ACT-CCREC Core Research Program: Study Questions and Design

Ty M. Cruce

WP-2015-01
January, 2015

ACT Working Paper Series

ACT working papers document preliminary research. The papers are intended to promote discussion and feedback before formal publication. The research does not necessarily reflect the views of ACT.

ACT®

ACT-CCREC Core Research Program: Study Questions and Design

Abstract

This report provides a non-technical overview of the guiding research questions and research design for the ACT-led core research program conducted on behalf of the GEAR UP College and Career Readiness Evaluation Consortium (CCREC). The core research program is a longitudinal study of the effectiveness of 14 GEAR UP state grants on the academic achievement, college going and college retention of low-income and other at-risk students who are eligible to receive grant-funded services. This research program will follow Consortium students and a matched comparison group of non-participants over time using assessment data from ACT Explore®, ACT Plan® or ACT Aspire®, and The ACT®, and college enrollment data from the National Student Clearinghouse (NSC) to examine differences in the academic growth and educational outcomes of these two groups.

Keywords: GEAR UP, consortium, research design, college readiness, enrollment, retention

Introduction

Gaining Early Awareness and Readiness for Undergraduate Programs (GEAR UP) is a U.S. Department of Education discretionary grant program that focuses on increasing the number of low-income students who enter and successfully complete postsecondary education. The program provides six- or seven-year grants to states and partnerships to fund services such as tutoring, mentoring, financial aid counseling, and academic planning for students at high-poverty middle and high schools. Grantees can elect to offer grant-funded services either to a specific cohort of students followed over the life-cycle of the grant (a cohort model) or to any students in grades 7-12 who are designated in need of those services (a priority model).

Fourteen states that received a GEAR UP state grant during either the 2011 or 2012 award cycle have formed a member consortium known as the College and Career Readiness Evaluation Consortium (hereafter, Consortium).[1] The mission of the Consortium is to provide research and evaluation to inform program implementation and administration and to demonstrate the program's value-added impact and efficacy. To assist with this mission, the Consortium has joined together with the National Council for Community and Education Partnerships (NCCEP) to serve as managing partner, the National Student Clearinghouse (NSC) to construct a multi-state database to support the Consortium's evaluation and research initiatives, and ACT, Inc. to provide additional research support based on its longitudinal assessment system.

As a part of its contribution to the Consortium, ACT is leading two long-term research efforts:

[1] The 14 states comprising the consortium are Arizona, Idaho, Kentucky, Minnesota, Montana, Nevada, New Mexico, North Carolina, Oklahoma, Tennessee, Utah, Washington, Wisconsin, and Wyoming.

1. The core research program, addressed in greater detail in the remainder of this report, is a longitudinal study of the effectiveness of GEAR UP state grants on the academic achievement, college going and college retention of low-income and other academically at-risk students who are eligible to receive grant-funded services. This research program will follow Consortium students over five data collection periods from 8th grade to their second year of college using achievement data from a suite of assessments (ACT Explore®, ACT Plan® or ACT Aspire®, and The ACT®) and college enrollment data from NSC to examine the academic growth and educational outcomes of these students. A matched comparison group of non-participants will also be followed over time using assessment and college enrollment data, and the differences in the educational outcomes of these two groups will be used to estimate the overall effect of GEAR UP state grants on eligible students.
2. The supplemental research program is a longitudinal study of the impact of good educational practices delivered through two commonly-offered GEAR UP-funded services (i.e., tutoring and mentoring) on the academic achievement and non-cognitive outcomes (e.g., academic self-confidence and commitment to school) of GEAR UP participants. Unlike the core research program, the supplemental research program is not intended to address the causal effect of GEAR UP. Instead, the supplemental research program is intended to help consortium member states and other GEAR UP programs better understand the relationships between the structure and quality of services commonly offered by GEAR UP programs and the relationships among service structure and quality, academic achievement, and non-cognitive outcomes.

Guiding Research Questions for the Core Research Program

The general causal question of interest to the core research program is: Are students who are eligible to receive services funded by GEAR UP state grants better off educationally than they would have been in absence of that funding? This broader question will be addressed by answering the following research questions, listed by the five data collection periods.

Base Year (8th Grade)

The research questions for the base year are not intended to address the causal effect of GEAR UP. Rather, they are asked for the purposes of establishing a baseline for estimating the effect of GEAR UP and informing the selection of a matched comparison group of non-participants.

BY.1: To what extent do GEAR UP participants and non-participants differ in their academic achievement levels in the areas of English, mathematics, reading and science as measured by ACT Explore?

BY.2: To what extent do GEAR UP participants and non-participants differ in being on track for readiness as measured by their attainment of the ACT Explore College Readiness Benchmarks in the areas of English, mathematics, reading and science?

BY.3: To what extent do GEAR UP participants and non-participants differ regarding their high school and postsecondary educational plans as self-reported on ACT Explore?

First Follow-up (10th Grade)

F1.1: To what extent do GEAR UP participants and a baseline-matched group of non-participants differ in their academic achievement levels in the areas of English, mathematics, reading and science as measured by ACT Plan or ACT Aspire?

F1.2: To what extent do GEAR UP participants and a baseline-matched group of non-participants differ in being on track for readiness as measured by their attainment of ACT Plan or ACT Aspire College Readiness Benchmarks in the areas of English, mathematics, reading and science?

F1.3: To what extent do GEAR UP participants and a baseline-matched group of non-participants differ regarding their postsecondary educational plans as self-reported on ACT Plan or ACT Aspire?

Second Follow-up (11th/12th Grade)

F2.1: To what extent do GEAR UP participants and a baseline-matched group of non-participants differ in their academic achievement levels in the areas of English, mathematics, reading and science as measured by The ACT?

F2.2: To what extent do GEAR UP participants and a baseline-matched group of non-participants differ in their readiness for college as measured by their attainment of the ACT College Readiness Benchmarks in the areas of English, mathematics, reading and science?

F2.3: To what extent do GEAR UP participants and a baseline-matched group of non-participants differ regarding their postsecondary educational plans as self-reported on The ACT?

Third Follow-up (First Year of College)

F3.1: To what extent do GEAR UP participants and a baseline-matched group of non-participants differ in their likelihood of attending college?

F3.2: Among those students who enrolled in college, to what extent do GEAR UP participants and a baseline-matched group of non-participants differ in the type of college attended (e.g., 2-year vs. 4-year, less selective vs. more selective)?

Fourth Follow-up (Second Year of College)

F4.1: Among those students who were enrolled in college during the third follow-up, to what extent do GEAR UP participants and a baseline-matched group of non-participants differ in their likelihood of being retained (a) at the same college and (b) at any college?

F4.2: Among those students who transferred to a different college between the third and fourth follow-ups, to what extent do GEAR UP participants and a baseline-matched group of non-participants differ in their type of transfer (i.e., vertical, horizontal, or reverse)?

F4.3 Among those students who were not enrolled in college during the third follow-up, to what extent do GEAR UP participants and a baseline-matched group of non-participants differ in their likelihood of attending college during the fourth follow-up (i.e., delayed college entry)?

Causal Model Framework

With the exception of those questions answered by baseline data, the research questions addressed by the core research program are guided by the Rubin Causal Model (Holland, 1986). According to this model, a necessary condition for causality is that an action, such as a treatment or intervention, is applied to a unit, such as an individual person or group of people, at a particular point in time. This condition for causality assumes that at the point in time the treatment is received the unit could have otherwise been subjected to some different action, such as an alternative (or control) treatment. For each of these actions (treatment or control) there is a potential outcome for the unit. The causal effect of the treatment on the unit is defined as the

difference between the potential outcome under the treatment and the potential outcome under the control. For this study, the causal effect of interest is whether or not students who are eligible to receive direct services funded by GEAR UP state grants are better off educationally (e.g., higher academic achievement, higher educational expectations, higher likelihood of attending college, higher likelihood of persisting in college) than they would have been in absence of that funding. These two potential outcomes can be written as follows:

$$\textit{potential outcome} = \begin{cases} \mathrm{Y}_{1i} & \text{if } \mathrm{GEARUP}_i = 1 \\ \mathrm{Y}_{0i} & \text{if } \mathrm{GEARUP}_i = 0 \end{cases}$$

The fundamental problem with inferring a causal effect is that for any particular person we can observe only *one* of the potential outcomes—either under the treatment or under the control. We cannot know what the outcome of GEAR UP participants would have been had they *not* received GEAR UP services. Likewise, we cannot know what the outcome of non-participants would have been had they received GEAR UP services. A common approach to address this missing data problem is to examine the difference in the observed outcomes between students who received the treatment and those who received the control. The relationship between this difference in observed outcomes and the causal effect can be represented as follows:

$$\underbrace{E\left[\mathrm{Y}_i \,\middle|\, \mathrm{GEAR\,UP}_i = 1\right] - E\left[\mathrm{Y}_i \,\middle|\, \mathrm{GEAR\,UP}_i = 0\right]}_{\text{Observed difference in educational outcome}} = \underbrace{E\left[\mathrm{Y}_{1i} \,\middle|\, \mathrm{GEAR\,UP}_i = 1\right] - E\left[\mathrm{Y}_{0i} \,\middle|\, \mathrm{GEAR\,UP}_i = 1\right]}_{\text{average GEARUP effect on the treated}}$$
$$+ \underbrace{E\left[\mathrm{Y}_{0i} \,\middle|\, \mathrm{GEAR\,UP}_i = 1\right] - E\left[\mathrm{Y}_{0i} \,\middle|\, \mathrm{GEAR\,UP}_i = 0\right]}_{\text{selection bias}}$$

such that the observed difference in the average outcome between GEAR UP participants and non-participants is equal to the average effect of GEAR UP on those who participated plus some degree of selection bias. Selection bias, defined below, is represented as the difference in the average outcome under conditions of the control (i.e., the absence of GEAR UP services) between GEAR UP participants and non-participants. If selection bias is negative, GEAR UP participants would have fared worse than non-participants in absence of the funding; this means that the observed difference in the outcome between the two groups *underestimates* the effect of GEAR UP. Conversely, if selection bias is positive, GEAR UP participants would have fared better than non-participants in absence of the funding; this means that the observed difference in the outcome between the two groups *overestimates* the effect of GEAR UP. If selection bias is zero, GEAR UP participants would have fared the same as non-participants in absence of the funding, and the observed difference is equivalent to the effect of GEAR UP on its participants.

Regardless of whether there is an underestimation or an overestimation of the treatment effect, selection bias is present in the observed difference between participants and non-participants because assignment to the treatment is not independent of the outcome. In other words, students are selected (or select themselves) into the treatment because of the potential outcome (whether realized or not) that they associate with that treatment. In the case of GEAR UP, specific schools and students are selected to receive these grant-funded services because they are deemed in need of and likely to benefit from the services.

One way to ensure independence between treatment assignment and the outcome—and thus solve the selection bias problem—is to randomly assign units to the treatment. This type of assignment found in a randomized controlled trial provides some assurance that the control and treatment groups are balanced—meaning that the characteristics of these two groups are

equivalent—prior to the treatment. Since schools and/or students participating in the GEAR UP program were not selected at random, a goal of the core research program is to create a research design that best approximates a randomized controlled trial in order to achieve equivalent groups prior to treatment. This process will minimize selection bias such that the observed differences in educational outcomes between GEAR UP participants and non-participants will be better estimates of the causal effects of grant-funded services on the participants.

Propensity Score Methods

Among the summative evaluation design approaches listed in the *FY 2011 Application for Grants under the Gaining Early Awareness and Readiness for Undergraduate Programs (GEAR UP) State Grants* (U.S. Department of Education Office of Postsecondary Education, 2011) is the carefully matched comparison group design. As a quasi-experimental research design that attempts to approximate a randomized controlled trial, this approach matches participants with non-participants

> "based on key characteristics that are thought to be related to the outcome. These characteristics may include, but are not limited to: (1) prior test scores and other measures of academic achievement (preferably, the same measures that the study will use to evaluate outcomes for the two groups); (2) demographic characteristics, such as age, disability status, gender, English proficiency, ethnicity, poverty level, parents' educational attainment, and single- or two-parent family background; (3) the time period in which the two groups are studied (e.g., the two groups are children entering kindergarten in the same year as opposed to sequential years); and (4) methods used to

collect outcome data (e.g., the same test of reading skills administered in the same way to both groups)" (p. 85).

Propensity score methods will be used to "match" GEAR UP participants with non-participants for all questions of a causal nature that are addressed by the core research program. A propensity score is defined as the unit's probability of receiving the treatment (versus the control) conditional on observed characteristics of the unit (Austin, 2011). Propensity scores are produced by a statistical model—commonly, logistic regression—that predicts membership in the treatment or control group based on background characteristics and other known criteria that were used to assign units to the treatment. Units in each group are then either matched or weighted by the propensity score to achieve a desired balance in the observed characteristics between the treatment and control groups. By achieving this balance, treatment assignment will be conditionally independent of the potential outcomes given the observed characteristics of the units. This means that the estimates of the treatment effects will no longer have bias due to assignment based on the *observable* characteristics of students. Sensitivity analysis will be performed to assess the potential for bias in the treatment effect due to remaining *unobservable* student characteristics.

Because propensity score methods are based on a statistical model of the assignment to the treatment, different units of analysis and different baseline assignment criteria may be used for matching GEAR UP participants and non-participants in states operating primarily under cohort models and in states operating primarily under priority models. For states employing cohort models, a statistical model of treatment assignment could include *both* school and student characteristics. For states employing priority models, predicting assignment to treatment will be based only on student characteristics. Propensity scores will be used to achieve balance between

the treatment and control groups on pre-treatment characteristics such as the students' gender, race/ethnicity, parent education level, and 8th grade achievement scores (i.e., ACT Explore) and on school characteristics such as the percent of underserved minorities, the percent receiving free or reduced-price lunch, and the average 8th grade achievement score.

Intention-to-Treat Design

Due to observed and unobserved differences both within and between states regarding the types, characteristics, delivery, and intensity of GEAR UP services offered, and due to an absence of student-specific data on their involvement patterns in the treatment, all analyses within the core research program will focus on the intent of the grant to treat eligible students as opposed to the documented receipt of grant-funded treatment by eligible students. Intention-to-treat (ITT) analyses include every student who was selected to receive the treatment regardless of whether or not that student complied with the treatment, received a deviation in the treatment, or withdrew from the treatment. With an ITT approach to analysis, the research question of interest is the effectiveness of *prescribing* the treatment—its "use-effectiveness"—as opposed to *receiving* the treatment—its "method-effectiveness" (Sheiner & Rubin, 1995). In other words, the results of the core research program will provide some evidence of the impact of GEAR UP-funded services on a general population of low income students knowing that not all of those students would initially accept, comply with, or successfully complete those services.

For the core research program, the independent variable will be binary, coded 1 if the student was eligible to receive direct services funded by GEAR UP state grants prior to taking ACT Explore in 8th grade and coded 0 if the student was not eligible to receive GEAR UP services at that time. The primary advantage of ITT analysis is that it provides an unbiased

estimate of the treatment effect while preserving sample size to maintain statistical power. A disadvantage of this approach is that the results of an ITT analysis are often conservative due to the dilution of the treatment effect from participant noncompliance or withdrawal.

Study Sample

The sample for the core research program will comprise those students in the 14 consortium states who are eligible to receive GEAR UP services prior to taking ACT Explore in the 8th grade. Ten consortium states administer either a cohort-only model or a hybrid of cohort and priority models (See Table 1). Within these states, only those students initially eligible to receive GEAR UP services under the cohort model will be followed by the core research program. Four states administer a priority-only model, but assignment of students to GEAR UP services under these models varies across the states. Given the focus of the core research program on the academic growth, educational plans, and eventual college enrollment and retention of students eligible to receive GEAR UP services, the core research program will incorporate priority students into the longitudinal design by similarly following a research "cohort" of priority students in these states who were first eligible to receive GEAR UP services before taking ACT Explore in 8th grade. Although this approach excludes a number of priority students from the study, the focus of the core research program is on understanding the benefits of GEAR UP for those students who were eligible to receive the services over the full course of the grant (grades 7-12). Non-participants for this study will be drawn from the combined database for ACT Explore, ACT Plan or ACT Aspire, and The ACT using the propensity score methods previously mentioned.

Timing of Data Collection and Deliverables

All student outcomes data supporting the core research program will be collected by ACT either from one of its assessments or from its partnership with NSC. Over the course of this longitudinal study, five separate data collection periods will occur: the base year in 8th grade (ACT Explore), the first follow-up in 10th grade (ACT Plan or ACT Aspire), the second follow-up in 11th or 12th grade (The ACT), the third follow-up in the first year of college (NSC), and the fourth follow-up in the second year of college (NSC). At the end of each data collection period a research report will be prepared that answers the aforementioned research questions listed for that collection period.

The timeline for data collection and reporting is provided in Table 2. Among the consortium states, 9 were awarded grants in FY2011 and 5 were awarded grants in FY2012 (See Table 1). Due to the different starting years for the consortium states, two waves of students will be followed longitudinally by the core research program. This creates a one-year lag in the timing of report publication to ensure that all consortium states are included in the study. Some consortium states that received grants in 2011 award cycle that are serving either priority students only or more than one cohort have elected to also participate in the second wave of data collection by providing ACT's longitudinal assessment system to a second wave of students.

References

Austin, P. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46,* 399-424.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960.

Sheiner, L. B., & Rubin, D. B. (1995). Intention to treat analysis and the goals of clinical trials. *Clinical Pharmacology & Therapeutics*, *57*, 6-15.

U.S. Department of Education Office of Postsecondary Education. (2011). *FY 2011 Application for grants under the gaining early awareness & readiness for undergraduate programs (GEAR UP) State Grants*. (OMB No. 1840-0821) Washington, DC: Author.

Table 1. College and Career Readiness Evaluation Consortium

State	FY Funding	Model	Data Collection	
			Wave 1	Wave 2
Arizona	2012	Cohort		X
Idaho	2011	Cohort	X	X
Kentucky	2011	Cohort	X	X
Minnesota	2011	Hybrid	X	X
Montana	2011	Hybrid	X	X
Nevada	2012	Cohort		X
New Mexico	2012	Priority		X
North Carolina	2012	Hybrid		X
Oklahoma	2011	Priority	X	X
Tennessee	2012	Hybrid		X
Utah	2011	Priority	X	X
Washington	2011	Hybrid	X	
Wisconsin	2011	Priority	X	
Wyoming	2011	Priority	X	

Table 2. Core Research Program Data Collection and Reporting Timeline

		Data Collection		
Data Collection	Grade	Wave 1 (FY2011)	Wave 2 (FY2012)	Report Delivered
Base Year	8	2012-13	2013-14	2015
1st Follow-up	10	2013-14	2015-16	2017
2nd Follow-up	11/12	2015-16	2017-18	2019
3rd Follow-up	Year 1 College	2017-18	2018-19	2020
4th Follow-up	Year 2 College	2018-19	2019-20	2021