Health Profession Opportunity Grants (HPOG) Impact Study Amendment to the Technical Supplement to the Evaluation Design Report

OPRE Report 2017-07

January 2017

Health Profession Opportunity Grants (HPOG) Impact Study: Amendment to the Technical Supplement to the Evaluation Design Report

OPRE Report 2017-07

January 2017

Eleanor L. Harvill, Shawn Moulton and Laura R. Peck, Abt Associates, 4550 Montgomery Avenue, Suite 800 North, Bethesda, MD 20814

Submitted to:
Hilary Forster, Contracting Officer's Technical Representative
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract No. HHSP23320095624WC, Task Order HHSP23337012T

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at www.acf.hhs.gov/opre.

Project Director: Gretchen Locke
Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

Overview

Randomized experiments provide researchers with a powerful method for understanding a program's effectiveness. Once they know the direction (favorable or unfavorable) and magnitude (small or large) of a program's impact, the next question is *why* the program produced its effect. Programs that operate in many locations—and the multi-site evaluations that accompany them—offer an opportunity to "get inside the black box" and explore that question. That is, what is it about how the program is configured and implemented that leads its impact to vary?

In the *Technical Supplement to the Evaluation Design Report: Impact Analysis Plan* (OPRE Report No. 2015-80; Harvill, Moulton & Peck, 2015), we described an approach to analyzing the influence of program characteristics (which include program components, implementation features, participant composition, and local context measures) on average impacts of the Health Profession Opportunity Grants (HPOG) program. That material appeared in the *Analysis Plan*'s Chapter 6, and its purpose was to explain how the evaluation would answer its *Research Question 3: Which locally adopted program components influence average impacts?*

Since then, the study team had the opportunity to "pre-test" the analytic method. In doing so, we learned that the approach did not improve upon more common practice (see Bell, Harvill, Moulton & Peck, 2017). In response, the team is revising its analytic methods, and this document should be thought of as replacing the previously stated plans.

Primary Research Questions

One of the HPOG Impact Study's main research questions is, "Which locally adopted program components influence average impacts?" This Amendment documents the planned approach for answering that question.

Purpose

This *Amendment to the Technical Supplement to the Evaluation Design Report* describes the HPOG Impact Team's plans for estimating the influence of locally adopted program components and implementation features on average impacts. Section 1 describes the plan for estimating the impact of select randomized program components. Section 2 discusses the plan to exploit programs' administrative division-level (natural, not randomly induced) variation in implementation features and program-level variation in program components to estimate the relationship between these and impact magnitude. Section 3 describes the plan for selecting which program components, implementation features, participant composition, and local context measures will be included in the model relating those measures to impact magnitude.

Key Findings & Highlights

In response to the limited number of variables we can include at the division-, program-, and local context-level, we use a combination of theory and an empirical approach to select which program components, implementation features, participant composition measures, and local context measures are included in the model relating those measures to impact magnitude. We first identified lists of candidate variables based on our expectations regarding their relationship to the effectiveness of the program. We then prioritized all candidate measures based on their theorized ability to produce policy relevant findings. We use a combination of theory and empirically-based approaches to define a number of candidate versions of the model specification. We plan to use Akaike Information Criterion (AIC) to

determine which of the candidate models to use as the full specification of the model used to report findings.

Methods

We plan to test experimentally the impact of three promising program enhancements: facilitated peer support groups, emergency assistance, and non-cash incentives. Randomly assigning individuals to two variants of the program within designated program sites allows for the comparison of different programmatic scenarios: one in which the program adopts the component of interest, and an alternative where the program does not adopt the component of interest. The contrast in outcomes between these two treatment groups estimates the contribution of the component as an add-on to the main program. That is, it shows the difference in impact that adding the component as an enhancement causes, given the already-existing features of the program.

The study will also take advantage of the naturally occurring variation in the specific services offered by programs (program components) and in how these services are delivered (implementation features) to extend findings about how these characteristics of the intervention may influence impacts. We exploit natural (i.e., non-experimental) cross-division and cross-program variation in these measures to produce non-experimental estimates of the relationship between these program characteristics and impact magnitude. This document explains the approach for executing this analysis.

Revised Plan for Analyzing the Influence of Program Characteristics on Average Impacts

In the *Technical Supplement to the Evaluation Design Report: Impact Analysis Plan* (OPRE Report No. 2015-80; Harvill, Moulton & Peck, 2015), we described an approach to analyzing the influence of program characteristics (which include program components, implementation features, participant composition, and local context measures) on average impacts of the Health Profession Opportunity Grants (HPOG) program. That material appeared in the *Analysis Plan*'s Chapter 6, and its purpose was to explain how the evaluation would answer its *Research Question 3: Which locally adopted program components influence average impacts?*

Since then, the study team had the opportunity to "pre-test" the analytic method. In doing so, we learned that the approach did not improve upon more common practice (see Bell, Harvill, Moulton & Peck, 2017). In response, the team is revising its analytic methods, and this document should be thought of as replacing the previously stated plans.

Section 1 describes the plan for estimating the impact of select randomized program components. As fully elaborated in the study's prior documents (Harvill, Moulton & Peck, 2015; Peck et al., 2014), randomly assigning individuals to two variants of the program within designated program sites allows for the comparison of different programmatic scenarios: one in which the program adopts the component of interest, and an alternative where the program does not adopt the component of interest. The contrast in outcomes between these two treatment groups estimates the contribution of the component as an add-on to the main program. That is, it shows the difference in impact that adding the component as an enhancement causes, given the already-existing features of the program. This is the best information for deciding whether to include the selected component as part of the standard program model going forward.

Section 2 discusses the plan to exploit division-level (natural, not randomly induced) variation in implementation features and program-level variation in program components to estimate the relationship between these and impact magnitude. The *Analysis Plan*'s Exhibit 5.2 lists the notation used in these models, and we replicate it here as Exhibit 2 because it may help the reader to understand the analytic model terms used. The analysis will be conducted for all confirmatory outcomes, as described in the *Analysis Plan*'s Section 2.2 (Harvill, Moulton & Peck, 2015). In contrast to the study's main impact analyses, these analyses are non-experimental in nature because they rely on non-randomly occurring variation across programs and divisions.

Section 3 describes the plan for selecting which program components, implementation features, participant composition, and local context measures will be included in the model relating those measures to impact magnitude. Finally, we provide a sample table shell that can be used for reporting results.

1. Analysis of Randomly Assigned Program Enhancements

In this section, we describe the plan to test experimentally the impact of three promising program enhancements on all confirmatory and secondary outcomes: facilitated peer support groups, emergency assistance, and non-cash incentives. HPOG staff and program participants in programs with a strong peer support component have noted that the support and associated accountability is considered to be one of the most important program elements. Program staff cite unanticipated financial need as a major reason for program dropout, and believe that easier access to emergency funds could buffer participants in times of crisis and improve program retention and completion. Non-cash incentives may also lead to improved participant outcomes by motivating desirable in- and out-of-program behaviors. For example, in a job

retention and advancement program, a results-based incentive might reward those individuals who stay employed for six months, while a behavior-based program might reward individuals who achieve perfect attendance.

Exhibit 1 shows the grantees in which enhancement components were randomly assigned, with final sample sizes. In the exhibit, "NV" refers to "natural variation," indicating whether the designated program component existed naturally within the grantee before it was added as a randomized-to enhancement in some places.

Exhibit 1: Grantee with Experimental Tests of Enhancement Components, by Type

HPOG-Impact Grantee	Peer Support	Emergency Assistance	Non-Cash Incentives
Bergen Community College		(9 Programs)[a] T=490 E=357	(Essex CC)[b] T=195 E=55
Eastern Gateway Community College	NV	NV	NV
Kansas Department of Commerce			
Schenectady County Community College		NV	NV
New Hampshire Office of Minority Health	T=256 E=218	NV	
Milwaukee Area WIB			
South Carolina Department of Social Services			T=201 E=125
Buffalo and Erie County WDC	T= 357 E=60	NV	
Gateway Community and Technical College (KY)		NV	T=118 E=65
Central Community College			NV
Suffolk County Department of Labor		NV	T=261 E=92
Pensacola State College		NV	
WIB SDA-83 Inc. (LA)			
Research Foundation of CUNY-Hostos Comm. Coll.		T=276 E=196	
Will County WIB		(some)[c] NV	
Full Employment Council	NV	T=144 E=122	NV
Central Susquehanna Intermediate Unit		NV	
The WorkPlace	T=158 E=112	NV	
Alamo Comm. Coll. District and Univ. Health System			T= 115 E=62
Edmonds Community College		NV	
HPOG/PACE Grantee	**Peer Support**	**Emergency Assistance**	**Non-Cash Incentives**
Pima County Community College District			
San Diego Workforce Partnership		NV	
Workforce Dev. Council of Seattle-King County		NV	

Source: HPOG-Impact Evaluation Design Implementation Plans.
Notes: "T" refers to the grantee's treatment group. "E" refers to the grantee's enhanced treatment group. Black cells indicate that a sufficient contrast exists and the grantee is implementing the enhancement for an experimental test of its effectiveness. Gray cells indicate that there is not sufficient contrast ("NV" indicates that these programs might be used to explore the natural variation that exists on this program component). White cells indicate that sufficient contrast exists for such a test, but the grantee is not implementing an enhancement.
[a] Nine HPOG programs within the Bergen Community College grantee are implementing the enhancement.
[b] The Essex Community College program within the Bergen Community College grantee is implementing the enhancement.
[c] There is not a sufficient contrast at a subset of the grantee's programs.

Our ability to obtain statistically significant findings showing that an enhancement component affects impact magnitude, when in fact it does, will be limited by available sample sizes—particularly by the number of programs and divisions that randomly assigned cases to both enhanced and standard programs (plus a control group). A null finding should not be interpreted as evidence of no effect, as an effect too small to detect may exist. While the number of individuals included in the groups, as shown in Exhibit 1,

may be adequate to support reasonable power at that level, limitations at other levels exist for each of the enhancement components, as follows:

- 7 divisions within 3 programs implemented three-arm random assignment with facilitated peer support as their enhancement;
- 15 divisions within 11 programs implemented three-arm random assignment with emergency assistance as their enhancement; and
- 10 divisions within 5 programs implemented three-arm random assignment with non-cash incentives as their enhancement.

1.1 Model Specification

For the experimental analysis of the effect of the selected component, we will use a three-level model to estimate program impacts controlling for program and individual factors. The unit of analysis for level one is the individual sample member; the unit of analysis for level two is the division; and the unit of analysis in level three is the program. Compared to the model in the *Analysis Plan*'s Section 5.5 used to answer *Research Question 1 (What impact do the HPOG programs as a group have on outcomes for participants and their families?)*, to estimate the impact of the enhancement (using three-armed randomization) we include an added incremental impact term for the randomly assigned to enhancement feature. Model terms are defined in Exhibit 2.

Exhibit 2: Definitions of Model Terms

Name	Definition
Outcome and Covariates	
Y_{kji}	The outcome measure for individual *i* from division *j* and program *k*
T_{kji}	The standard HPOG program treatment indicator (1 for those individuals assigned to the standard HPOG treatment; 0 for the control group individuals; this is labelled "T" for "treatment")
E_{kji}	The enhanced HPOG program treatment group indicator (1 for only those individuals assigned to the enhanced HPOG treatment group; 0 otherwise; this is labelled "E" for "enhanced" treatment)
TE_{kji}	The HPOG program treatment group indicator (1 for those individuals assigned to the standard HPOG treatment or enhanced HPOG treatment groups; 0 for the control group individuals; this is labelled "TE" for the combination of standard "treatment" and "enhanced" treatment groups)
IC_{ckji}	Individual baseline characteristic *c* for individual *i* from division *j* and program *k* (grand mean centered), *c* = 1, . . ., *C* (this is labelled "IC" for "individual characteristics")
I_{gkj}	implementation feature *g* for division *j* in program *k* (grand mean centered), *g* = 1, . . ., *G* (these are labelled "I" for "implementation")
$\bar{I}_{gk}$	Implementation feature *g* averaged across divisions within program *k* (grand mean centered)
PC_{dkj}	Participant composition variable *d* for division *j* in program *k* (grand mean centered), *d* = 1, . . ., *D*; this is a division-level aggregation of the individual characteristics (ICs) (these are labelled "PC" for "participant composition")
P_{mk}	Program component *m* for program *k* (grand mean centered), *m* = 1, . . ., *M* including the experimentally varied enhancement components P_{Sk}, P_{Ak}, and P_{Ik} (these are labelled "P" for "program")
LC_{qk}	Local context variable *q* for program *k* (grand mean centered), *q* = 1, . . ., *Q* (these are labelled "LC" for "local context")
F_k	Omitted program-level factor F_k (which could be the aggregation to the program level of an omitted division-level factor) that influences treatment impact magnitudes (this is labelled "F" for "factor")

Name	Definition
Model Coefficients	
α_{kj} (alpha)	The control group mean outcome (counterfactual) in division *j*
α_k	The control group mean outcome (counterfactual) in program *k*
α_0	The grand mean control group outcome
β_{kj} (beta)	The conditional impact of being offered the standard HPOG program for each division *j*
β_k	The conditional impact of being offered the standard HPOG program for each program *k*
β_0	The grand mean impact of the standard HPOG Treatment
δ_c (delta)	The effect of individual characteristic *c* on the mean outcome, *c* = 1, . . ., *C*
γ_c (gamma)	The influence of individual characteristic *c* on impact magnitude, *c* = 1, . . ., *C*
π_{ekj} (pi)	The impact of being offered an enhanced HPOG program that includes component *e* relative to the standard HPOG program for each division; this and the other subscripted "pi"s are program component impacts
π_{ek}	The impact of being offered the enhanced HPOG program, inclusive of component *e*, rather than the standard HPOG program without e, for each program
π_e	The grand mean impact of being offered the enhanced HPOG program inclusive of component *e*, rather than the standard HPOG program without *e*
π_m	The influence of program component *m* on impact magnitude, *m* = 1, . . ., *M*
ζ_q (zeta)	The influence of local context variable *q* on impact magnitude, *q* = 1, . . ., *Q*
κ_q (kappa)	The effect of local context variable *q* on control group mean outcome
φ_g (phi)	The influence of implementation feature *g* on impact magnitude, *g* = 1, . . ., *G*
τ_d (tau)	The influence of participant composition variable *d* on impact magnitude, *d* = 1, . . ., *D*
λ (lambda)	The amount by which a one-unit change in F (an omitted confounder) alters impact magnitude
Error Terms	
ε_{kji} (epsilon)	A random component of the outcome for each individual
v_{kj} (nu)	A random component of control group mean outcome for each division
v_k	A random component of control group mean outcome for each program
u_{kj} (upsilon)	A random component of the standard program impact for each division
u_k	A random component of the standard program impact for each program
ω_{kj} (omega)	A random component of the enhanced program's incremental impact for each division
ω_k	A random component of the enhanced program's incremental impact for each program

The level one regression equation depicted by Equation (1) below uses data from individuals to model the relationship between an outcome *Y* and an overall HPOG treatment indicator, *TE*, (which denotes whether the participant was assigned to either the standard HPOG treatment or enhanced HPOG treatment) and an enhanced treatment indicator, *E*, while controlling for individual characteristics. The equation also includes the conditional control group mean and treatment impact for each division. The conditional impact estimates (β_{kj} and π_{ekj}) and control group means (α_{kj}) for each program provide the dependent

variables for level two of the model, as depicted in Equations (2), (3), and (4). The π_{ekj} parameter varies with the sample of sites being analyzed: sites with random assignment to facilitated peer support ($e = S$), sites with random assignment to emergency assistance ($e = A$), and sites with random assignment to non-cash incentives ($e = I$).

Level One: Individuals

$$Y_{kji} = \alpha_{kj} + \beta_{kj}TE_{kji} + \pi_{ekj}E_{kji} + \sum_c \delta_c IC_{ckji} + \varepsilon_{kji} \quad \text{(eq. 1)}$$

Level Two: Divisions

$$\beta_{kj} = \beta_k + u_{kj} \quad \text{(eq. 2)}$$

$$\pi_{ekj} = \pi_{ek} + \omega_{kj} \quad \text{(eq. 3)}$$

and:

$$\alpha_{kj} = \alpha_k + v_{kj} \quad \text{(eq. 4)}$$

Level Three: Programs

$$\beta_k = \beta_0 + u_k \quad \text{(eq. 5)}$$

$$\pi_{ek} = \pi_e + \omega_k \quad \text{(eq. 6)}$$

and:

$$\alpha_k = \alpha_0 + v_k \quad \text{(eq. 7)}$$

We can simplify the above three-level model by substituting Equations (2) through (7) into Equation (1), which produces the following model:

$$Y_{kji} = \alpha_0 + \beta_0 TE_{kji} + \pi_e E_{kji} + \sum_c \delta_c IC_{ckji} + \{\varepsilon_{kji} + v_k + v_{kj} + u_k TE_{kji} + u_{kj} TE_{kji} + \omega_k E_{kji} + \omega_{kj} E_{kji}\}, \quad \text{(eq. 8)}$$

Here, π_e is the primary coefficient of interest: it provides an estimate of the impact of being offered the enhanced HPOG program relative to the standard HPOG program. Conducting the analysis separately for facilitated peer support enhancement programs, emergency assistance enhancement programs, and non-cash incentives enhancement programs provides the experimental estimates of the contribution of those program components to the overall impact magnitude, the various π_e terms for the selected enhancements. We define these estimates as follows:

- $\hat{\pi}_S^X$ provides an experimental estimate of π_e when data from programs that randomly assign to facilitated peer support are analyzed;
- $\hat{\pi}_A^X$ provides an experimental estimate of π_e when data from programs that randomly assign to emergency assistance are analyzed; and
- $\hat{\pi}_I^X$ provides an estimate of π_e when data from sites that randomly assign to non-cash incentives are analyzed.

We plan to use maximum likelihood procedures (which assume joint normal distributions for the error terms) to estimate the above model. More specifically, we plan to use SAS's proc mixed to estimate the model, including weights to adjust for survey nonresponse.

2. Examining the Role of Non-Randomized Program Characteristics

In addressing *Research Question 3*, our goal is to understand how program components and implementation features influence the magnitude of intervention impacts so that stronger program designs can be developed and adopted in the future. The programs that randomize to three experimental arms provide the best evidence on these questions but only for samples of limited size and only for the three HPOG components being tested experimentally as program enhancements. The study will also take advantage of the naturally occurring variation in the specific services offered by programs (program components) and in how these services are delivered (implementation features) to extend findings about how these characteristics of the intervention may influence impacts.

This section describes our analytic approach to estimating the influence of division-level implementation features and participant composition measures, as well as program-level program component and local context measures on impact magnitude. We exploit natural (i.e., non-experimental) cross-division and cross-program variation in these measures to produce non-experimental estimates of the relationship between these program characteristics and impact magnitude. We are particularly interested in how program components and implementation features relate to impact magnitudes, because these characteristics could be incorporated into future programs. The extent to which impact varies by participant composition and local context is important for understanding how, when, and for whom the program works. Although these latter measures are not elements of the designed intervention, they may be of policy relevance in terms of program targeting. Moreover, it is important to control for these measures because they likely associate with grantees' choices regarding program design and implementation.

2.1 Model Specification

To relate these various program characteristics to impact magnitude, we extend the multi-level model in the *Analysis Plan*'s Section 5.5 (Harvill, Moulton & Peck, 2015) by interacting the treatment indicator with measures of program characteristics. We will conduct this analysis on the study's confirmatory outcome, educational progress. Later analyses may expand the number of outcomes explored in this analysis. We will use a combination of theory and an empirical approach to select which program characteristics are included in the analysis. We plan to estimate the model using the combined sample of all individuals in the standard HPOG treatment group or in the control group across all 23 grantees that are part of the main impact analysis.[1] As in the Section 1 model used to estimate the impact of randomized program enhancements, in this model the unit of analysis at level one is the individual sample member; the unit of analysis at level two is the division; and the unit of analysis at level three is the program. Exhibit 2 defined the terms.

Level 1: Individuals

$$Y_{kji} = \alpha_{kj} + \beta_{kj} T_{kji} + \sum_c \delta_c IC_{ckji} + \varepsilon_{kji} \qquad \text{(eq. 9)}$$

Level 2: Divisions

$$\beta_{kj} = \beta_k + \sum_g \varphi_g I_{gkj} + \sum_d \tau_d PC_{dkj} + u_{kj} \qquad \text{(eq. 10)}$$

[1] As noted, we plan to use maximum likelihood procedures (which assume joint normal distributions for the error terms) to estimate the above model; but if this three-level model fails to converge, then we may alter the specification of the random effects (v_{kj}, v_k).

and:

$$\alpha_{kj} = \alpha_k + v_{kj} \qquad \text{(eq. 11)}$$

Level 3: Programs

$$\beta_k = \beta_0 + \sum_m \pi_m P_{mk} + \sum_q \zeta_q LC_{qk} + u_k \qquad \text{(eq. 12)}$$

and:

$$\alpha_k = \alpha_0 + \sum_q \kappa_q LC_{qk} + v_k \qquad \text{(eq. 13)}$$

Combining the elements of the above three-level model produces the following:

$$Y_{kji} = \alpha_0 + \beta_0 T_{kji} + \sum_c \delta_c IC_{ckji} + \sum_q \kappa_q LC_{qk} + \sum_m \pi_m P_{mk} T_{kji} + \sum_q \zeta_q LC_{qk} T_{kji} + \sum_g \varphi_g I_{gkj} T_{kji} + \sum_d \tau_d PC_{dkj} T_{kji} + \{\varepsilon_{kji} + v_k + v_{kj} + u_k T_{kji} + u_{kj} T_{kji}\} \qquad \text{(eq. 14)}$$

In Equation (14), the local context measures (LC_{qk}), program components (P_{mk}), implementation features (I_{gkj}) and participant composition measures (PC_{dkj}) are all multiplied by the treatment indicator. These interaction terms capture the influence of the measure on impact magnitude. In addition, the local context measures enter the model directly, capturing the influence of the economic environment on control group outcomes. This specification does not allow participant composition measures to affect control group outcomes, assuming that the individual-level characteristics (IC_{ckji}) included in the model are more salient to the outcomes of individuals in the control group (and the aggregate participant composition measures may not directly affect control group outcomes). Measures of program components and implementation features are not allowed to affect control group outcomes, because control group members did not access these services.

The extent to which program components and implementation features relate to impact magnitudes is our primary interest; yet we recognize the importance of controlling for local context and participant composition measures. Local context and participant composition measures may be correlated with how the intervention is designed and implemented (e.g., which program components are offered) and may also affect impact magnitude (e.g., via the availability of jobs or through peer effects). In the equations above, the coefficients π_m for $m = 1,\ldots, M$ capture the relationship between impact magnitude and program components. The coefficients φ_g for $g = 1,\ldots, G$ capture the relationship between implementation features and impact magnitudes. Because estimates of these parameters are identified by the naturally occurring (i.e., non-randomized) variation in program components and implementation features, these estimates are non-experimental. As such, they will be interpreted accordingly, with appropriate caveats related to the potential for omitted variable bias to affect the estimates of interest.

3. Selecting Covariates for Inclusion in Model Relating Non-Randomized Variables to Impact Magnitude

In response to the limited number of variables we can include at the division-, program-, and local context-level, we use a combination of theory and an empirical approach to select which program components, implementation features, participant composition measures, and local context measures are included in the Equation (14) model.[2] We first identified lists of candidate variables based on our

[2] This analysis focuses on measures defined at the division-, program- and local context-levels, and therefore the sample sizes at each of these levels determine the degrees of freedom available when testing whether the coefficient on a given program characteristic is statistically different from zero.

expectations regarding their relationship to the effectiveness of the program.[3] There are four types of candidate variables as follows:

- **Program components** defined at the program-level, as described in the *Analysis Plan*'s Exhibit 4.1;
- **Implementation features** defined at the division-level, as described in the *Analysis Plan*'s Exhibit 4.2;
- **Participant composition** measures, which are division-level aggregations of the individual-level baseline characteristics listed in the *Analysis Plan*'s Exhibits 2.1–2.5; and
- **Local context** measures defined at the local context-level, as described in the *Analysis Plan*'s Exhibit 4.3.[4]

3.1 Priority Level of Candidate Measures

Exhibit 3 denotes the priority level of each candidate measure. We prioritized all candidate measures based on their theorized ability to produce policy relevant findings (based on both the literature as documented in Chapter 4 of the *Analysis Plan* as well as the expertise of Abt staff); the amount of variation in the candidate measures across divisions and programs; and missing data rates. To ensure that the Equation (14) model produces findings that are policy relevant, we plan to automatically include all *priority 1* candidate measures in the analysis.

Exhibit 3: Priority Level of Candidate Measures

Priority	Type	Domain	Variable Description
1	Program Component	Presence of Career Pathways Principles	Extent to which available offerings and program content is based on principles of the career pathways framework
3	Program Component	Case Management	Average caseload for FTE (estimated full time equivalent) case managers
1	Program Component	Case Management	Number of services that case managers and counselors deliver that meet the needs of participants
3	Program Component	Comprehensive Services	Access to social and other services: social and other services delivered that meet participants' needs
1	Program Component	Comprehensive Services	Access to and delivery of tuition and other financial services: tuition coverage plus financial services offered that meet participant needs
2	Program Component	Comprehensive Services	Access to childcare and transportation: accessibility via public transportation plus childcare and transportation services offered that meet participant needs
3	Program Component	Comprehensive Services	Location of services: number of services co-located with the training site
3	Program Component	Employment Supports	Number of employment supports that are offered that meet participants' needs
3	Program Component	Behavioral Incentives	Non-cash incentives: whether the program provides offer non-cash incentives to participants for achieving program milestone
3	Program Component	Peer Support	Offer of facilitated peer support
3	Program Component	Emergency Assistance	Access to emergency funds to meet needs stemming from imminent eviction from housing, utility shutoff, vehicle repair needs, etc.

[3] Section 4.1 of the *Analysis Plan* discusses our hypotheses regarding these relationships and provides citations to the literature.

[4] Local context measures can be thought of as approximately grantee-level. In Equations (12) and (14) above, the local context variables enter the model at the program-level. However, observations for programs within the same local context area have the same value of these measures, because these programs experience the same economic environment. Appendix A of the *Analysis Plan* provides more detail on these local context measures, which are obtained for programs and grantees from Census and BLS data sources.

Priority	Type	Domain	Variable Description
1	Implementation Feature	Management/Staff Focus	Extent to which program is employment focused
1	Implementation Feature	Management/Staff Focus	Extent to which program is education focused
3	Implementation Feature	Staff Experience	Percentage of management/staff at the division level with at least five years of experience
2	Implementation Feature	Staff Discretion/Autonomy	Staff perception of autonomy, including authority to carry out responsibility, ability to try different techniques, trust in staff professional judgment and not too many rules
3	Participant Composition	*All Domains*	*All Variables*
3	Local Context	*All Domains*	*All Variables*

The analysis's degrees of freedom are limited by the number of observations at the division, program-, and local-context levels. A standard guideline is to include at least five observations per covariate (Green, 1991). Based on this, if we only include division-level measures, we could include at most about 17 measures. This study does not have 17 measures at the division level, and we also want to include program and local context measures in the model for both policy and bias reduction reasons. When we do include program-level and local-context measures as well, then we need to account for the degrees of freedom at each of those levels, where we have fewer observations. For example, a model could accommodate two local context, three program and three division-level variables. Alternatively, we could include only one local context measure in order to accommodate as many as four program-level and five division-level measures, and these latter two levels are those where the policy triggers are. Given these constraints, the approach we describe below suggests various possible model specifications, taking into consideration the degrees of freedom tradeoffs across levels and also desire for policy-relevant information.

3.2 Candidate Model Specifications

To determine which *priority 2* and *priority 3* program component, implementation feature, participant composition, and local context candidate measures are included in the analysis, we use a combination of theory and empirically-based approaches to define a number of various candidate versions of the specification. One of these candidate model specifications will be selected as the full specification that will be used to report findings. Note that *all* candidate model specifications will include an intercept, treatment indicator, individual-level baseline covariates, and random components as depicted in Equation (14). Additionally, all candidate models include the five pre-selected *priority 1* measures. The candidate models differ with respect to which subset of the *priority 2* and *priority 3* program component, implementation feature, participant composition, and local context measures are included.

3.2.1 Theory-Based Candidate Models

We begin by defining candidate models based on theory.[5] Below we describe which *priority 2* and *priority 3* measures are included in the theory-based models. Note that four of the six theory-based

[5] Note that all models (including those based on theory) include an intercept, treatment indicator, individual-level baseline covariates, and random components as depicted in Equation (14), as well as the pre-selected *priority 1* measures.

models include the unemployment rate in the model, emphasizing the hypothesized importance of this local context measure.

- *Candidate Model 1t:* Includes 9 candidate measures:
  - all *priority 1* candidate measures
  - all *priority 2* candidate measures
  - average caseload for FTE case managers
  - unemployment rate

  This candidate model specification is designed to be an approximation of the model used by Bloom, Hill & Riccio (2003) to non-experimentally estimate the relationship between program characteristics and impact magnitude.[6]

- *Candidate Model 2t:* Includes 7 candidate measures:
  - all *priority 1* candidate measures
  - all *priority 2* candidate measures

- *Candidate Model 3t:* Includes 8 candidate measures:
  - all *priority 1* candidate measures
  - all *priority 2* candidate measures
  - unemployment rate

- *Candidate Model 4t:* Includes 10 candidate measures:
  - all *priority 1* candidate measures
  - all *priority 2* candidate measures
  - unemployment rate
  - the following two participant composition measures:
    - Average quarterly wage received during the four quarters prior to the quarter of random assignment (average across all study participants in a given division)
    - Proportion of quarters employed during the four quarters prior to the quarter of random assignment (average across all study participants in a given division)

- *Candidate Model 5t:* Includes 9 candidate measures:
  - all *priority 1* candidate measures
  - all *priority 2* candidate measures
  - number of employment supports that are offered that meet participants' needs
  - unemployment rate

---

[6] The findings reported in Table 8 of Bloom, Hill & Riccio (2003) indicate that the following local program characteristics have a statistically significant predictors of program impacts at the $p<0.10$ significance level: emphasis on moving clients into jobs quickly; emphasis on personalized client attention; basic education service differential; staff caseload size; and the unemployment rate. Whether the program places an emphasis on moving clients into jobs quickly is captured by the following *priority 1* measure: e*xtent to which program is employment focused*. Additionally, emphasis on personalized client attention is captured by the following *priority 1* measure: *number of services that case managers and counselors deliver that meet the needs of participants*. *Candidate Model 1t* does not directly capture the basic education service differential, but does include a measure of the *extent to which available offerings and program content is based on principles of the career pathways framework* (a *priority 1* measure), which captures a variety of educational opportunities available to the client. Finally, the *Candidate Model 1t* explicitly includes the staff caseload size and unemployment rate.

- *Candidate Model 6t:* Includes 9 candidate measures:
  - all *priority 1* candidate measures
  - all *priority 2* candidate measures
  - number of employment supports that are offered that meet participants' needs
  - average caseload for FTE (estimated full time equivalent) case managers

3.2.2 Unconstrained Empirically Derived Candidate Models

Next, we define seven candidate models using an empirical approach as follows:

- *Candidate Model 0:* Includes 5 candidate measures:
  - all *priority 1* candidate measures

- *Candidate Model 1u:* Includes 6 candidate measures:
  - all *priority 1* candidate measures
  - one *priority 2* or *priority 3* candidate measure with the lowest p-value of all remaining candidate measures (each model is estimated separately with one of the remaining candidate measures not already included in the model)

- *Candidate Model 2u:* Includes 7 candidate measures:
  - The 6 covariates included in *Candidate Model 1u*
  - one additional *priority 2* or *priority 3* candidate measure with the lowest p-value of all remaining candidate measures.

- *Candidate Model 3u:* Includes 8 candidate measures:
  - the 7 covariates included in *Candidate Model 2u*
  - one additional *priority 2* or *priority 3* candidate measure with the lowest p-value of all remaining candidate measures

- *Candidate Model 4u:* Includes 9 candidate measures:
  - the 8 covariates included in *Candidate Model 3u*
  - one additional *priority 2* or *priority 3* candidate measure with the lowest p-value of all remaining candidate measures

- *Candidate Model 5u:* Includes 10 candidate measures:
  - the 9 covariates included in *Candidate Model 4u*
  - one additional *priority 2* or *priority 3* candidate measure with the lowest p-value of all remaining candidate measures

- *Candidate Model 6u:* Includes11 candidate measures:
  - the 10 covariates included in *Candidate Model 5u*
  - one additional *priority 2* or *priority 3* candidate measure with the lowest p-value of all remaining candidate measures

3.2.3 Constrained Empirically Derived Candidate Models

We define an additional six candidate models (*Candidate Models 1c–6c*) using the same empirical approach used above to generate *Candidate Models 1u–6u*, but constraining the approach such that no more than two participant composition measures (division level) are selected and no more than one local

context measure is selected for inclusion in each candidate model. The intent here is to constrain the empirical approach such that more covariate slots are available for program component and implementation feature measures, which are of potentially greater policy interest. To the extent that the empirical approach selects candidate models that have no more than two participant composition measures and no more than one local context measure, a subset of *Candidate Models 1c–6c* will be identical to *Candidate Models 1u–6u.*

Exhibit 4 summarizes the models that we propose to analyze in the process of determining which will be the main model used to report findings from this analysis.

Exhibit 4: Candidate Model Specifications

Model	Priority 1 Variables Included	Priority 2 Variables Included	Priority 3 Variables Included	Total Variables Included
Theory-based Models				
1t. BRH replica	ALL	ALL	average caseload unemployment rate	9
2t.	ALL	ALL	NONE	7
3t.	ALL	ALL	unemployment rate	8
4t.	ALL	ALL	unemployment rate aggregate participant wages aggregate prior employment	10
5t.	ALL	ALL	employment supports unemployment rate	9
6t.	ALL	ALL	employment supports average caseload	9
Unconstrained Empirically-derived Models				
0. Base model	ALL	NONE	NONE	5
1u.	ALL	the one measure with the lowest p-value		6
2u.	ALL	the 1u. measure plus the next measure with the lowest p-value		7
3u.	ALL	the 2u. measures plus the next measure with the lowest p-value		8
4u.	ALL	the 3u. measures plus the next measure with the lowest p-value		9
5u.	ALL	the 4u. measures plus the next measure with the lowest p-value		10
6u.	ALL	the 5u. measures plus the next measure with the lowest p-value		11
Constrained Empirically-derived Models				
1c.	ALL	the same as 1u		6
2c.	ALL	the 1u. measure plus the next measure with the lowest p-value, given variable selection constraints[a]		7
3c.	ALL	the 2c. measures plus the next measure with the lowest p-value, given variable selection constraints[a]		8
4c.	ALL	the 3c. measures plus the next measure with the lowest p-value, given variable selection constraints[a]		9
5c.	ALL	the 4c. measures plus the next measure with the lowest p-value, given variable selection constraints[a]		10
6c.	ALL	the 5c. measures plus the next measure with the lowest p-value, given variable selection constraints[a]		11

Notes: See Exhibit 3 for list of variables and priority designation. In model numbers, "t" refers to theory-based, "u" refers to unconstrained empirically-based, and "c" refers to constrained empirically-based.
[a] Selection of *priority 2* and *priority 3* variables is constrained such that no more than one local context and two participant composition measures are included in the model.

3.3 Method for Selecting Candidate Model Used to Report Findings

We plan to use Akaike Information Criterion (AIC) to determine which of the candidate models to use as the full specification of the Equation (14) model used to report findings. Given that some of our proposed models push degrees of freedom limits, should any model fail to converge, it will be dropped from consideration. Among the candidate models, the one with the smallest AIC is considered the best (although the AIC value itself is not meaningful). AIC rewards goodness of fit, but it also includes a penalty that is an increasing function of the number of estimated parameters, which discourages overfitting. When estimating each candidate model, prepackaged SAS commands compute the AIC value used to determine which of the candidate models to use as the final Equation (14) used to report findings.[7]

4. Table Shells for Reporting Findings

Exhibit 5 provides a sample table shell for reporting experimental impacts of HPOG program enhancements. This table is similar to those planned for the study's other impact analyses (e.g., see the *Analysis Plan*'s Exhibit 5.3). Exhibit 6 provides a sample table shell for reporting estimates of the contribution of non-randomized program characteristics to impact magnitude. The Exhibit also presents standard errors of the impact estimates, and the sample size of individuals and clusters.

Exhibit 5: Sample Table Shell—Estimates of the Contribution of Randomized Program Enhancements to Impact Magnitude

	Standard Treatment Group Mean (1)	Enhanced Treatment Group Mean (2)	Impact of Enhancement (3)	Relative Impact (4)
Outcome Domain1				
Measure1	A	B	A-B*	%
FILL IN…				
Outcome Domain2				
Measure2				

Notes:
* Statistically significant, p<0.01
** Statistically significant, p<0.05
*** Statistically significant, p<0.10

[7] One consequence of this model selection strategy is that the AIC criterion may result in a set of independent variables with bias and incorrect variances due to failure to account for the fact that the chosen set of independent variables was one of multiple possible sets of independent variables used during the model selection process. This is known in the literature as "pre-test" bias and, as noted by Kennedy (2003), standard practice is to ignore the pre-test bias problem. As Kennedy (2003, page 221) notes:

> Most econometricians ignore the pre-test bias problem; in fact, few even admit its existence. The main counter-argument to pre-test bias is that without pre-testing we must rely on an assumption concerning what variables are included in the set of independent variables. Is the probability that pre-testing yields an incorrect set of independent variables greater than or less than the probability that the econometrician has selected the "correct" assumption? Pre-testing is simply a means of providing additional evidence to aid the econometrician in selecting the appropriate set of independent variables. So long as the econometrician views this as evidence to be evaluated sensibly in light of other considerations (such as economic theory), rather than as a mechanical procedure, pre-test bias should not be of great concern.

We follow common practice in ignoring pre-test bias, as we rely not only on an empirical approach, but also theory to derive the model used to report findings.

Exhibit 6: Sample Table Shell—Estimates of the Contribution of Non-Randomized Program Characteristics to Impact Magnitude

	Impact (1)	Standard Error (2)	Significance Level (3)
Program Components			
Program Component 1			
Program Component 2			
Implementation Features			
Implementation Feature 1			
Implementation Feature 2			
Local Context Measures			
Local Context Measure 1			
Local Context Measure 2			
Participant Composition Measures			
Participant Composition Measure 1			
Participant Composition Measure 2			
Sample Size			
Individuals			
Divisions			
Programs			

Notes:
* Statistically significant, p<0.01
** Statistically significant, p<0.05
*** Statistically significant, p<0.10

Works Cited

Bell, Stephen H., Eleanor L. Harvill, Shawn R. Moulton, and Laura Peck. (2017). *Using Within-Site Experimental Evidence to Reduce Cross-Site Attributional Bias in Connecting Program Components to Program Impacts*. OPRE Report #2017-13. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Bloom, H.S., Hill, C.J., & Riccio, J.A. (2003). Linking program implementation and effectiveness: Lessons from a pooled sample of welfare-to-work experiments. *Journal of Policy Analysis and Management*, 22, 551–575.

Green, S.B. (1991). How Many Subjects Does It Take To Do A Regression Analysis? *Multivariate Behavioral Research*, 26(3): 499-510.

Harvill, Eleanor L., Shawn Moulton, and Laura R. Peck. (2015). *Health Profession Opportunity Grants Impact Study Technical Supplement to the Evaluation Design Report: Impact Analysis Plan*. OPRE Report #2015-80, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Available at <<http://www.acf.hhs.gov/opre/resource/hpog-impact-study-technical-supplement-to-the-evaluation-design-report-impact-analysis>>

Kennedy, Peter. (2003). *A Guide to Econometrics*. Cambridge, MA: MIT Press.

Peck, Laura R., Alan Werner, Alyssa Rulf Fountain, Jennifer Lewis Buell, Stephen H. Bell, Eleanor Harvill, Hiren Nisar, David Judkins and Gretchen Locke. (2014). *Health Profession Opportunity Grants Impact Study Design Report*. OPRE Report #2014-62. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Available at <<http://www.acf.hhs.gov/opre/resource/health-profession-opportunity-grants-hpog-impact-study-design-report>>