For example, the overall rating is highly correlated with the ARPA rating and with educational performance measures (e.g., math gains, the GED rate, and the vocational completion rate), but less so with employment outcomes. However, differences in the correlations could partly reflect differences in the weights given in aggregating components to construct the overall rating; these weights are shown in Appendix Table A.1.

Because the pairwise correlations are not universally large, it appears that different performance measures are capturing different dimensions of center performance. Thus, our analysis uses the overall performance measure as well as each component measure.

In results not shown, we also estimated the Spearman rank correlations across different years (for the same components). The correlation in center rankings across years is generally high. For the overall rating, the correlations between the 1994, 1995, and 1996 rankings range from 0.53 to 0.73. Across all of the components, the rank correlations range from 0.31 (GED, 1994 and 1996) to 0.77 (wage, 1995 to 1996), with one very low outlier (a 0.09 correlation between the 1994 and 1996 quality placement rankings). Because the correlation across years is generally high, year-by-year results are unlikely to differ substantially from the multiyear averages.

Table IV.2. Rank Correlations between Unadjusted Center Performance Components, Multiyear Averages

	Center Ranking									
Center Ranking	Overall	Reading Gains	Math Gains	GED	Vocational Completion	Placement	Average Wage	Quality Placement	Full-Time	ARPA Rating
Overall	1.00									
Reading Gains	0.69	1.00								
Math Gains	0.76	0.84	1.00							
GED Rate	0.66	0.48	0.53	1.00						
Vocational Completion Rate	0.87	0.53	0.62	0.52	1.00					
Placement Rate	0.61	0.31	0.40	0.29	0.36	1.00				
Average Wage	0.44	0.08	0.21	0.19	0.36	0.36	1.00			
Quality Placement	0.57	0.23	0.31	0.17	0.39	0.46	0.56	1.00		
Full-Time	-0.07	-0.09	-0.09	-0.35	-0.01	-0.08	0.20	0.19	1.00	
ARPA Rating	0.81	0.60	0.63	0.54	0.60	0.53	0.19	0.44	-0.13	1.00

Sample Size = 100 centers

Source: Performance measure data.

Notes: All centers are weighted equally. Table shows the Spearman rank correlation based on a multiyear average of the center's performance ratio. For reading gains, math gains, full-time, and ARPA rating, this is a two-year average (see Table II.1 for available years). For other performance measures, this is a three-year average (PYs 1994-1996).

3. Relationship between Control Group Outcomes and Baseline Characteristics

For there to be a possibility of regression adjustment improving the correspondence between performance measures and impacts, there must be a strong relationship between control group outcomes and baseline characteristics. Intuitively, the use of baseline data to adjust the performance measures is supposed to remove the influence of, or "control" for, participant characteristics; then, if such characteristics are not predictive of control group outcomes, regression adjustment will not move performance measures closer to impacts estimated in the NJCS. Therefore, we assessed the relationship between center-level *outcomes for the control group* and center-level baseline participant characteristics.

Table IV.3 reports the R^2 and adjusted R^2 values from regressions of center-level NJCS control group outcomes on center-level baseline characteristics. The R^2 is a measure of the goodness of fit;

its value is the proportion of the variance in the outcome measure that is explained by the covariates. If the R^2 value is close to 1, the covariates explain a large share (close to 100 percent) of the variance, which would suggest that the covariates are highly predictive of the outcome. Adjusted R^2 values can be interpreted the same way; the adjustment accounts for the number of variables included in the model.

We show results for two different sets of baseline characteristics. First, we show R^2 values from a stepwise regression of center-level control group outcomes (from the NJCS follow-up data) on baseline characteristics from the NJCS and local area characteristics from the ARF. In the second row, we show R^2 values from a regression (not stepwise) of center-level control group outcomes on center-level averages of participant characteristics from the ETA-652 form and the ARF. In the third and fourth rows, we show the corresponding adjusted R^2 values.

The R^2 values are relatively large and suggest that the characteristics of participants and their local areas are predictors of these outcomes; adjusted R^2 values are somewhat lower, particularly for the ETA-652 adjustment. Together with Table IV.1, these results indicate that there may be some scope for regression adjustment to improve the correspondence between center-level impacts and center performance measures.

Table IV.3. Regression R^2 Values from Regressions of Center-Level Control Group Outcomes on Center-Level Baseline Characteristics

	Dependent Variable: Center Level Control Group Outcome						
Independent Variables	Any Education and Training	Hours of Education and Training	GED Receipt	Vocational Certificate Receipt	Ever Arrested	1997 Annual Earnings	1998 Annual Earnings
Regression R^2							
NJCS Baseline Characteristics and Local Area Characteristics	0.56	0.70	0.69	0.85	0.62	0.72	0.55
ETA-652 Baseline Characteristics and Local Area Characteristics	0.68	0.72	0.64	0.71	0.71	0.58	0.58
Regression Adjusted R^2							
NJCS Baseline Characteristics and Local Area Characteristics	0.48	0.62	0.59	0.77	0.57	0.62	0.46
ETA-652 Baseline Characteristics and Local Area Characteristics	0.31	0.41	0.23	0.37	0.38	0.09	0.09

Sample Size = 100 centers.

Sources: NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: Table reports R^2 values from regressions of center-level average outcomes for the NJCS control group on center-level average baseline characteristics. The regression that controls for NJCS baseline characteristics is a forward selection stepwise regression with inclusion and exclusion p-value thresholds of 0.20. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight; center-level control group outcomes are weighted using the follow-up weight.

4. Distributions of Performance Measures and Estimated Impacts across Centers

An important issue for assessing the extent to which the regression adjustment of the performance measures can make a difference is to examine the variation in the performance measures and the estimated impacts across centers. For instance, all else equal, we might expect regression adjustment to have more of an effect on the ranking of performance measures if the unadjusted performance measures are close together. If the unadjusted performance measures are spread out, adjustment may be less likely to alter the rankings. Similarly, we might expect to find stronger associations between impacts and center performance if the estimated impacts vary somewhat across centers than if they do not.

We find that the multiyear average performance measures do not vary substantially across the 100 study centers (Table IV.4). Median values range from 1 to 1.26, and the measures range from 0.56 to 2.19. There are, however, some differences across measures. For instance, the overall, full-time placement, and ARPA rating measures vary less than the other measures, and the reading gains, math gains, and GED rate measures vary the most. These results suggest that small changes in the performance measures due to regression adjustment could have a nontrivial effect on the center performance rankings.

Table IV.4. Summary Statistics for Unadjusted Performance Measures and Center-Level Impact Estimates

	Min	1st Quartile	Mean	Median	3rd Quartile	Max	Standard Deviation
Performance Measures (Multiyear Averages)							
Overall	0.87	1.04	1.09	1.10	1.14	1.34	0.09
Reading Gains	0.56	0.96	1.15	1.13	1.33	1.85	0.26
Math Gains	0.59	1.06	1.20	1.20	1.35	1.87	0.24
GED Rate	0.62	0.92	1.06	1.04	1.14	2.19	0.25
Vocational Completion Rate	0.67	0.98	1.09	1.10	1.20	1.48	0.16
Placement Rate	0.90	1.04	1.10	1.11	1.16	1.23	0.08
Average Wage	0.89	0.99	1.03	1.02	1.06	1.16	0.06
Quality Placement	0.95	1.18	1.26	1.26	1.36	1.59	0.13
Full-Time	0.93	1.07	1.10	1.09	1.13	1.21	0.05
ARPA Rating	0.80	0.95	0.99	1.00	1.03	1.10	0.07
Center-Level Impact Estimates							
Any Educational Services[a]	-4.6	20.3	30.0	29.0	39.2	73.1	15.7
Hours Educational Services	-329	617	945	965	1,232	1,849	463
GED Receipt[a]	-23.2	9.1	21.3	20.2	32.5	115.8	19.5
Vocational Certificate Receipt[a]	-20.3	23.0	31.6	32.1	41.8	69.7	15.4
Arrested[a]	-60.2	-17.2	-6.5	-5.2	3.8	38.9	17.5
1997 Annual Earnings[b]	-8,274	-1,130	494	460	2,607	7,205	3128
1998 Annual Earnings[b]	-8,566	-601	1,415	1,307	3,774	10,908	3448

Sample Size = 100 centers.

Sources: Performance measure data, NJCS follow-up surveys.

[a] Impacts are measured in percentage points.

[b] Impacts are measured in 1995 dollars.

The center-level impact estimates vary considerably more than do the performance measures (Table IV.4). For instance, impacts on 1998 annual earnings range from -$8,566 to $10,908, and impacts on the receipt of a vocational certificate range from -20.3 to 69.7 percentage points. These results could be partly due to relatively small sample sizes in some centers. Thus, in many of our analyses below, we group centers into performance terciles to help reduce this potential measurement error problem.

In Appendix Table A.2, we show summary statistics for performance measures and impact estimates using the 40 centers for which there are 100 or more observations at follow-up. As expected, for these centers, the standard deviations of the impact estimates are smaller than for the full sample; the minimum and maximum values also suggest a tighter distribution. In Appendix Table A.3, we show means and standard deviations of each performance measure by year (for the full sample of centers), and find that these statistics are relatively stable across years. Compared to the multiyear averages, however, annual measures tend to have slightly larger standard deviations.

B. Comparing Adjusted and Unadjusted Performance Measures

This section presents evidence that although the regression adjustment process changes somewhat the center performance measures and rankings, the unadjusted and adjusted measures are nonetheless highly correlated.

1. Regression Adjusting Performance Measures

To calculate adjusted performance measures, we regressed unadjusted performance measures (for all years and all components) on baseline characteristics from the NJCS baseline survey and ETA-652 program intake form following equation (1). It is the residual from this regression that is the adjusted performance measure. In regressions based on the NJCS data, we used stepwise procedures to select independent variables with explanatory power. As described in Chapter III, we calculated center-level averages of baseline characteristics among both treatment and control group members predicted to be assigned to each center using the NJCS baseline sample.[12]

Table IV.5 shows R^2 values from a representative set of the estimation regressions (Appendix Tables A.4 and A.5 show parameter estimates and which variables were selected by the stepwise procedure). We show results for five different performance measures (all three-year averages): (1) overall, (2) the GED rate, (3) the vocational completion rate, (4) the average wage, and (5) the placement rate. For each measure, we show R^2 results for four different sets of independent variables: (1) the NJCS baseline variables and local area characteristics from the ARF; (2) the NJCS baseline variables, local area characteristics, and center characteristics; (3) the ETA-652 baseline variables and local area characteristics; and (4) the ETA-652 baseline variables, local area characteristics, and center characteristics.

[12] Although our impact estimates are calculated using the follow-up sample, using the baseline sample at this stage has several advantages. First, because the baseline sample is larger, it is likely to provide less noisy estimates of center-level average participant characteristics. Second, the baseline sample is likely to mimic the sample that would be available to the Job Corps performance measurement system if regression adjustment were adopted. Because the Job Corps performance measurement system would not be able to distinguish between potential Job Corps participants who would and would not be reached for follow up, the baseline sample is likely to better approximate the sample available to the performance measurement system.

The results indicate that the center-level baseline characteristics have significant explanatory power in the models, although the R^2 values vary somewhat across performance measures and specifications (Table IV.5). Depending on the model and performance measure, between 41 percent and 89 percent of the variance is explained by the covariates.[13] This suggests that the regression adjustment process might influence center performance rankings.

An unexpected finding is that the R^2 values tend to be slightly larger using the ETA-652 data than using the NJCS data. This could be due to the stepwise regression procedure that was used to select the model covariates in the specifications that used the NJCS survey data; this process was not used for specifications that used the ETA-652 data (where all variables were included in the models). As shown in Appendix Tables A.4 and A.5, more ETA-652 variables are included in the models than NJCS baseline variables. *Adjusted* R^2 values (that account for the number of variables included in the models) are larger using the NJCS variables (not shown). It is striking, however, that the two data sources have similar predictive power.

Table IV.5. Regression R^2 Values from Regressions of Three-Year Average Unadjusted Center Performance Ratings on Center-Level Baseline Characteristics

	Regression R^2				
	Dependent Variable: Unadjusted Performance Rating				
Independent Variables	Overall	GED	Vocational Completion	Average Wage	Placement Rate
NJCS Baseline Characteristics and Local Area Characteristics	0.66	0.76	0.41	0.79	0.70
NJCS Baseline Characteristics, Local Area Characteristics, and Center Characteristics	0.58	0.65	0.59	0.79	0.76
ETA-652 Baseline Characteristics and Local Area Characteristics	0.72	0.80	0.58	0.84	0.76
ETA-652 Baseline Characteristics, Local Area Characteristics, and Center Characteristics	0.81	0.86	0.71	0.89	0.81

Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, center characteristics, 2008 ARF.

Notes: All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Table reports R^2 values from regressions of three-year average unadjusted performance measures on center-level average baseline characteristics. The regressions that control for NJCS baseline characteristics are forward-selection stepwise regressions with inclusion and exclusion p-value thresholds of 0.20.

[13] We had also initially planned to assess goodness of fit by estimating the regression model for a randomly selected subsample of the overall sample, predicting performance using the holdout sample, and then comparing predicted performance to actual performance. However, restricting the sample size limits the number of covariates that can be included in the regression model. Because we cannot estimate the same model for a 50 percent subsample, the relevance of this analysis is limited.

2. Correspondence between Unadjusted and Adjusted Performance Measures

To get a sense for how regression-adjusted performance measures relate to the unadjusted measures, we identified low, medium, and high performers according to each measure and used these groupings to construct three-by-three contingency tables. We then examined cell counts in the diagonal and off-diagonal entries.

Regression adjustment has some influence on center performance rankings. Using the three-year average overall rating, nearly half of all centers are classified into new terciles after adjustment (Table IV.6). Using the NJCS adjustment, 52 centers have adjusted and unadjusted three-year average overall ratings that are in the same tercile, and 48 have ratings in different terciles. Using the ETA-652 adjustment, 50 centers have adjusted and unadjusted three-year average overall ratings in the same tercile, and 50 have ratings in different terciles. (The correspondence between unadjusted and adjusted terciles is similar when adjustment also includes center characteristics, as shown in Appendix Table A.6.)

Table IV.6. Unadjusted Center Performance Tercile and Adjusted Center Performance Tercile, Overall Three-Year Average Rating

	NJCS-Adjusted Performance Tercile			ETA-652-Adjusted Performance Tercile		
Unadjusted Performance Tercile	Low	Medium	High	Low	Medium	High
Low	19	10	4	22	8	3
Medium	10	13	10	6	12	15
High	4	10	20	5	13	16
Number of Centers	33	33	34	33	33	34

Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF.

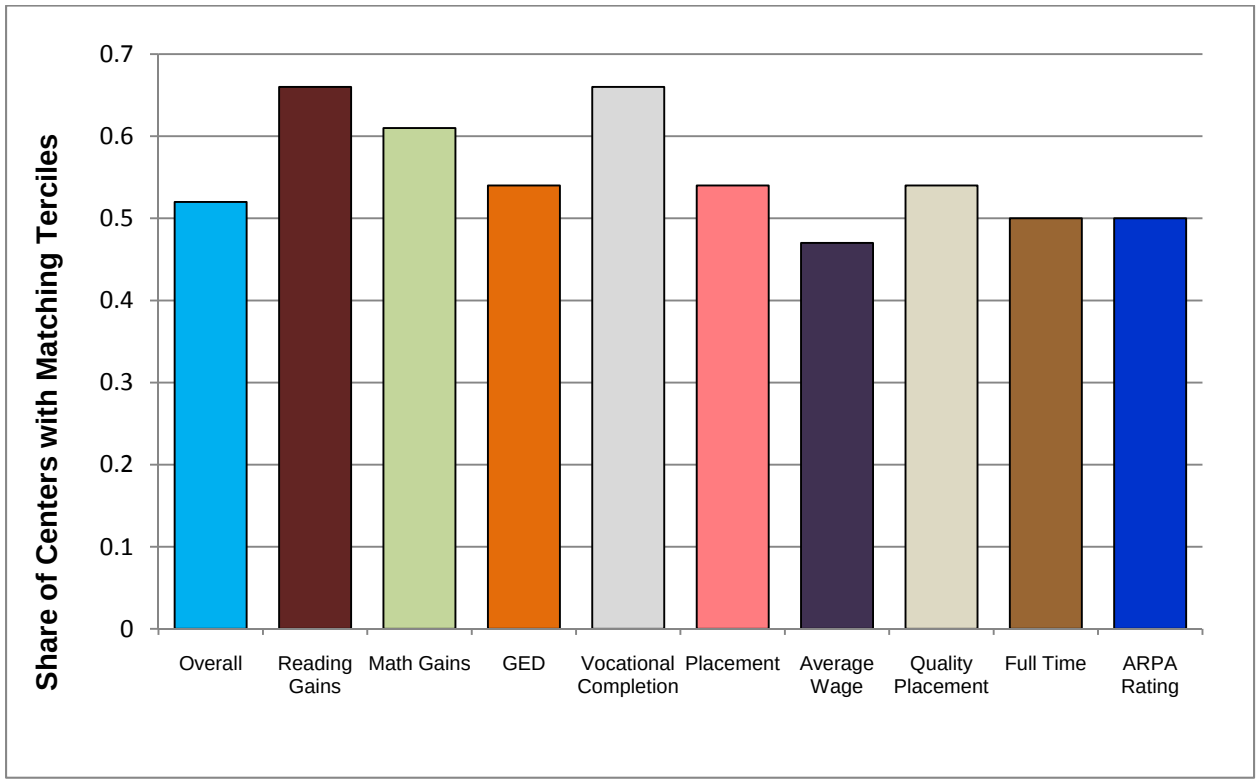
Notes: Table shows terciles of the three-year average overall performance rating. NJCS-adjusted and ETA-652-adjusted performance terciles are terciles based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics.

The share of centers with matching terciles is similar for other performance measures as well (Figure IV.1). For all multiyear average center performance measures, the share of centers that have unadjusted and NJCS-adjusted performance measures in the same tercile (i.e., those that would have been along the diagonal in Table IV.6) is between 0.47 and 0.66. The adjustment changes the performance tercile for more than one third of all centers. Unadjusted and NJCS-adjusted performance measure terciles are more often different for the average wage rating, the ARPA rating, and the full-time rating, and are more often the same for the reading gains rating, the math gains rating, and the vocational completion rating.

In addition to comparing terciles of the performance measure distributions, we calculated the correlations between unadjusted and adjusted performance measures. Table IV.7 shows those correlations for each performance measure and each year (PY 1994, PY 1995, PY 1996, or the two- or three-year average). We show results using the NJCS adjustment as well as the ETA-652 adjustment.

The correlations between adjusted and unadjusted ratings are relatively high; all correlations are statistically significant (Table IV.7). The correlation ranges from 0.39 to 0.82, depending on the measure, year, and adjustment. Surprisingly, correlations are generally higher using the NJCS adjustment than the ETA-652 adjustment. However, as discussed, this difference could reflect the use of the stepwise procedure for the NJCS adjustment, rather than the data source.

Figure IV.1. Correspondence between Terciles of Unadjusted and NJCS-Adjusted Center Performance Ratings, Multiyear Averages



Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, 2008 ARF.

Notes: Figure shows the share of centers with unadjusted and NJCS-adjusted multiyear average performance ratings in the same tercile. NJCS-adjusted performance terciles are terciles based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics.

Table IV.7. Correlations between Unadjusted Center Performance and Adjusted Center Performance, All Components in All Years

	Correlation Between Unadjusted and Adjusted Performance Measures							
	NJCS-Adjusted				ETA-652-Adjusted			
Performance Measure	PY94	PY95	PY96	Multiyear Average	PY94	PY95	PY96	Multiyear Average
Overall	0.54	0.69	0.65	0.58	0.48	0.58	0.59	0.53
Reading Gains	0.54	0.74	--	0.74	0.61	0.67	--	0.64
Math Gains	0.66	0.82	--	0.75	0.55	0.64	--	0.56
GED Rate	0.53	0.64	0.64	0.49	0.39	0.54	0.57	0.45
Vocational Completion Rate	0.63	0.71	0.71	0.77	0.63	0.65	0.67	0.65
Placement Rate	0.55	0.56	0.51	0.55	0.48	0.51	0.52	0.49
Average Wage	0.59	0.40	0.48	0.46	0.45	0.41	0.48	0.40
Quality Placement	0.74	0.66	0.65	0.63	0.55	0.53	0.54	0.49
Full-Time	--	0.55	0.65	0.56	--	0.50	0.54	0.52
ARPA Rating	0.72	0.62	--	0.66	0.60	0.62	--	0.58

Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF.

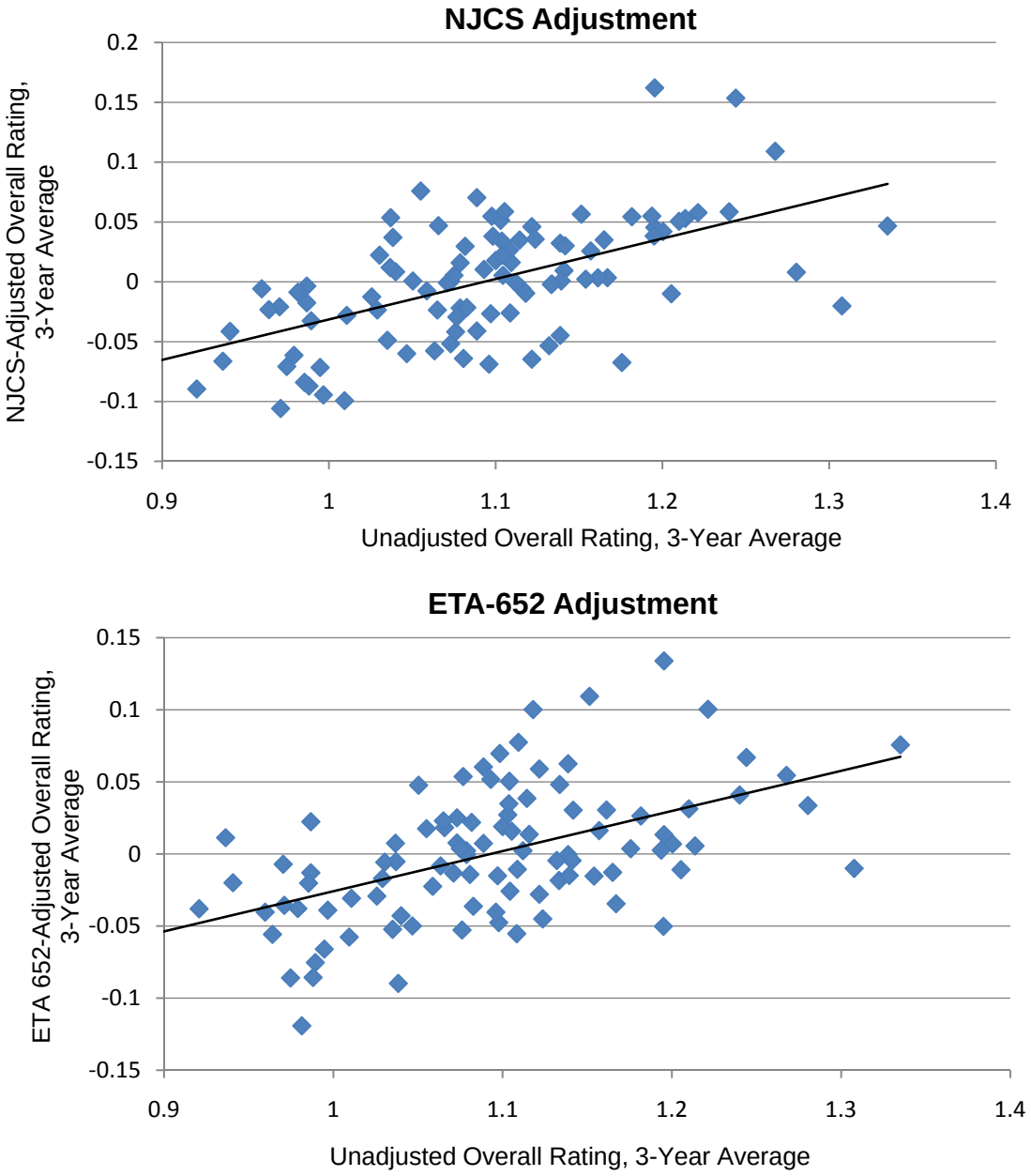
Notes: All correlations are statistically significant at the 1 percent level. NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight.

Finally, in Figures IV.2, IV.3, and IV.4, we present scatter plots showing the relationship between unadjusted and adjusted performance measures. In each plot, we also show a fitted regression line through these points. In Figure IV.2, we show the three-year average overall rating when adjusted by NJCS baseline characteristics and local area characteristics (top panel) and when adjusted using ETA-652 data and local area characteristics (bottom panel). Appendix Figure A.1 shows the analogous comparison with adjustment for center characteristics; the overall relationship between unadjusted and adjusted performance measures is quite similar whether or not center characteristics are included.

In Figure IV.3, we show comparisons of the unadjusted and adjusted performance measures for the GED rate, vocational completion rate, average wage, and quality placement ratings. In all cases, we use multiyear averages and adjust by NJCS baseline characteristics and local area characteristics. In Figure IV.4, we show the correspondence between the unadjusted and adjusted overall performance rating separately by year, with adjustment by NJCS baseline characteristics and local area characteristics.

Consistent with Table IV.7, the figures show that the unadjusted and adjusted performance measures are positively correlated. However, there is some scatter around these regression lines, which indicates that performance measures were, to some degree, changed by adjustment.

Figure IV.2. Unadjusted and Adjusted Center Performance, Three-Year Average Overall Rating

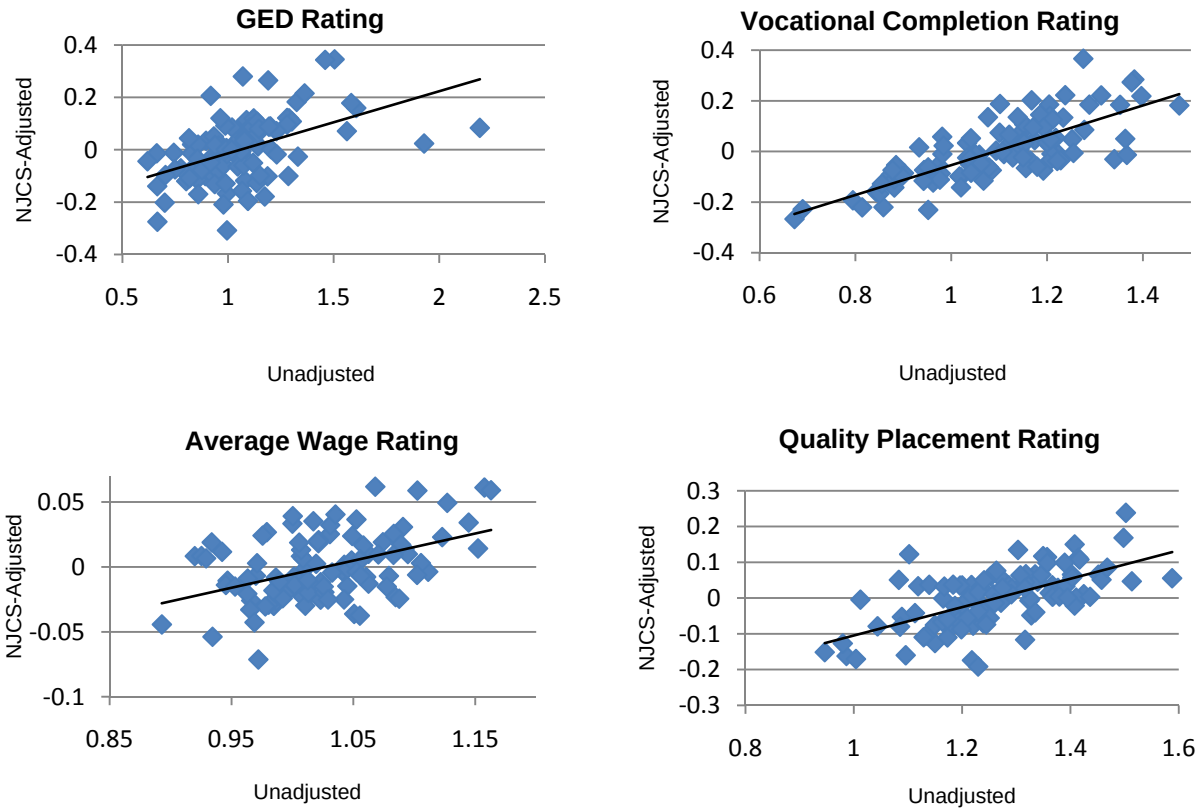


Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. In both graphs, the slopes are statistically significant at the 1 percent level.

Figure IV.3. Unadjusted and NJCS-Adjusted Center Performance, Three-Year Average Component Ratings

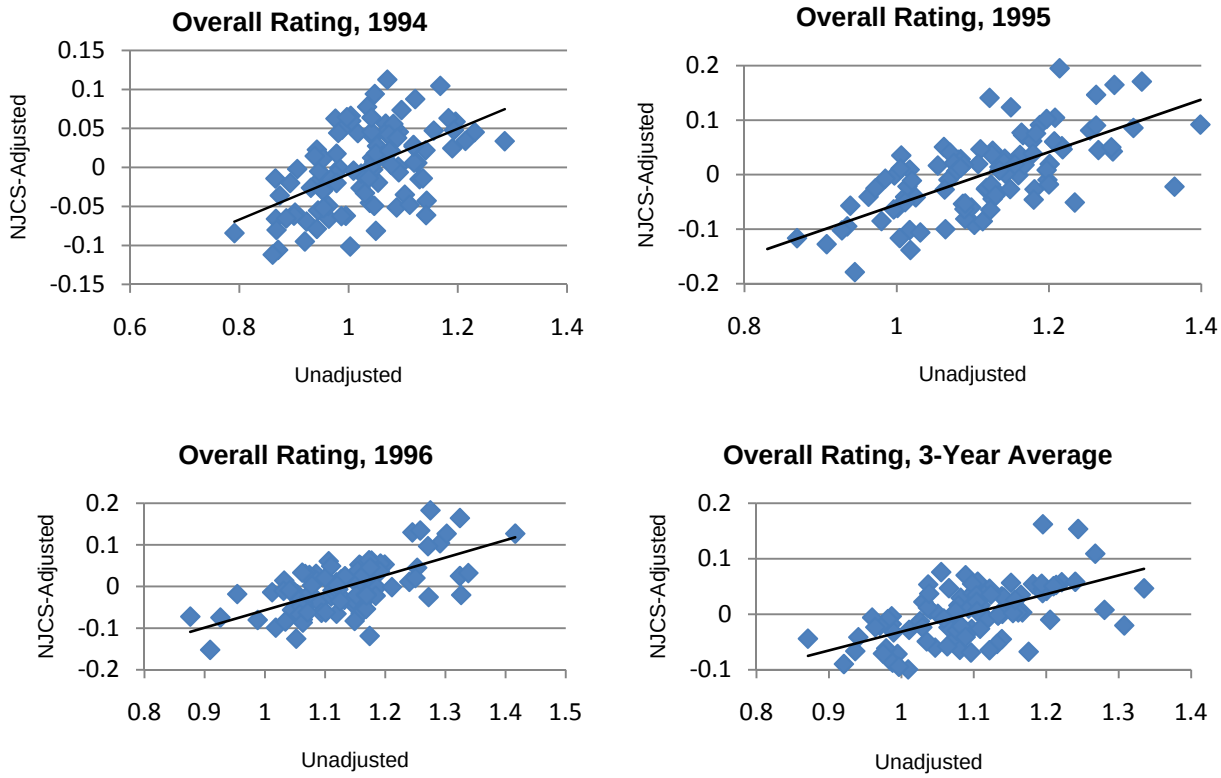


Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, 2008 ARF.

Notes: All adjusted performance measures use the NJCS adjustment. NJCS-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. In all four graphs, the slopes are statistically significant at the 1 percent level.

Figure IV.4. Unadjusted and NJCS-Adjusted Center Performance, Overall Rating by Year



Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, 2008 ARF.

Notes: All adjusted performance measures use the NJCS Adjustment. NJCS-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. In all four graphs, the slopes are statistically significant at the 1 percent level.

Finally, we also analyzed whether very high-performing centers—based on NJCS-adjusted performance measures—were the same across different performance measure components. There are 26 centers with at least one top-five ranking; however, there are only 4 centers with three or more top-five rankings. When we calculated top-ten rankings for each component, we found that there are 47 centers with at least one top-ten ranking, but only 11 centers with three or more top-ten rankings. Therefore, there is not much consistency in center rankings across components using the NJCS-adjusted performance measures.

C. Comparing Performance Measures and Impact Estimates

In the last section, we presented evidence that the regression adjustment process had some effect on center-level performance scores. As discussed in this section, however, the regression adjustment process did not improve the association between the center-level performance measures and impact estimates.

1. Scatter Plots of Performance Measures and Impacts

To measure the effect of regression adjustment, we compared the center-level impact estimates ($impact_c$) to adjusted and unadjusted performance measures ($\hat{\varepsilon}_c$ and PM_c) using descriptive analyses and regression models. First, we plotted the performance measures against the impact estimates for each center and plotted a regression line through these points. In addition to providing an overall picture of this relationship, these scatter plots illustrate whether there are differences in the relationship across different regions of the performance distribution and whether outliers are driving the results (they are not).

Figures IV.5 through IV.9 show the relationship between center performance measures (adjusted and unadjusted) and center-level program impacts. Each figure shows the unadjusted, the NJCS-adjusted, and the ETA-652-adjusted performance measure in relation to center level impacts on a selected outcome.

Using the three-year average, we find the correlation between centers' overall ratings and their 1998 earnings impacts to be negative (but not significant) across all three performance measures: unadjusted, NJCS-adjusted, and ETA-652-adjusted (Figure IV.5). Adjustment does not affect the relationship between overall (three-year average) performance and impacts on 1998 earnings. It does seem to have a small effect on the relationship between the overall performance rating and impacts on arrests (Figure IV.6)—the relationship becomes more negative with adjustment. However, the correlation is still quite weak.

We also analyzed the relationship between performance measure components (adjusted and unadjusted) and well-matched impact estimates. Figures IV.7 and IV.8 show GED performance/GED impacts and vocational degree performance/vocational certificate impacts, respectively. In both cases, adjustment affects the relationship between performance and impacts, but the correlation remains quite weak. The correlation between reading and math performance and impacts on receipt of educational services is likewise weak (Figure IV.9); the relationship is changed little by regression adjustment.

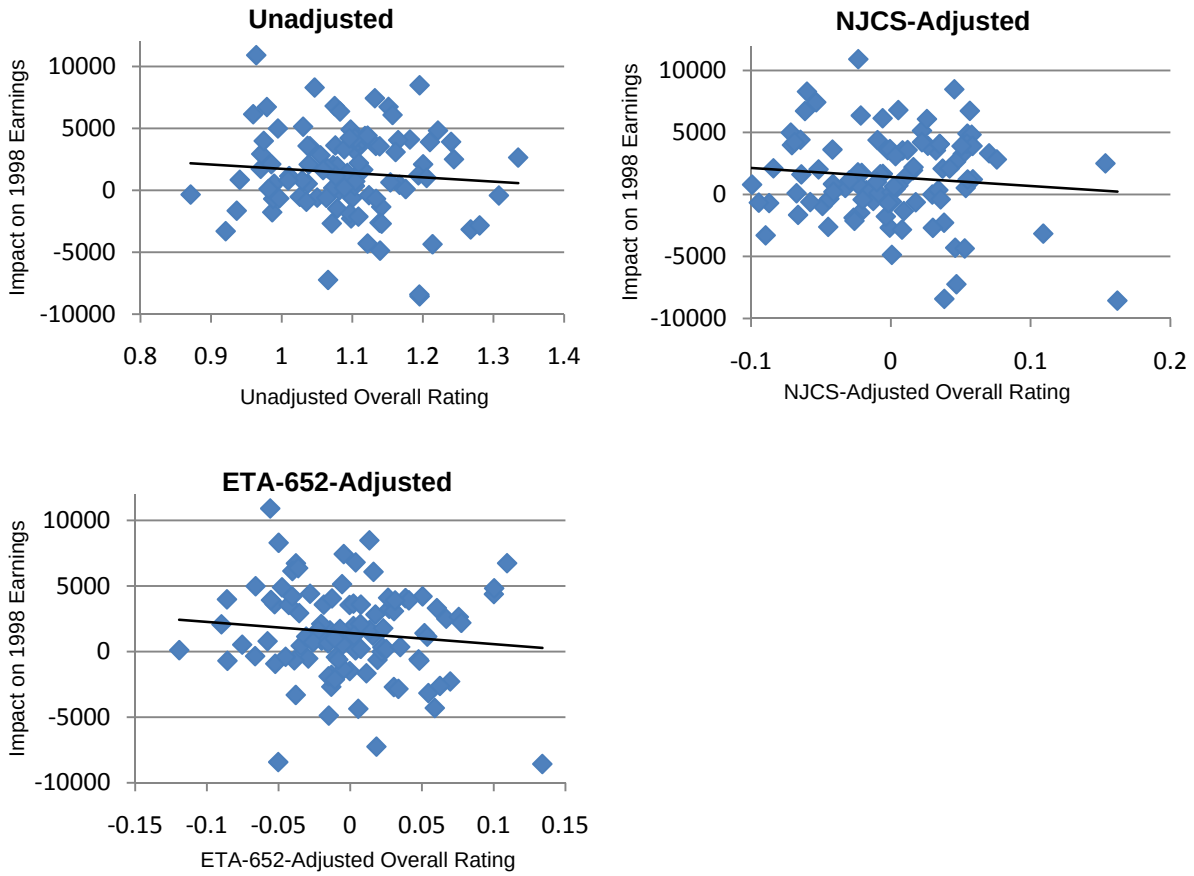
2. Correlations between Performance Measures and Impacts

To summarize further the relationship between performance measures and impact estimates, we calculated the correlations between regression-adjusted performance measures and impact estimates and compared them to the correlations using unadjusted performance measures. Table IV.8 shows estimates of the correlations between center-level impacts and several categories of multiyear average performance measure ratings: the overall rating, the GED rating, the vocational completion rating, the average wage rating, and the placement rating. For each performance measure and impact estimate, we estimated the correlation between the impact estimate and each of the unadjusted, NJCS-adjusted, and ETA-652-adjusted performance ratings.

When we tested for significance of the correlation coefficients, we found the correlations to be generally weak and insignificant (Table IV.8). One exception is the positive (and marginally significant) correlation between unadjusted performance measures (overall, vocational completion rate, and placement) and impacts on hours of educational services. The other notable exception is the negative correlation between several performance measures (adjusted and unadjusted) and impacts on 1997 and 1998 earnings. The negative correlation between earnings and performance measures appears for different adjustments and different measures of performance.

Another approach to summarizing this relationship is to estimate the rank correlations between performance measures and impact estimates. In Table IV.9, we present results that show the Spearman rank correlations between different center-level impact estimates and different performance measures (adjusted and unadjusted). The rank correlations are similar to the correlations shown in Table IV.8. For several measures, center rankings based on impacts on 1997 and 1998 earnings are negatively correlated with center performance rankings.

Figure IV.5. Three-Year Average Overall Center Performance Rating and 1998 Annual Earnings Center-Level Impacts (Unadjusted, NJCS-Adjusted, and ETA-652-Adjusted Performance)

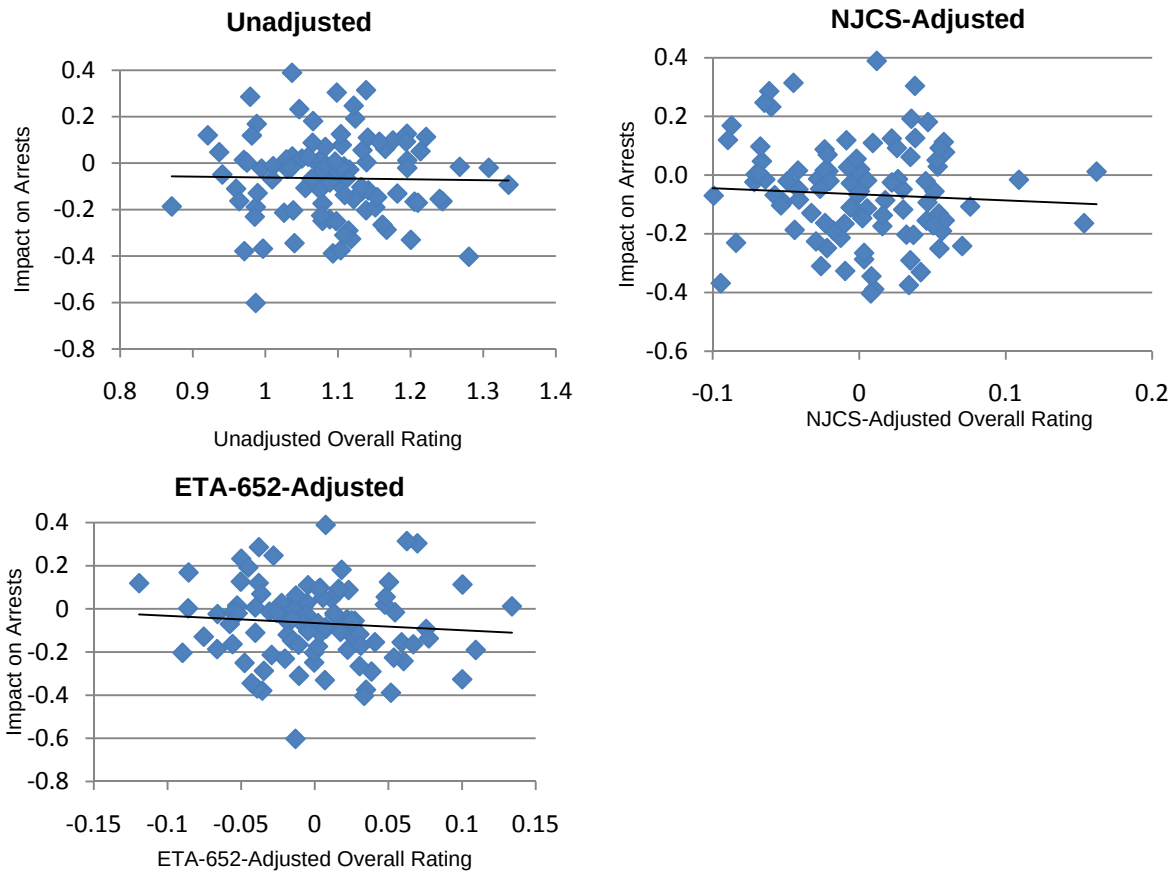


Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups. In all three graphs, the slopes are not statistically significant.

Figure IV.6. Three-Year Average Overall Center Performance Rating and Arrests Center-Level Impacts (Unadjusted, NJCS-Adjusted, and ETA-652-Adjusted Performance)

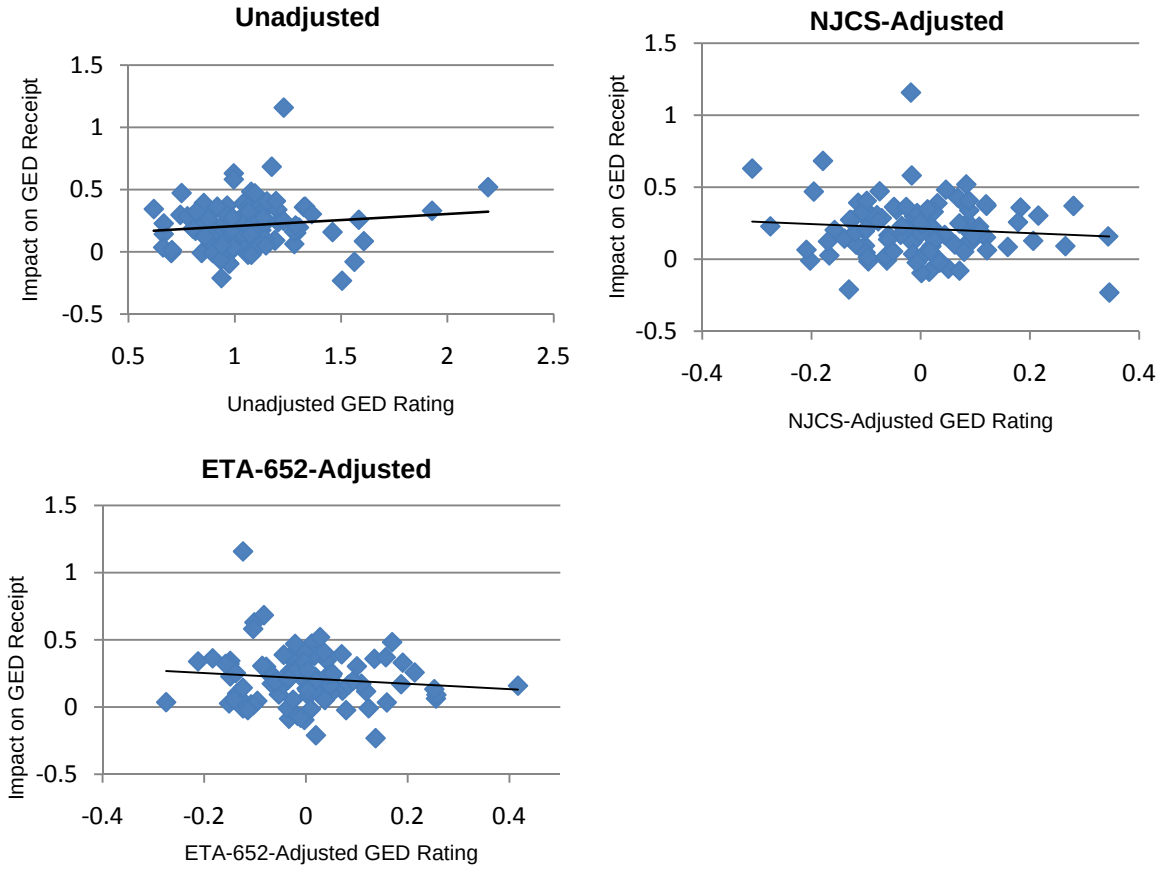


Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups. In all three graphs, the slopes are not statistically significant.

Figure IV.7. Three-Year Average GED Center Performance Rating and GED Receipt Center-Level Impacts (Unadjusted, NJCS-Adjusted, and ETA-652-Adjusted Performance)

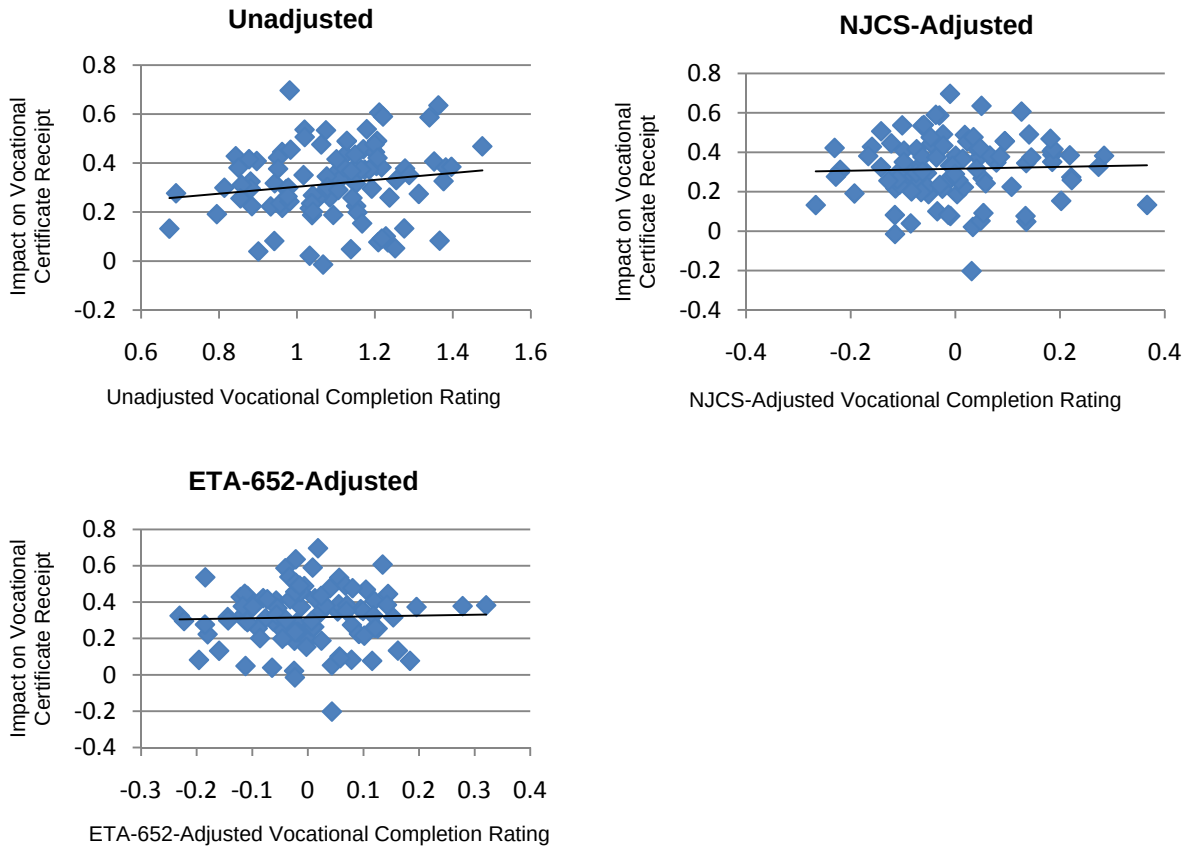


Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups. In all three graphs, the slopes are not statistically significant.

Figure IV.8. Three-Year Average Vocational Completion Center Performance Rating and Vocational Certificate Receipt Center-Level Impacts (Unadjusted, NJCS-Adjusted, and ETA-652-Adjusted Performance)

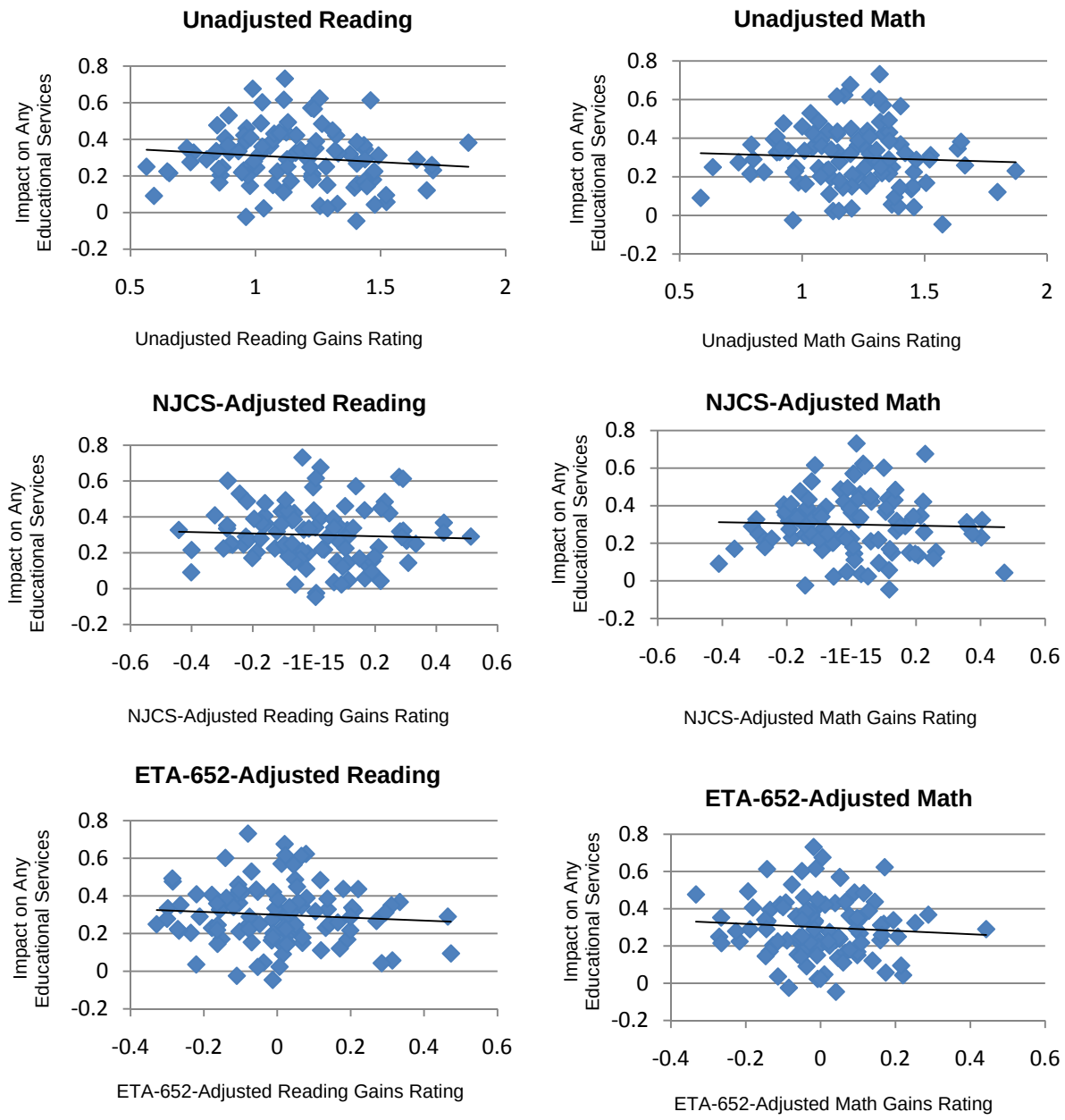


Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups. In all three graphs, the slopes are not statistically significant.

Figure IV.9. Two-Year Average Reading and Math Gains Center Performance Ratings and Any Educational Services Center-Level Impacts (Unadjusted, NJCS-Adjusted, and ETA-652-Adjusted Performance)



Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups. In all six graphs, the slopes are not statistically significant.

Table IV.8. Correlations between Center-Level Impacts and Multiyear Average Performance Ratings (Unadjusted, NJCS-Adjusted, and ETA-652-Adjusted)

	Overall Rating			GED Rating			Vocational Completion Rating			Average Wage Rating			Placement Rating		
Outcome for Impact Estimate	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj
Any Educational Services	-0.02	-0.06	-0.01	-0.08	0.08	-0.02	0.08	0.06	0.07	0.09	0.13	0.15	0.05	-0.75	0.04
Hours of Educational Services	0.17*	-0.03	0.08	0.02	-0.03	-0.03	0.19*	0.13	0.06	-0.01	-0.04	-0.04	0.19*	0.16	0.12
GED Receipt	0.15	-0.08	-0.10	0.12	-0.10	-0.11	0.13	0.02	-0.12	0.05	0.10	-0.08	0.13	0.06	-0.10
Vocational Certificate Receipt	0.13	-0.04	-0.01	0.00	-0.06	-0.14	0.14	0.04	0.03	-0.11	0.05	-0.07	0.23**	0.12	0.08
Ever Arrested	-0.02	-0.06	-0.09	0.02	0.07	-0.03	-0.06	-0.10	-0.12	-0.04	-0.08	-0.01	-0.04	-0.04	0.03
1997 Annual Earnings	-0.14	-0.19*	-0.22**	-0.22**	-0.32***	-0.25**	-0.08	-0.05	-0.13	0.07	0.03	-0.18*	0.03	0.04	-0.01
1998 Annual Earnings	-0.09	-0.11	-0.11	-0.28***	-0.32***	-0.23**	-0.02	-0.01	-0.01	0.05	0.02	-0.18*	0.08	0.11	0.03

Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: Table shows the correlation based on a multiyear average of the center's performance rating and the center-level impact estimate. NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

Table IV.9. Rank Correlations between Center-Level Impacts and Multiyear Average Performance Ratings (Unadjusted, NJCS-Adjusted, and ETA-652-Adjusted)

	Overall Rating			GED Rating			Vocational Completion Rating			Average Wage Rating			Placement Rating		
Outcome for Impact Estimate	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj	Unadj	NJCS-Adj	ETA-Adj
Any Educational Services	-0.05	-0.11	-0.04	-0.10	0.06	-0.05	0.04	0.06	0.06	0.04	0.11	0.16	-0.01	-0.10	0.01
Hours of Educational Services	0.21**	-0.06	0.06	0.04	-0.03	-0.08	0.21**	0.14	0.08	0.00	-0.02	-0.02	0.20**	0.17*	0.08
GED Receipt	0.11	-0.08	-0.10	0.13	-0.01	-0.06	0.08	-0.01	-0.09	0.03	0.09	-0.10	0.10	0.06	-0.05
Vocational Certificate Receipt	0.19*	0.00	0.00	-0.04	-0.03	-0.12	0.17*	0.08	0.05	-0.02	0.06	-0.07	0.23**	0.17	0.09
Ever Arrested	-0.03	-0.07	-0.11	0.05	0.07	-0.07	-0.09	-0.19*	-0.11	-0.04	-0.04	0.00	-0.05	-0.04	0.02
1997 Annual Earnings	-0.11	-0.14	-0.21**	-0.21**	-0.25**	-0.20***	-0.09	-0.02	-0.17*	0.08	0.02	-0.18*	0.04	0.07	0.00
1998 Annual Earnings	-0.02	0.04	-0.07	-0.23**	-0.21**	-0.11	0.01	0.08	-0.03	0.06	0.05	-0.16	0.09	0.17*	0.05

Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: Table shows the Spearman rank correlation based on a multiyear average of the center's performance rating and the center-level impact estimate. NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

Table IV.10. 1998 Annual Earnings Center-Level Impacts and Three-Year Average Overall Adjusted Performance Terciles

1998 Annual Earnings Impacts Tercile	NJCS-Adjusted Three-Year Average Overall Performance Tercile			ETA-652-Adjusted Three-Year Average Overall Performance Tercile		
	Low	Medium	High	Low	Medium	High
Low	12	11	10	9	12	12
Medium	12	12	9	10	13	10
High	9	10	15	14	8	12
Number of Centers	33	33	34	33	33	34

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

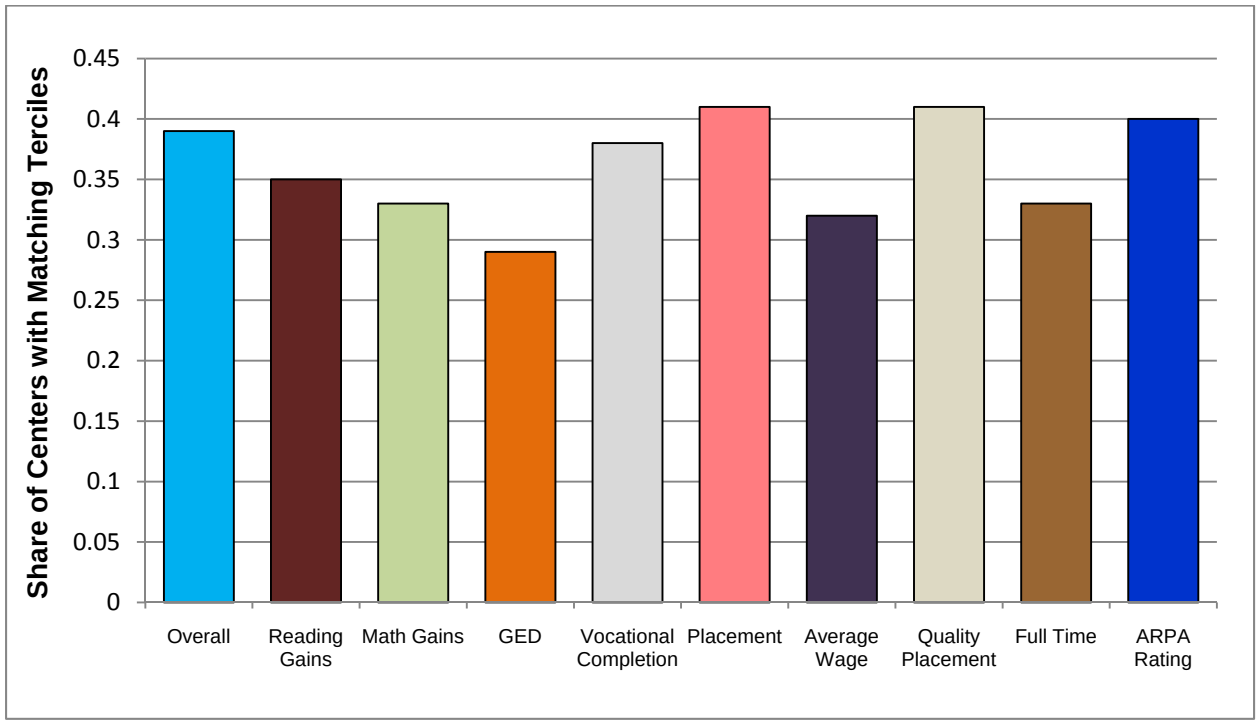
Table IV.10 is a contingency table that shows the correspondence between low, medium, and high centers based on their 1998 earnings impacts and three-year average adjusted overall performance rankings. Using the NJCS adjustment, there are 39 centers with 1998 earnings impacts and three-year average adjusted overall performance ratings in the same tercile, with 61 centers in different terciles. Using the ETA-652 adjustment, there is little correlation between impact tercile and the overall rating tercile, with 34 centers in the same tercile. Statistical chi-square tests confirm the independence of the impact and performance tercile counts.

For other performance measures, the share of centers with NJCS-adjusted performance and 1998 earnings impacts in matching terciles is similar to or smaller than the share for the overall rating (Figure IV.10). Depending on the performance measure, between 29 percent and 41 percent of centers have NJCS-adjusted performance measures and 1998 earnings impacts in the same tercile. Because the correspondence is not appreciably different from what we would expect between two unrelated measures, this suggests that the correlation between center performance and 1998 earnings impacts is relatively weak, which confirms the findings in Figure IV.5 and Tables IV.8 and IV.9.

3. Correlations between Performance Measures and Impacts by Center Characteristics

The analysis presented thus far has pooled results across all centers; however, it is possible that there are stronger relationships between impacts and performance measures for centers with particular characteristics. For instance, it could be that for centers in a particular region, the characteristics used for adjustment include the ones that differ across centers in that region; in another region, there may be more unobservable characteristics that are not included in the adjustment process. If so, adjusted performance measures may more closely mimic impacts for centers in the first region, but may be less closely related in the other region.

Figure IV.10. Correspondence between Terciles of 1998 Annual Earnings Impacts and NJCS-Adjusted Center Performance Ratings, Multiyear Averages



Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, 2008 ARF, NJCS follow-up surveys.

Notes: Figure shows the share of centers with NJCS-adjusted multiyear average performance ratings and 1998 annual earnings impacts in the same tercile. NJCS-adjusted performance terciles are terciles based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

Therefore, in addition to looking at results overall, we also calculated performance-impact correlations for subgroups defined by center characteristics, in particular (1) size (number of slots), (2) type of management (privately operated or operated by a Federal agency), and (3) region.

Table IV.11 shows the correlations between center-level impacts and the three-year average overall performance rating by center size. (The table shows correlations for both the NJCS adjustment and the ETA-652 adjustment, both of which also include adjustment for local area characteristics.) The correlation between the three-year average overall rating and impacts is generally small and insignificant. Consistent with the pooled results, we see some evidence that 1997 earnings impacts are negatively related to the three-year average overall performance rating.

Looking separately across privately-operated centers and centers operated by a Federal agency, the correlations between center-level impacts and the three-year average adjusted overall performance rating are small and insignificant for most outcomes (Table IV.12). However, among privately-operated centers, the three-year average overall rating is negatively correlated with impacts on 1997 annual earnings. Among centers operated by a Federal agency, when using the ETA-652

adjustment, the correlation is likewise negative and similar in magnitude, but insignificant. These results are generally consistent with the overall pooled results, which show weak correlations between impacts and performance measures, with the exception of an unexpected negative relationship between earnings impacts and performance measures.

Table IV.11. Correlations between Center-Level Impacts and Three-Year Average Overall Performance Rating (NJCS-Adjusted and ETA-652-Adjusted), by Center Size

	NJCS-Adjusted Overall Rating			ETA-652-Adjusted Overall Rating		
Outcome for Impact Estimate	Small Centers	Medium Centers	Large Centers	Small Centers	Medium Centers	Large Centers
Any Educational Services	-0.14	0.02	-0.02	-0.05	-0.06	0.32
Hours of Educational Services	0.10	-0.07	-0.05	0.20	-0.02	0.47*
GED Receipt	-0.02	-0.02	-0.42	-0.14	0.02	-0.29
Vocational Certificate Receipt	-0.17	-0.07	0.31	-0.10	0.02	0.30
Ever Arrested	0.02	-0.05	-0.27	0.06	-0.21	-0.15
1997 Annual Earnings	-0.14	-0.25*	-0.05	-0.30*	-0.10	-0.28
1998 Annual Earnings	-0.15	-0.07	-0.03	-0.18	0.01	-0.36
Number of Centers	36	48	16	36	48	16

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, center characteristics, 2008 ARF, NJCS follow-up surveys.

Notes: Center size is based on the total number of slots in the center, not the number of NJCS participants. Small centers are those with fewer than 226 students, medium centers have between 226 and 495 students, and large centers have more than 495. NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

Finally, in Table IV.13, we show the correlations between center-level impacts and the three-year average overall performance rating by center region. The negative correlation between earnings impacts and the overall rating seems to be concentrated in a few regions, especially Region 2. Although there are a handful of significant correlations in different regions for different outcomes, there do not seem to be regions in which impacts and the overall rating are positively and significantly correlated across impact outcomes (or negatively correlated, in the case of arrests).

Together, these results suggest that the performance measurement system—even after adjusting performance measures for participant characteristics—does not mimic center-level impact estimates, even for particular types of centers.

Table IV.12. Correlations between Center-Level Impacts and Three-Year Average Overall Performance Rating (NJCS-Adjusted and ETA-652-Adjusted), by Center Management

	NJCS-Adjusted Overall Rating		ETA-652-Adjusted Overall Rating	
Outcome for Impact Estimate	Privately Operated	Operated by a Federal Agency	Privately Operated	Operated by a Federal Agency
Any Educational Services	-0.02	-0.14	0.06	-0.20
Hours of Educational Services	0.02	-0.18	0.08	0.10
GED Receipt	-0.10	-0.05	-0.07	-0.16
Vocational Certificate Receipt	-0.01	-0.19	-0.01	0.01
Ever Arrested	-0.01	-0.18	-0.03	-0.18
1997 Annual Earnings	-0.26**	-0.02	-0.21*	-0.27
1998 Annual Earnings	-0.17	0.12	-0.12	-0.09
Number of Centers	71	29	71	29

Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, center characteristics, 2008 ARF, NJCS follow-up surveys.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

Table IV.13. Correlations between Center-Level Impacts and Three-Year Average NJCS-Adjusted Overall Performance Rating, by Center Region

	NJCS-Adjusted Overall Rating								
	Center Region								
Outcome for Impact Estimate	1	2	3	4	5	6	7	9	10
Any Educational Services	-0.84	0.65	-0.23	0.03	0.51	-0.32	-0.29	-0.58	-0.54
Hours of Educational Services	-0.91*	-0.07	0.17	0.08	0.02	0.21	-0.24	-0.32	-0.44
GED Receipt	-0.56	-0.32	0.13	-0.29	0.25	-0.02	0.24	-0.80**	-0.04
Vocational Certificate Receipt	-0.84	0.11	0.49	-0.12	0.24	-0.13	-0.15	-0.44	0.24
Ever Arrested	0.30	-0.24	0.19	-0.33	0.01	0.10	-0.17	-0.34	0.12
1997 Annual Earnings	0.23	-0.92***	0.03	-0.05	0.06	-0.33	0.17	-0.44	-0.12
1998 Annual Earnings	0.17	-0.82**	-0.56*	0.40*	0.29	-0.34	0.43	-0.44	0.12
Number of Centers	4	6	12	22	10	15	13	8	10

Sources: Performance measure data, NJCS baseline survey, center characteristics, 2008 ARF, NJCS follow-up surveys.

Notes: NJCS-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

4. Correlations between Performance Measures and Impacts by Performance Tercile

In addition to looking at results separately by center characteristics, we can also look at how the correlations vary across different parts of the performance measure distribution. In Table IV.14, we show the correlations between center-level impacts and the three-year average overall performance rating by adjusted rating tercile. We show results for both the NJCS and ETA-652 adjustments (both with local area characteristics).

The correlation between impacts and performance ratings is generally weak within terciles as well (Table IV.14). The correlation between the NJCS-adjusted overall rating and 1997 earnings impacts is negative and significant for the highest tercile. However, this relationship is not consistent across adjustments: the correlation between the ETA-652-adjusted overall rating and 1998 earnings impacts is positive and significant for the middle tercile.

Table IV.14. Correlations between Center-Level Impacts and Three-Year Average Overall Performance Rating (NJCS-Adjusted and ETA-652-Adjusted), by Adjusted Rating Tercile

	NJCS-Adjusted Overall Rating			ETA-652-Adjusted Overall Rating		
Outcome for Impact Estimate	Low	Medium	High	Low	Medium	High
Any Educational Services	0.00	0.03	0.21	0.10	-0.07	0.17
Hours of Educational Services	-0.09	0.05	0.15	0.27	0.00	0.29*
GED Receipt	-0.25	-0.17	0.07	0.14	0.10	0.17
Vocational Certificate Receipt	-0.13	-0.25	-0.06	0.10	0.20	0.23
Ever Arrested	-0.02	-0.01	-0.04	-0.13	0.22	0.04
1997 Annual Earnings	-0.05	-0.02	-0.31*	-0.14	-0.06	-0.07
1998 Annual Earnings	-0.01	-0.02	-0.27	0.06	0.32*	-0.04
Number of Centers	33	33	34	33	33	34

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, 2008 ARF, NJCS follow-up surveys.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

D. Sensitivity Analyses

Our core analyses, described above, found that the relationship between performance measures and impacts is weak, whether or not performance measures have been adjusted for participant characteristics. However, because of sampling variability, center-level impacts, performance measures, and baseline characteristics may be noisy estimates of the parameters of interest. To assess the degree to which measurement error affects our findings, we conducted several robustness checks.

First, we restricted the analysis to centers with more NJCS observations, for which center-level measures should be estimated with greater precision. If we observe a stronger correlation between impacts and performance measures among centers with more observations, it would suggest that measurement error may have muted the estimated correlation in the full sample. (Appendix Table A.7 confirms that the standard errors of impact estimates are smaller, on average, for centers with more observations.)

When the sample is restricted to centers with 100 or more NJCS follow-up observations, the negative correlation between earnings impacts and performance measures is no longer significant (Table IV.15). Across the 70 correlations in Table IV.15, there are a handful of significant relationships. The most robust is the positive correlation between the NJCS-adjusted average wage

rating and impacts on educational services (receipt and hours) and vocational certificate receipt. This suggests that the NJCS-adjusted average wage rating may do a better job of tracking impacts than is evident using the full sample; this could be either because the 60 excluded centers (with fewer than 100 observations) are different from centers with more observations, or because measurement error attenuates the correlation.

Table IV.15. Correlations between Center-Level Impacts and Multiyear Average Center Performance Rating (NJCS-Adjusted), Sample Restricted to Centers with at Least 100 Observations

	NJCS-Adjusted Rating									
Outcome for Impact Estimate	Overall	Reading Gains	Math Gains	GED	Vocational Completion	Placement	Average Wage	Quality Placement	Full-Time	ARPA Rating
Any Educational Services	-0.01	0.00	-0.11	0.08	0.07	0.08	0.48***	0.07	0.02	-0.01
Hours of Educational Services	-0.10	-0.14	-0.25	-0.07	0.09	0.13	0.39**	-0.04	-0.08	-0.06
GED Receipt	-0.16	-0.08	-0.21	-0.04	0.00	0.07	0.08	0.10	-0.26	-0.12
Vocational Certificate Receipt	0.19	0.27*	0.24	0.13	0.28*	0.22	0.42***	0.28*	-0.15	0.18
Ever Arrested	-0.10	-0.17	-0.21	0.06	-0.26	-0.01	0.07	0.16	0.05	-0.01
1997 Annual Earnings	-0.05	-0.04	-0.09	-0.14	0.02	0.24	-0.02	0.16	-0.24	0.00
1998 Annual Earnings	0.10	0.16	0.04	-0.17	0.15	0.27*	-0.09	0.15	-0.11	0.13

Sample size = 40 centers.

Sources: Performance measure data, NJCS baseline survey, 2008 ARF, NJCS follow-up surveys.

Notes: Sample restricted to the 40 centers with 100 or more observations in the NJCS follow-up data. NJCS-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

An alternative approach to testing for the role of measurement error is to estimate the relationship between center-level impacts and performance measures, giving more weight to larger centers rather than giving each center equal weight (our benchmark approach).

In Table IV.16, we show correlations—calculated using the full sample—between NJCS-adjusted performance measures and impact estimates. However, rather than weighting each center equally, we weighted it by the sum of the NJCS baseline weights among participants predicted to

attend that center. This approach gives more weight to centers in which impacts and participant characteristics are more precisely measured. However, our results do not look altogether different from our main approach; most correlations are small and insignificant. Consistent with our main results, there is some evidence of a negative correlation between the GED rating and impacts on earnings. Finally, the positive correlation between the average wage rating and impacts on educational outcomes (observed in Table IV.14, when the sample was restricted to centers with more observations) is present, but is significant for only one impact outcome.

Table IV.16. Correlations between Center-Level Impacts and Multiyear Average Center Performance Rating (NJCS-Adjusted), Centers Weighted by Number of Observations

	NJCS-Adjusted Rating									
Outcome for Impact Estimate	Overall	Reading Gains	Math Gains	GED	Vocational Completion	Placement	Average Wage	Quality Placement	Full-Time	ARPA Rating
Any Educational Services	-0.07	-0.07	-0.07	0.03	0.03	-0.07	0.25**	0.00	-0.01	-0.05
Hours of Educational Services	-0.04	-0.07	0.01	-0.08	0.13	0.07	0.14	-0.01	-0.03	-0.02
GED Receipt	-0.10	-0.03	-0.01	-0.11	0.01	0.07	0.12	0.04	-0.10	-0.05
Vocational Certificate Receipt	0.03	0.12	0.09	0.04	0.13	0.20**	0.15	0.14	-0.08	-0.08
Ever Arrested	-0.07	-0.03	-0.07	0.04	-0.18*	-0.07	-0.01	0.03	-0.13	0.05
1997 Annual Earnings	-0.14	-0.08	-0.15	-0.25**	-0.02	0.14	-0.02	0.00	-0.08	-0.12
1998 Annual Earnings	-0.04	0.02	-0.05	-0.23**	0.04	0.18*	-0.05	0.03	-0.08	0.00

Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, 2008 ARF, NJCS follow-up surveys.

Notes: Centers are weighted by the sum of the NJCS baseline weights. NJCS-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) but not center characteristics. When constructing center-level averages, baseline characteristics are weighted using the baseline weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

E. Analytical Extensions

To explore the implications of our findings, we also investigated the relationship between center-level impact estimates and *alternative* measures of center performance from the NJCS follow-up interview data. These exploratory analyses were conducted to help frame our findings and put them in context.[14]

The performance measurement system gauges outcomes of Job Corps participants. Using the NJCS *follow-up survey data,* we constructed estimates that are similar in spirit to these performance measures. In particular, we used, as a substitute for the performance measures, center-level average outcomes for the subset of NJCS treatment group members who participated in the Job Corps program. The correlation between impacts and treatment participant outcomes could differ from the correlation between impacts and the Job Corps performance measures, either because the data may be collected differently (such as over a different time frame or for a broader sample) or because different outcomes may be measured.

If center-level treatment group participant averages of outcomes based on the follow-up survey data are highly correlated with center-level impact estimates, this may suggest potential additions or modifications to the Job Corps performance measurement system. For instance, if the correlation between impacts and outcomes depends on the data source (NJCS followup or the performance measurement system), the mode of data collection may matter. Likewise, if the correlation between impacts and outcomes is strong for outcomes not currently collected through the performance measurement system, those outcomes may be viable alternative measures of center performance that could be collected moving forward. To keep the analysis manageable, our analysis focused on the outcomes that were used in generating impact estimates, with one addition (length of stay in a Job Corps center).

An interesting finding is that center-level treatment participant outcome averages are positively correlated with impacts for several outcome and impact categories (Table IV.17). In particular, center-level treatment participant averages of hours of educational services, GED receipt, vocational certificate receipt, and length of stay in a Job Corps center are all positively and significantly correlated with impacts on hours of educational services, GED receipt, and vocational certificate receipt. Treatment participant rates of vocational certificate receipt are also positively and significantly correlated with impacts on 1998 earnings. In addition, both earnings outcomes are positively and significantly correlated with both earnings impacts.

[14] We had also planned to estimate the optimal weighting scheme by identifying the linear combination of components that was most highly correlated with program impacts. However, because the individual components are generally uncorrelated with impacts, we did not pursue this extension.

Table IV.17. Correlations between Center-Level Impacts and Center-Level Treatment Participant Outcomes

	Outcome for Treatment Participants						
Outcome for Impact Estimate	Hours of Educational Services	GED Receipt	Vocational Certificate Receipt	Ever Arrested	1997 Annual Earnings	1998 Annual Earnings	Length of Stay in Job Corps Center
Any Educational Services	-0.12	-0.01	0.01	0.19*	0.18*	0.01	-0.03
Hours of Educational Services	0.56***	0.33***	0.40***	0.00	0.16	0.11	0.54***
GED Receipt	0.42***	0.66***	0.41***	-0.09	0.18*	0.15	0.42***
Vocational Certificate Receipt	0.30***	0.33***	0.64***	0.00	0.23**	0.15	0.43***
Ever Arrested	0.07	-0.10	0.08	0.29***	-0.14	-0.12	-0.02
1997 Annual Earnings	0.00	0.11	0.14	0.10	0.45***	0.37***	0.12
1998 Annual Earnings	0.08	0.14	0.25**	0.10	0.36***	0.53***	0.14

Sample size = 100 centers.

Sources: NJCS baseline survey, NJCS follow-up surveys.

Notes: "Any educational services" is excluded as an outcome because 100 percent of treatment group participants received educational services (by definition, because they participated in Job Corps). Average center-level outcomes for the treatment group are estimated among Job Corps participants in the treatment group and are weighted using the follow-up weight. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

The correlation between treatment participant outcomes and impact estimates is relatively strong, and is noticeably different from the weak relationship between the Job Corps performance measures and impact estimates (see Section IV.C above). To explore these differences, we analyzed the relationship between treatment participant "performance measures", the Job Corps performance measures, and the NJCS outcomes for the treatment and control groups. Because our impact estimates are similar to treatment-control differences, the correlation between performance measures and impacts can be approximated by the difference between two components: the correlation between performance measures and *treatment* group outcomes and the correlation between performance measures and *control* group outcomes. If performance measures are highly correlated with treatment group outcomes and less correlated with control group outcomes, we will observe a correlation between performance measures and impacts. On the other hand, if both treatment and control group outcomes are equally correlated (or uncorrelated) with performance measures, the correlation between performance measures and impacts will be small.

In Table IV.18, we show the correlations between treatment participant performance measures and (1) treatment group outcomes (top panel) and (2) control group outcomes (bottom panel). Comparing the top and bottom panels reveals that the positive and significant correlations in Table

IV.17 are likely the result of large, positive correlations between treatment participant performance measures and treatment group outcomes, and smaller correlations between treatment participant performance measures and control group outcomes. For example, the correlation between vocational certificate receipt among treatment participants and 1998 annual earnings among treatment group members is 0.22, whereas the correlation between vocational certificate receipt among treatment participants and 1998 annual earnings among control group members is -0.08 (and insignificant); it is essentially the difference between these correlations that drives the positive and significant correlation between vocational certificate receipt among treatment participants and impacts on 1998 annual earnings (Table IV.17).

Table IV.18. Correlations between Center-Level Treatment Group/Control Group Outcomes and Center-Level Treatment Participant Outcomes

	Outcome for Treatment Participants						
	Hours of Educational Services	GED Receipt	Vocational Certificate Receipt	Ever Arrested	1997 Annual Earnings	1998 Annual Earnings	Length of Stay in Job Corps Center
Treatment Group Outcomes							
Any Educational Services	0.14	-0.01	0.01	0.05	-0.13	0.05	0.08
Hours of Educational Services	0.86***	0.32***	0.48***	-0.16	-0.09	0.05	0.73***
GED Receipt	0.42***	0.89***	0.44***	-0.09	0.27***	0.31***	0.50***
Vocational Certificate Receipt	0.47***	0.38***	0.91***	-0.01	0.23**	0.24**	0.53***
Ever Arrested	-0.26***	-0.07	-0.06	0.92***	0.17	0.10	-0.23**
1997 Annual Earnings	-0.07	0.23**	0.25**	0.20**	0.89***	0.79***	0.16
1998 Annual Earnings	0.03	0.26***	0.22**	0.11	0.69***	0.93***	0.22**
Control Group Outcomes							
Any Educational Services	0.18*	0.03	0.00	-0.22**	-0.25**	-0.02	0.04
Hours of Educational Services	0.16	-0.04	-0.02	-0.22**	-0.31***	-0.16	0.03
GED Receipt	-0.11	0.12	-0.12	-0.02	-0.02	0.06	-0.04
Vocational Certificate Receipt	0.06	-0.02	0.06	-0.09	-0.08	-0.01	-0.07
Ever Arrested	-0.27***	0.03	-0.13	0.43***	0.26***	0.19*	-0.15
1997 Annual Earnings	-0.09	0.06	0.05	0.05	0.25**	0.27***	-0.01
1998 Annual Earnings	-0.09	0.09	-0.08	-0.04	0.24**	0.27***	0.05

Sample size = 100 centers.

Sources: NJCS baseline survey, NJCS follow-up surveys.

Notes: "Any educational services" is excluded as a treatment participant outcome because 100 percent of treatment group participants received educational services (by definition, because they participated in Job Corps). Average center-level outcomes for each group (treatment participants, treatment members, and control members) are weighted using the follow-up weight. In the top panel, correlations between corresponding outcomes are not equal to 1 because the comparison is between the treatment group as a whole and treatment group participants.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

We can similarly decompose the correlations between the Job Corps performance measures and impacts. The correlations between the unadjusted Job Corps performance measures and center-level treatment group outcomes are sometimes positive and significant (top panel of Table IV.19), though of smaller magnitude than the correlations from above in the top panel of Table IV.18. Unadjusted Job Corps performance measures and center-level control group outcomes exhibit low-to-moderate correlations (bottom panel of Table IV.19). Thus, the key difference between the Job Corps performance measure-impact relationship and the treatment participant performance measure-impact relationship seems to be that the correlations between Job Corps performance measures and treatment group outcomes are smaller.[15]

This analysis suggests that the weak correlations between the Job Corps performance measures and NJCS impacts reflect the fact that the Job Corps performance measures and treatment group outcomes are not highly correlated (but are only moderately correlated). In contrast, when we substitute NJCS treatment participant outcomes for performance measures, the correlations are higher (even across different outcomes). These differences could reflect differences in the samples used to construct the measures, in the time frame over which data are collected, or in how outcomes are defined and reported. These differences could also be spurious: because NJCS performance measures and impacts are based on overlapping, small samples, the positive correlations could reflect measurement error.[16] Further exploring these differences is beyond the scope of this study; this analysis would require Job Corps management information system and performance data on individual treatment group members that could then be compared to the survey data.

[15] We also looked separately at the correlations between treatment group performance measures and unadjusted Job Corps performance measures. As shown in Appendix Table A.8, the correlations are generally positive and significant, with a few exceptions. The correlations between treatment participant earnings and the GED receipt performance measure are negative and significant; likewise, the correlations between educational outcomes and the full-time performance measures are negative and significant.

[16] Performance measures are calculated based on outcomes among all treatment participants (subject to response rates), whereas NJCS treatment participant outcomes are based on NJCS participants. If the NJCS sample is sufficiently large, average outcomes for NJCS participants should match average outcomes for all treatment participants. However, if the NJCS sample sizes are small, there may be measurement error in center-level measures. The positive correlations between impacts and treatment participant performance measures documented here could reflect spurious correlations if impact estimates and treatment group participant average outcomes are both measured with error due to sampling variability. One way to test whether this is a concern is to split the sample, estimating the NJCS performance measures using half of the sample and estimating impacts using the other half of the sample. Though this ensures that the samples do not overlap, this test has the drawback that it further reduces the sample sizes, introducing uncorrelated noise in the estimates, potentially attenuating the correlations. We find that the correlations between NJCS performance measures and impacts are considerably smaller under the split sample approach, suggesting that measurement error is indeed a concern.

Table IV.19. Correlations between Center-Level Treatment Group/Control Group Outcomes and Unadjusted Performance Ratings

	Unadjusted Performance Rating									
	Overall	Reading Gains	Math Gains	GED	Vocational Completion	Placement	Average Wage	Quality Placement	Full-Time	ARPA Rating
Treatment Group Outcomes										
Any Educational Services	0.16	-0.05	-0.02	0.11	0.19*	0.07	0.22**	0.10	-0.04	0.12
Hours of Educational Services	0.39**	0.21**	0.20*	0.39***	0.29***	0.22**	0.04	0.29***	-0.21**	0.39***
GED Receipt	0.28**	0.08	0.22**	0.21**	0.20**	0.37***	0.26***	0.21**	-0.27***	0.26**
Vocational Certificate Receipt	0.34**	0.17*	0.25**	0.13	0.27***	0.29***	0.10	0.32***	-0.13	0.45***
Ever Arrested	-0.13	-0.09	-0.01	-0.29***	-0.04	-0.13	0.22**	0.00	0.20**	-0.19*
1997 Annual Earnings	0.02	-0.09	0.03	-0.24**	-0.01	0.31***	0.41***	0.20*	0.15	0.02
1998 Annual Earnings	0.11	-0.02	0.05	-0.20**	0.10	0.32***	0.44***	0.25**	0.07	0.08
Control Group Outcomes										
Any Educational Services	0.07	0.11	0.05	0.10	-0.02	-0.02	-0.03	0.06	-0.09	0.10
Hours of Educational Services	0.15	0.18*	0.12	0.34***	0.04	-0.04	-0.01	0.01	-0.20**	0.12
GED Receipt	0.07	-0.01	0.08	0.02	0.04	0.22**	0.19*	0.05	-0.14	0.01
Vocational Certificate Receipt	0.12	0.13	0.18*	0.06	0.07	-0.02	0.20**	0.01	-0.17	0.19*
Ever Arrested	-0.09	-0.07	-0.02	-0.24**	0.01	-0.06	0.21**	0.01	0.27***	-0.19*
1997 Annual Earnings	0.16	0.09	0.19*	0.02	0.07	0.24**	0.29***	0.15	-0.04	0.16
1998 Annual Earnings	0.19*	0.06	0.12	0.10	0.11	0.22**	0.39***	0.27***	-0.03	0.12

Sample size = 100 centers.

Sources: NJCS baseline survey, NJCS follow-up surveys.

Notes: Average center-level outcomes for each group (treatment and control) are weighted using the follow-up weight.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

V. CONCLUSIONS

This study extends previous work analyzing the relationship between Job Corps performance measures and center-level impact estimates from the NJCS (Schochet and Burghardt 2008). That earlier work found no systematic relationship between a summary measure of Job Corps center performance and program impacts. Here, we build on the previous study, constructing new measures of center performance that may or may not be a better reflection of impacts. In particular, we constructed regression-adjusted performance measures—which adjust for differences in individual and local area characteristics of center participants—to measure the component of center performance that is not explained by these characteristics.

We compared adjusted and unadjusted performance measures (including different components and different program years) to center-level impacts on a range of outcomes, including educational services, educational attainment, arrests, and earnings. Our key findings are summarized as follows:

- Students in high-performing centers are significantly different from students in low-performing centers; though the characteristics that are more common among students in high-performing centers are generally associated with better outcomes, there are some exceptions. Differences in the characteristics of students served by centers in different performance terciles are relatively small in magnitude.
- Regression-adjusting for characteristics changes center performance rankings, but not dramatically. Regression-adjusted and unadjusted performance measures are positively correlated, although there are differences.
- Regression-adjusted performance measures are no better than unadjusted performance measures at distinguishing between centers with larger impacts and those with smaller impacts. The correlations between impacts and performance measures are generally weak and insignificant. Similar results apply using the ETA-652 data and the more detailed NJCS baseline survey data.
- Our findings hold for overall measures of performance as well as components of center performance and different program years; that is, the relationship between impacts and different performance measure components is also generally weak.
- Among the subgroups we analyzed, there are not particular groups of centers for which performance measures track impacts.
- Outcomes for treatment group participants as measured using the *NJCS follow-up survey data*—which are conceptually similar to performance measures—are positively correlated with impacts. Exploring these findings may be a promising avenue for future research.

Overall, we find that the relationship between performance measures and impacts is weak, whether or not they have been adjusted for participant characteristics. Although regression adjustment had some effect on the performance rankings, it did not change their ability to mimic impacts.

Our results may not be surprising when considered in relation to previous studies that have found that impact estimates based on nonexperimental methods often do not mimic those based on experimental methods. Here, regression-adjusted performance measures—which are analogous to nonexperimental impact estimates—are not strongly associated with impact estimates from the

NJCS, an experimental study. While the baseline covariates did explain some of the variance in the performance measures (and the control group outcomes), there are likely to be important unobserved differences between students attending different centers. If those unobserved factors are associated with outcomes, we would not expect regression-adjusted performance measures to reflect the impact of individual centers.

There is also reason to think that measurement error may be influencing our results. First, the impact estimates may be imprecisely estimated due to relatively small sample sizes in some of the 100 centers; Appendix Table A.7 confirms that the standard errors on center impact estimates are, for smaller centers, relatively large. However, our analyses that grouped centers into performance terciles helped adjust for this imprecision and corroborated our overall correlational analyses. Another potential source of measurement error stems from the possible mismatch of the NJCS survey data that was used to construct the impact estimates and the Job Corps performance data. This theory is supported by our finding that the associations between the center impact estimates were much stronger using new "performance measures" from the NJCS survey data than the actual Job Corps performance measures. An important area for future research will be to resolve these measurement differences by comparing the two sets of performance measures by obtaining Job Corps performance data on individual treatment group members.

The weak association between performance measures and impacts could also be related to the fact that performance measures do not vary a great deal across centers. For example, the overall Job Corps performance measure covering the PY 1994 to 1996 periods varies from 0.87 to 1.34, and the interquartile range is only 0.10 points. Though this implies that there may be scope for the regression adjustment procedure to reorder centers, it has the drawback that regression adjustment may actually highlight small differences between center performance measures that may be not meaningful.

Finally, Job Corps performance measures may not be closely related to impacts because in-program and shorter-term employment outcomes used in the Job Corps performance measurement system are only relatively weakly associated with the longer-term employment outcomes that we used to measure NJCS impacts. For example, the correlation between the vocational certificate attainment rate and 1998 earnings is only about 0.25 using the NJCS treatment group sample; using the NJCS treatment group, the correlation between hours of educational services and 1998 earnings is 0.05. Similarly, as discussed in Schochet and Burghardt (2008), the post-program employment-related performance measures appear to have low correlations with longer-term labor market outcomes. Using an earnings measure from the survey data corresponding to the performance measure that was used to calculate the placement rate, Schochet and Burghardt (2008) found that the correlation of this measure with 1997 and 1998 earnings was only about 0.12. Similarly, the correlation between the hourly wage at placement (for workers) and their earnings is only 0.30. These low correlations result because there is considerable movement in and out of jobs for Job Corps youth. Similarly, the progression in wages over time varies greatly among youth, with the result that the correlation between wages at placement and later earnings is weak.

REFERENCES

Bloom, Howard S. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review,* vol. 8, 1984, pp. 225-246.

Schochet, Peter Z. "National Job Corps Study: Methodological Appendixes on the Impact Analysis." Princeton, NJ: Mathematica Policy Research, Inc., June 2001.

Schochet, Peter Z., and John A. Burghardt. "Do Job Corps Performance Measures Track Program Impacts?" *Journal of Policy Analysis and Management,* vol. 27, no. 3, summer 2008, pp. 556-576. See also the Employment and Training Administration (ETA) Research and Evaluation monograph 01-L, "National Job Corps Study: Impacts by Center Characteristics."

Schochet, Peter Z., John A. Burghardt, and Sheena M. McConnell. "Does Job Corps Work? Impact Findings from the National Job Corps Study." *American Economic Review,* vol. 68, no. 5, December 2008, pp. 1864-1886.

APPENDIX A

ADDITIONAL TABLES AND FIGURES

Table A.1. Job Corps Center Performance Measurement System for Program Years 1994, 1995, and 1996

	Program Year 1994				Program Year 1995				Program Year 1996			
Area/Measure	Pool[a]	Measure	Standard[b]	Weight[c]	Pool[a]	Measure	Standard[b]	Weight[c]	Pool[a]	Measure	Standard[b]	Weight[c]
Program Accomplishments												
Reading Gains	Scored less than **8.5 on TABE 5/6 total reading test** at program entry (or did not take test)	Percentage of students in pool who gain two grades or score 8.5 on follow-up TABE reading test	30 percent	.056	Scored less than **8.5 on TABE 5/6 total reading test** at program entry (or did not take test)	Percentage of students in pool who gain two grades or score 8.5 on follow-up TABE reading test	35 percent	.067	Scored less than **6.7 on TABE 7/8 total reading test** at program entry (or did not take test)	Percentage of students in pool who gain two grades or score **6.7 on follow-up TABE reading test**	40 percent	0
Math Gains	Scored less than **8.5 on TABE 5/6 total math test** at program entry (or did not take test)	Percentage of students in pool who gain two grades or score 8.5 on follow-up TABE math test	33 percent	.056	Scored less than **8.5 on TABE 5/6 total math test** at program entry (or did not take test)	Percentage of students in pool who gain two grades or score 8.5 on follow-up TABE math test	35 percent	.067	Scored less than **7.4 on TABE 7/8 total math test** at program entry (or did not take test)	Percentage of students in pool who gain two grades or score **6.7 on follow-up TABE math test**	45 percent	0
GED Rate	Without high school diploma and **scored 6.3 or above on TABE 5/6 total reading test** at program entry (or did not take test)	Percentage of students in pool who obtain GED/high school degree, including bonus for students who initially score low on test	Model-based	.056	Without high school diploma and **scored 6.3 or above on TABE 5/6 total reading test** at program entry (or did not take test)	Percentage of students in pool who obtain GED/high school degree, including bonus for students who initially score low on test	Model-based	.067	Without high school diploma and **scored 5.2 or above on TABE 7/8 total reading test** at program entry (or did not take test)	Percentage of students in pool who obtain GED/high school degree, including bonus for students who initially score low on test	Model-based	.20
Vocational Completion Rate	**Stayed at least 60 days and participated in a vocational program with an approved training achievement record (TAR)**	Percentage of students in pool who complete vocation at completer or advanced-completer level	56 percent	.167	**All terminees**	Percentage of students in pool who complete vocation at completer or advanced-completer level	45 percent	.20	**All terminees**	Percentage of students in pool who complete vocation at completer or advanced-completer level	45 percent	.20

	Program Year 1994				Program Year 1995				Program Year 1996			
Area/Measure	Pool[a]	Measure	Standard[b]	Weight[c]	Pool[a]	Measure	Standard[b]	Weight[c]	Pool[a]	Measure	Standard[b]	Weight[c]
Placement												
Placement Rate	All terminees plus Job Corps advanced training (AT) or advanced career training (ACT) transfers	Percentage of students in pool placed in job/military or school, with bonus for AT/ACT transfers	69 percent	.111	All terminees plus AT/ACT transfers	Percentage of students in pool placed in job/military or school, with bonus for AT/ACT transfers	70 percent	.16	All terminees plus AT/ACT transfers	Percentage of students in pool placed in job/military or school, with bonus for AT/ACT transfers	70 percent	.30
Average Wage	Students placed in a job/military	Average wage	Model-based	.111	Students placed in a job/military	Average wage	Model-based	.08	Students placed in a job/military	Average wage	Model-based	.10
Quality Placement/ Job Training Match Rate	**Vocational completers with a placement record and those with a record that was due but not received**	Percentage placed in a job-training match, with bonus for students placed in college or AT/ACT transfers	51 percent	.111	**All job/military completers**	Percentage placed in a job-training match **(no bonus for students placed in college or ACT)**	42 percent	.08	**All job/military completers**	Percentage placed in a job-training match **(no bonus for students placed in college or ACT)**	50 percent	.10
Full-Time	NA	NA	NA	NA	Students placed in a job/military	Percentage of students placed who are placed full-time	70 percent	.08	Students placed in a job/military	Percentage of students placed who are placed full-time	80 percent	.10
Quality/ Compliance												
ARPA Rating	NA	Regional office rating of center quality/ compliance	100	.333	NA	Regional office rating of center quality/ compliance	100	.20	NA	Regional office rating of center quality/ compliance	NA	0

Note: Reproduced from Schochet and Burghardt (2008). NA indicates not applicable or no change. Bold type shows elements that changed.

[a]Pool of students is the group included in the denominator of the measure.

[b]Standard is the target that centers are expect to meet.

[c]Weight is the share of the individual outcome measure in each center's overall performance score.

Table A.2. Summary Statistics for Unadjusted Performance Measures and Center-Level Impact Estimates, Sample Restricted to Centers with at Least 100 Observations

	Min	1st Quartile	Mean	Median	3rd Quartile	Max	Standard Deviation
Performance Measures (Multiyear Averages)							
Overall	0.92	1.01	1.08	1.08	1.14	1.27	0.09
Reading Gains	0.59	0.92	1.14	1.18	1.40	1.64	0.28
Math Gains	0.59	1.00	1.16	1.20	1.32	1.52	0.23
GED Rate	0.66	0.86	1.02	1.01	1.13	1.58	0.20
Vocational Completion Rate	0.67	0.95	1.07	1.07	1.18	1.40	0.17
Placement Rate	0.90	1.05	1.11	1.12	1.17	1.22	0.08
Average Wage	0.95	0.99	1.02	1.01	1.05	1.16	0.05
Quality Placement	0.98	1.17	1.26	1.24	1.37	1.51	0.14
Full-Time	0.99	1.07	1.09	1.09	1.13	1.19	0.05
ARPA Rating	0.85	0.93	0.99	1.00	1.05	1.10	0.08
Center-Level Impact Estimates							
Any Educational Services[a]	2.4	16.7	26.5	27.1	35.6	61.3	14.0
Hours Educational Services	124	679	952	988	1243	1709	362
GED Receipt[a]	-8.7	12.7	21.1	21.5	26.8	63.0	12.9
Vocational Certificate Receipt[a]	-1.5	23.0	29.5	29.6	38.1	48.8	10.8
Arrested[a]	-37.9	-11.3	-3.8	-1.5	3.8	12.5	11.3
1997 Annual Earnings[b]	-5679	-836	408	460	1762	4669	2286
1998 Annual Earnings[b]	-4358	-300	1302	1379	3597	6737	2683

Sample Size = 40 centers.

Sources: Performance measure data, NJCS follow-up surveys.

Notes: Sample restricted to the 40 centers with 100 or more observations in the NJCS follow-up data.

[a] Impacts are measured in percentage points.

[b] Impacts are measured in 1995 dollars.

Table A.3. Summary Statistics for Unadjusted Performance Measures by Year

	1994		1995		1996	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Overall	1.03	0.09	1.11	0.11	1.14	0.09
Reading Gains	1.06	0.31	1.25	0.28	--	--
Math Gains	1.07	0.26	1.33	0.27	--	--
GED Rate	1.00	0.30	1.11	0.32	1.06	0.24
Vocational Completion Rate	1.00	0.16	1.07	0.20	1.20	0.19
Placement Rate	1.06	0.11	1.09	0.11	1.15	0.08
Average Wage	1.01	0.06	1.03	0.06	1.03	0.06
Quality Placement	1.18	0.17	1.32	0.18	1.30	0.16
Full-Time	--	--	1.16	0.07	1.03	0.06
ARPA Rating	0.98	0.08	0.99	0.08	--	--

Sample Size = 100 centers.

Sources: Performance measure data.

Table A.4. Results for Regressions of Selected Performance Measures on Center-Level Baseline Characteristics from the NJCS Baseline Survey and Area Resource File

NJCS Baseline Characteristics	Overall	GED	Vocational Completion	Average Wage	Placement Rate
Demographic Characteristics					
Black					
Hispanic	-0.365***	-0.677**			
	(0.121)	(0.294)			
Other race	-0.240***				
	(0.0684)				
Female				-0.106***	
				(0.0299)	
High School Degree					
GED	-0.489**		-0.663		
	(0.237)		(0.454)		
Vocational Degree					
Ever Had a Full or Part-Time Job		0.758**			
		(0.337)			
Currently Working	0.249**		0.410*		
	(0.117)		(0.228)		
Married and Living Together	-0.816*		-2.675***		
	(0.443)		(0.864)		
Not Married and Living Together					
Separated, Divorced, or Widowed		1.686*			
		(0.906)			
Ever Arrested				0.184***	
				(0.0672)	
Native Language is Spanish	0.624***	1.095**	0.527***		
	(0.204)	(0.473)	(0.136)		
Native Language is Other Than English or Spanish		-0.947***			-0.448***
		(0.224)			(0.0756)
Has Child					
Pregnant				-0.370	
				(0.261)	
Months in Education Programs in Previous Year: 0-6				0.103	
				(0.0663)	
Months in Education Programs in Previous Year: > 6					
Had Physical or Emotional Problem That Limited Work		-1.491**			-0.514**
		(0.671)			(0.203)
Earnings in Past Year $1,000-$4,999					0.357***
					(0.102)

NJCS Baseline Characteristics	Overall	GED	Vocational Completion	Average Wage	Placement Rate
Earnings in Past Year $5,000-9,999					
Earnings in Past Year >= $10,000				0.397***	0.329*
				(0.103)	(0.172)
Age 18-20					
Age > 20					
Highest Grade Completed 9-11				-0.119*	
				(0.0624)	
Highest Grade Completed > 11					
Family on Welfare Some of the Time When Growing Up					
Family on Welfare Most of the Time When Growing Up				-0.326***	
				(0.0875)	
Received Welfare for Part of Past Year				0.467***	
				(0.158)	
Received Welfare All of Past Year				0.121**	
				(0.0578)	
Received Food Stamps for Part of Past Year	0.750***		1.931***		
	(0.257)		(0.481)		
Received Food stamps All of Past Year	-0.133				
	(0.0911)				
Used Marijuana Occasionally During Past Year					-0.259**
					(0.106)
Used Marijuana Frequently During Past Year	-0.567**	-1.633**		-0.484***	-0.933***
	(0.273)	(0.643)		(0.145)	(0.192)
Used Hard Drugs Occasionally During Past Year					
Used Hard Drugs Frequently During Past Year	1.595***		3.423***		
	(0.426)		(0.731)		
Ever in a Drug or Alcohol Treatment Program				0.490***	
				(0.105)	
Rents Housing					-0.150**
					(0.0698)
Owns Housing				0.0814	
				(0.0493)	
In Other Non-public, Non-subsidized Housing					
Worked in Previous year					
Months Employed in Previous Year: 0-3	0.180	0.471			
	(0.132)	(0.300)			
Months Employed in Previous Year: 4-9			-0.590**		
			(0.241)		
Months Employed in Previous Year: 10-12					

NJCS Baseline Characteristics	Overall	GED	Vocational Completion	Average Wage	Placement Rate
Lived in MSA at Time of Application			0.0725	-0.0442***	
			(0.0553)	(0.0156)	
Lived in PMSA at Time of Application					
Occupation: Services					
Occupation: Laborer and Construction					
Occupation: Sales		-0.515			
		(0.339)			
Occupation: Private Household					
Occupation: Mechanics, Repairers, Technicians					-0.427*
					(0.248)
Occupation: Administrative Support and Clerical	0.550*				
	(0.320)				
Occupation: Agriculture, Forestry, Fishing				-0.508***	
				(0.165)	
Occupation: Manufacturing	-1.293**	-3.058**		-0.744**	-2.096***
	(0.553)	(1.303)		(0.306)	(0.442)
Occupation: Other					
Region of Residence at Application: 2		0.320***		-0.0678***	-0.115***
		(0.0819)		(0.0181)	(0.0248)
Region of Residence at Application: 3		0.416***			
		(0.0592)			
Region of Residence at Application: 4	-0.0888***				-0.106***
	(0.0206)				(0.0145)
Region of residence at Application: 5		0.130**		0.0197	
		(0.0607)		(0.0147)	
Region of Residence at Application: 6		0.247***		-0.0393***	-0.173***
		(0.0556)		(0.0112)	(0.0170)
Region of Residence at Application: 7		0.319***			
		(0.0638)			
Region of Residence at Application: 8	0.144***	0.489***			
	(0.0352)	(0.0814)			
Region of Residence at Application: 9		0.414***		0.0504***	
		(0.0663)		(0.0144)	
Local Area Characteristics					
Percent of Population That Is White		0.636*			
		(0.340)			
Percent of Population That is Black					
Percent of Population in Juvenile Institutions	60.75*		167.8**		42.18
	(35.97)		(69.50)		(29.66)
Percent of Families with a Female Head	1.958***	5.959***			

NJCS Baseline Characteristics	Overall	GED	Vocational Completion	Average Wage	Placement Rate
	(0.451)	(1.015)			
Average Household Size	0.129*	0.765***			0.167***
	(0.0690)	(0.164)			(0.0454)
Percent of Population in Urban Areas				0.102***	
				(0.0340)	
Percent of Families in Poverty					
Median Household Income					
Unemployment Rate					
Deaths by Homicide and Legal Intervention (Rate)	-1,007***	-3,338***		-155.2*	
	(296.3)	(864.6)		(87.95)	
Percent of Population That Is Foreign Born					
Total Births	-8.49e-07*				
	(4.72e-07)				
Percent of Births to Teens (< 18 years)					
Percent of Households with Income $5,000-$9,999				3.566***	
				(0.495)	
Percent of Households with Income $10,000-$14,999					
Percent of Households with Income $15,000-$24,999	1.337**				
	(0.582)				
Percent of Households with Income $25,000-$49,999	0.876*				
	(0.467)				
Percent of Households with Income $50,000-$99,999				1.367***	
				(0.214)	
Percent of Households with Income > $100,000					
Constant	-0.0974	-2.960***	0.924***	0.405***	0.808***
	(0.343)	(0.734)	(0.0801)	(0.101)	(0.128)
Sample Size	100	100	100	100	100
R^2 Value	0.662	0.763	0.411	0.791	0.701

Sources: Performance measure data, NJCS baseline survey, 2008 ARF.

Notes: Standard errors in parentheses. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. The covariates were selected using stepwise regression methods using a covariate inclusion p-value cutoff value of 0.20. Binary covariates that were the "left-out" variables in the regressions are not shown.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level..

Table A.5. Results for Regressions of Selected Performance Measures on Center-Level Baseline Characteristics from the ETA-652 Intake Forms and Area Resource File

ETA-652 Baseline Characteristics	Overall	GED	Vocational Completion	Average Wage	Placement Rate
Demographic Characteristics					
Male	-0.0201	0.0727	-0.0774	0.136***	-0.00171
	(0.0888)	(0.215)	(0.199)	(0.0438)	-0.0801
Black	0.327**	0.396	0.554	-0.0125	-0.0624
	(0.162)	(0.391)	(0.362)	(0.0797)	(0.146)
Hispanic	0.269	0.227	0.350	-0.172	-0.192
	(0.220)	(0.534)	(0.494)	(0.109)	(0.199)
American Indian	-0.153	-0.686	-0.252	-0.0208	-0.463**
	(0.208)	(0.505)	(0.467)	(0.103)	(0.188)
Asian	0.660	0.171	0.882	0.279	0.309
	(0.410)	(0.995)	(0.920)	(0.203)	(0.370)
Weeks Since Employed Full-Time: 6-20	-0.0482	1.023	-0.0582	-0.0400	-0.442
	(0.396)	(0.959)	(0.887)	(0.195)	(0.357)
Weeks Since Employed Full-time: >20	-0.0428	1.124	-0.281	-0.141	-0.682*
	(0.419)	(1.015)	(0.938)	(0.207)	(0.378)
Weeks Since Employed Full-Time: Not Applicable	0.219	0.289	0.280	0.0203	-0.131
	(0.312)	(0.756)	(0.699)	(0.154)	(0.281)
Hourly Wage $4.50-$5.50	-0.0160	-0.864	0.0773	0.177	0.107
	(0.283)	(0.685)	(0.633)	(0.139)	(0.255)
Hourly Wage > $5.50	0.0578	-0.718	-0.177	0.478***	0.239
	(0.293)	(0.710)	(0.656)	(0.145)	(0.264)
Hourly Wage Not Applicable	-0.216	-0.519	-0.391	0.0760	0.0327
	(0.191)	(0.462)	(0.428)	(0.0941)	(0.172)
Lived in City of >50,000 People At Application	-0.190	0.133	-0.344	-0.0846	-0.0861
	(0.119)	(0.288)	(0.266)	(0.0587)	(0.107)
Number of Dependents 1-2	0.0809	-0.152	0.246	-0.147	0.197
	(0.469)	(1.137)	(1.052)	(0.232)	(0.423)
Number of Dependents > 2	0.630	1.021	1.459	0.516	-0.318
	(0.739)	(1.791)	(1.656)	(0.365)	(0.667)
Months Out of School 13-24	-0.481	0.201	-1.036	-0.310*	-0.649**
	(0.323)	(0.784)	(0.725)	(0.160)	(0.292)
Months Out of School > 24	0.119	0.294	0.285	-0.0797	-0.0445
	(0.328)	(0.795)	(0.735)	(0.162)	(0.296)
Months Out of School Not Applicable	-0.490*	-0.496	-0.950	-0.0777	-0.343
	(0.271)	(0.657)	(0.607)	(0.134)	(0.245)
Estimated Annual Income $401-$6,528	-1.267*	-0.934	-1.939	-0.168	-0.304
	(0.640)	(1.552)	(1.435)	(0.316)	(0.578)
Estimated Annual Income > $6,529	-1.039	-0.566	-1.925	-0.0788	-0.384
	(0.626)	(1.517)	(1.403)	(0.309)	(0.565)
Estimated Annual Income Not Applicable	-0.936	-0.584	-1.637	-0.0858	-0.383
	(0.621)	(1.504)	(1.391)	(0.306)	(0.560)
Military Service	1.013	0.121	1.748	0.500	0.980

ETA-652 Baseline Characteristics	Overall	GED	Vocational Completion	Average Wage	Placement Rate
	(0.806)	(1.953)	(1.806)	(0.398)	(0.727)
Legal Resident	0.264	-0.101	0.607	0.675	-0.287
	(0.836)	(2.027)	(1.874)	(0.413)	(0.755)
Ever Had Serious Illness or Injury	1.228	-1.662	2.934*	0.431	0.622
	(0.777)	(1.884)	(1.742)	(0.384)	(0.701)
Under Care of Health Care Provider in Past Year	0.395	0.251	0.437	-0.0220	0.262
	(0.443)	(1.074)	(0.993)	(0.219)	(0.400)
Currently Being Treated for Health Condition	-0.886	0.426	-1.735	-0.618**	-0.0112
	(0.582)	(1.412)	(1.305)	(0.287)	(0.526)
Currently Covered by Health Insurance or Medicaid	0.172	0.334	0.412	0.119*	0.182
	(0.137)	(0.333)	(0.308)	(0.0678)	(0.124)
Has a Child Care Plan	0.134	0.178	0.129	0.342**	0.163
	(0.342)	(0.829)	(0.767)	(0.169)	(0.309)
Family Receiving Public Assistance	-0.471**	-0.258	-0.946**	-0.0741	-0.110
	(0.198)	(0.481)	(0.445)	(0.0980)	(0.179)
Head of Family	-0.471***	-1.060***	-0.984***	0.0276	-0.190
	(0.158)	(0.382)	(0.353)	(0.0778)	(0.142)
Ever Convicted or Adjudged Delinquent	-0.125	0.0155	-0.234	0.121	-0.251
	(0.306)	(0.742)	(0.686)	(0.151)	(0.276)
Highest Grade Completed 9-11	-0.178	-0.511	-0.490	0.0520	-0.0503
	(0.227)	(0.550)	(0.509)	(0.112)	(0.205)
Highest Grade Completed > 11	0.0395	0.276	0.0343	0.232	-0.129
	(0.318)	(0.772)	(0.714)	(0.157)	(0.287)
Age 18-20	-0.101	-0.456	-0.340	0.00664	0.222
	(0.206)	(0.500)	(0.463)	(0.102)	(0.186)
Age > 20	-0.0586	-0.335	-0.139	-0.183	0.0506
	(0.361)	(0.874)	(0.808)	(0.178)	(0.325)
Local Area Characteristics					
Percent of Population That Is White	0.305	-0.102	0.342	-0.238	-0.674*
	(0.408)	(0.990)	(0.915)	(0.202)	(0.369)
Percent of Population That Is Black	0.00217	-0.557	-0.307	-0.284	-0.521
	(0.698)	(1.692)	(1.564)	(0.344)	(0.630)
Percent of Population In Juvenile Institutions	60.78	-105.5	104.2	19.53	18.96
	(54.03)	(131.0)	(121.1)	(26.67)	(48.77)
Percent of Families with a Female Head	-0.642	-0.303	0.204	0.625	0.697
	(1.581)	(3.832)	(3.543)	(0.780)	(1.427)
Average Household Size	0.0794	1.020**	0.114	-0.0194	0.0845
	(0.171)	(0.413)	(0.382)	(0.0842)	(0.154)
Percent of Population in Urban Areas	0.158	0.114	0.315	0.137	0.0643
	(0.210)	(0.509)	(0.471)	(0.104)	(0.190)
Percent of Families in Poverty	4.147*	4.954	7.246	0.358	0.262
	(2.061)	(4.996)	(4.620)	(1.017)	(1.860)
Median Household Income	-2.04e-05	7.37e-06	-4.58e-05	8.33e-06	1.41e-06

ETA-652 Baseline Characteristics	Overall	GED	Vocational Completion	Average Wage	Placement Rate
	(1.53e-05)	(3.70e-05)	(3.42e-05)	(7.54e-06)	(1.38e-05)
Unemployment Rate	-0.439	2.084	-4.003	-0.457	-0.600
	(1.840)	(4.461)	(4.126)	(0.908)	(1.661)
Deaths by Homicide and Legal Intervention (Rate)	871.6	3,526*	181.2	-602.3	-261.0
	(753.2)	(1,826)	(1,688)	(371.8)	(679.9)
Percent of Population That Is Foreign Born	-0.0554	-0.0574	-0.118	0.0472*	0.0124
	(0.0479)	(0.116)	(0.107)	(0.0236)	(0.0432)
Total Births	-1.23E-06	-3.49e-06	-1.74e-06	1.80e-07	-8.05e-07
	(8.94e-07)	(2.17e-06)	(2.00e-06)	(4.42e-07)	(8.07e-07)
Percent of Births to Teens (< 18 years)	-3.324	-12.91*	-2.313	2.606*	2.458
	(2.948)	(7.146)	(6.609)	(1.455)	(2.661)
Percent of Households with Income $5,000-$9,999	11.57***	30.51***	17.71*	2.784	2.953
	(3.993)	(9.678)	(8.950)	(1.971)	(3.604)
Percent of Households with Income $10,000-$14,999	5.093	-3.494	8.940	-1.092	7.184
	(6.438)	(15.61)	(14.43)	(3.178)	(5.811)
Percent of Households with Income $15,000-$24,999	5.730	20.46*	13.64	1.309	1.398
	(4.224)	(10.24)	(9.468)	(2.085)	(3.812)
Percent of Households with Income $25,000-$49,999	12.98***	20.31**	19.43**	1.081	4.775
	(3.688)	(8.940)	(8.267)	(1.821)	(3.329)
Percent of Households with Income $50,000-$99,999	3.451	3.328	7.916	0.208	2.261
	(2.804)	(6.796)	(6.285)	(1.384)	(2.531)
Percent of Households with Income > $100,000	17.12***	33.34***	31.95***	0.486	3.280
	(4.820)	(11.68)	(10.81)	(2.380)	(4.351)
Constant	-6.210**	-15.97**	-10.71*	-0.889	-1.284
	(2.839)	(6.881)	(6.363)	(1.401)	(2.562)
Sample Size	100	100	100	100	100
R^2 Value	0.721	0.796	0.581	0.836	0.758

Sources: Performance measure data, ETA-652 intake form, 2008 ARF.

Notes: Standard errors in parentheses. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. Binary covariates that were the "left-out" variables in the regressions are not shown.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

Table A.6. Unadjusted Center Performance Tercile and Adjusted Center Performance Tercile (including Center Characteristics), Overall Three-Year Average Rating

	NJCS- and Center Characteristics-Adjusted Performance Tercile			ETA-652- and Center Characteristics-Adjusted Performance Tercile		
Unadjusted Performance Tercile	Low	Medium	High	Low	Medium	High
Low	21	7	5	19	10	4
Medium	7	15	11	7	13	13
High	5	11	18	7	10	17
Number of Centers	33	33	34	33	33	34

Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, center characteristics, 2008 ARF.

Notes: Table shows terciles of the three-year average overall performance rating. NJCS-adjusted and ETA-652-adjusted performance terciles are terciles based on adjustments that also include local area characteristics (from the 2008 ARF) and center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight.

Table A.7. Standard Errors of Impact Estimates, by Observations per Center

	Observations Per Center				
Outcome for Impact Estimate	< 50	50-99	100-149	150-199	>200
Any Educational Services[a]	18.1	16.0	11.4	8.7	7.7
Hours of Educational Services	665	470	379	279	249
GED Receipt[a]	23.0	19.0	14.8	11.2	10.0
Vocational Certificate Receipt[a]	18.0	14.6	11.6	8.4	7.8
Ever Arrested[a]	20.6	16.9	12.2	9.6	8.1
1997 Annual Earnings[b]	3,497	3,177	2,301	1,670	1,536
1998 Annual Earnings[b]	3,990	3,420	2,654	1,866	1,800
Number of Centers	18	42	26	4	10

Sources: NJCS follow-up surveys.

Notes: Table shows the average standard error across centers in each category. Impacts are calculated using the follow-up weight and are adjusted for differences in participation across research groups. Groups are based on observations at follow-up.

[a] Impacts are measured in percentage points.

[b] Impacts are measured in 1995 dollars.

Table A.8. Correlations Between Unadjusted Performance Measures and Center-Level Treatment Participant Outcomes

Unadjusted Performance Measure	Outcome for Treatment Participants						
	Hours of Educational Services	GED Receipt	Vocational Certificate Receipt	Ever Arrested	1997 Annual Earnings	1998 Annual Earnings	Length of Stay in Job Corps Center
Overall	0.41***	0.34***	0.38***	-0.17	-0.02	0.11	0.45***
Reading Gains	0.28***	0.17*	0.21**	-0.11	-0.09	-0.04	0.28***
Math Gains	0.25**	0.27***	0.32***	-0.06	0.00	0.01	0.32***
GED Rate	0.43***	0.26***	0.17*	-0.31***	-0.27***	-0.21**	0.38***
Vocational Completion Rate	0.32***	0.27***	0.33***	-0.07	-0.06	0.09	0.35***
Placement Rate	0.21**	0.33***	0.33***	-0.15	0.26***	0.35***	0.33***
Average Wage	-0.02	0.23**	0.07	0.15	0.40***	0.46***	0.09
Quality Placement	0.27***	0.21**	0.29***	-0.05	0.19*	0.27***	0.34***
Full-Time	-0.25**	-0.23**	-0.17*	0.14	0.20**	0.15	-0.17*
ARPA Rating	0.44***	0.30***	0.47***	-0.19*	-0.02	0.08	0.42***

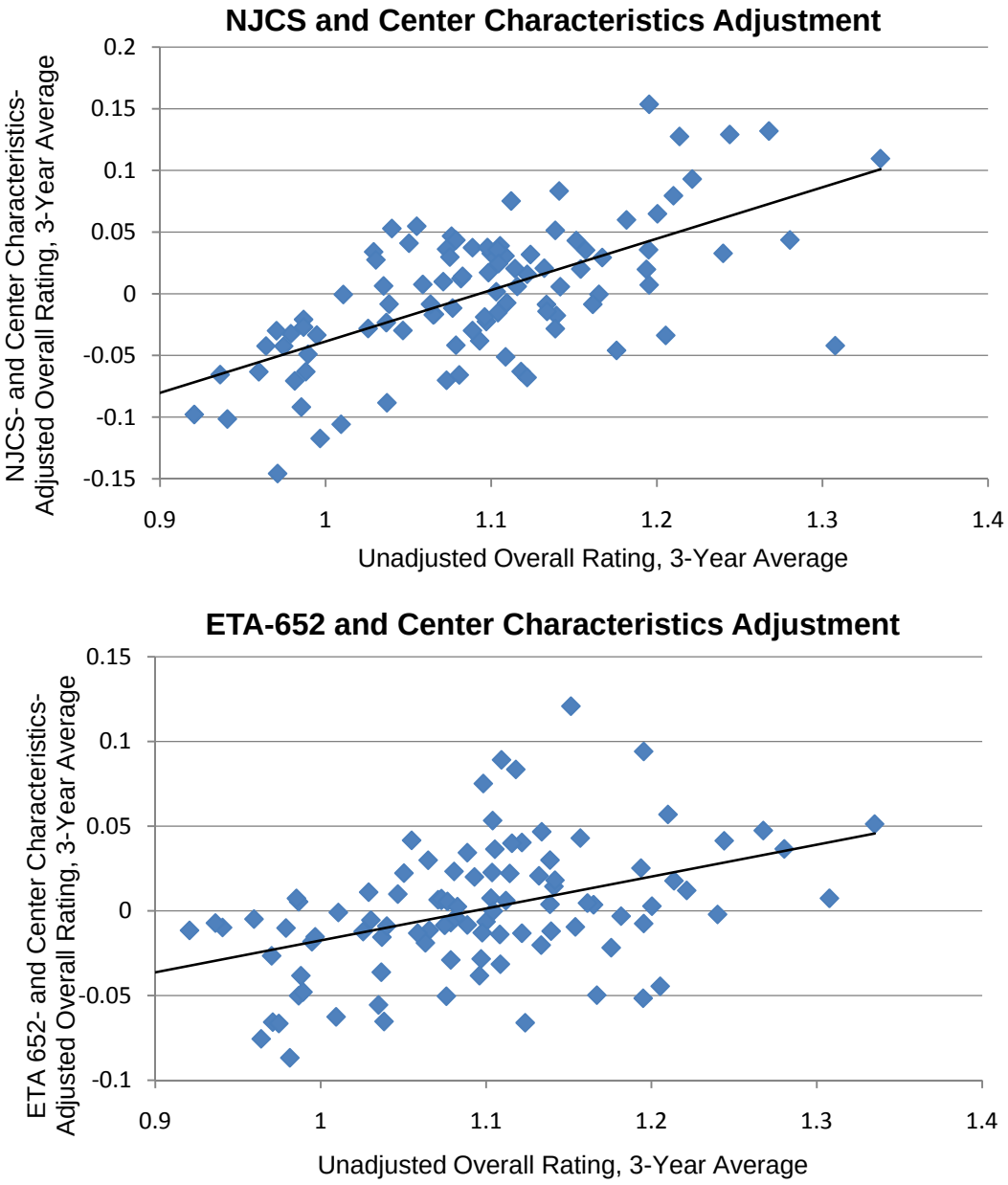
Sample size = 100 centers.

Sources: NJCS baseline survey, NJCS follow-up surveys.

Notes: "Any educational services" is excluded as an outcome because 100 percent of treatment group participants received educational services (by definition, because they participated in Job Corps).

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

Figure A.1. Unadjusted and Adjusted Center Performance (including Center Characteristics), Three-Year Average Overall Rating



Sample size = 100 centers.

Sources: Performance measure data, NJCS baseline survey, ETA-652 intake form, center characteristics, 2008 ARF.

Notes: NJCS-adjusted and ETA-652-adjusted ratings are based on adjustments that also include local area characteristics (from the 2008 ARF) and center characteristics. All centers are weighted equally; when constructing center-level averages, baseline characteristics are weighted using the baseline weight. In both graphs, the slopes are statistically significant at the 1 percent level.

www.mathematica-mpr.com

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC