Abstract Title Page

**Paper 4 Title:** Does Upward Bound have an Effect on Student Educational Outcomes? A Re-Analysis of the Horizons Randomized Control Trial Study

**Authors and Affiliations:** Alan B. Nathan University of Wisconsin-Madison and Douglas N. Harris Tulane University
Background / Context: High school graduation, college enrollment and college completion rates for disadvantaged students trail other groups (e.g., Asian, White, not impoverished) by a wide margin. A 2011 National Center for Education Statistics (NCES) report noted 63.5% of Black students and 65.9% of Hispanic students graduated from high school during the 2008-2009 school year, as compared to 82% of White and 92% of Asian/Pacific Islander students (Stillwell, Sable, and Plotts; 2011). Using multiple data sources including the 1997 National Longitudinal Survey of Youth (NLSY97), Heckman and LaFontaine (2007) found 65% of Black and Hispanic students graduated from high school as compared to 82% of White students.

Differences in college enrollment rates across racial groups are also evident, albeit less pronounced. Among 18-24 year olds, approximately 38% of Blacks and 32% of Hispanics enter college as compared to approximately 43% of Whites (Fry, 2010). Another recent NCES study completes the picture by reporting that the six-year Bachelor of Arts (B.A.) attainment percentages for students who entered college in 2002 was lowest among African-Americans (40%), trailing those for Hispanics (49%) and Whites (60%) (Aud, Hussar, Kena, Bianco, Frohlich, Kemp and Tahan, 2011).

If one recasts the attainments gaps in terms of differing levels of family income, the differences are even larger (Reardon, 2011). Bailey and Dynarski (2011) reported that 80% of all students who are in the highest income quartile attend college while 29% of all students in the lowest income quartile do so, resulting in a college entrance gap of 51 percentage points. Furthermore, they reported that 54% of all students who are in the highest income quartile complete college while 9% of all students in the lowest income quartile do so, resulting in a college completion gap of 45 percentage points.

Higher levels of education are associated with higher lifetime earnings (Mincer, 1974; Cameron and Heckman, 1993; Keane and Rouse, 1995). Therefore, any limits on the education received by disadvantaged groups may have negative economic consequences (Becker, 1962; Mincer, 1974; Cameron and Heckman, 1993) as well as negative health and social outcomes (Wolfe and Haveman, 2002) for these groups.

For much of the last fifty years policy-makers, instructional leaders, researchers and educators have devoted considerable resources to shrinking these educational gaps (Jencks and Phillips, 1998; Harris and Herrington, 2006; Gamoran, 2007; Magnuson and Waldfogel, 2011). One collection of initiatives intended to reduce the size of the problem are pre-college programs. In general, these programs are designed to lessen gaps in educational outcomes by providing low income and minority students with the academic background and cultural knowledge required for academic success.

Decades have passed since the launch of the first pre-college pilot programs. These initiatives have now matured to the point where several comprehensive studies have been commissioned to catalog and describe the landscape of pre-college interventions writ large (Swail, 2001; Gandara and Bial, 2001; Walker, Jurich and Estes, 2001). These reports have identified and chronicled common program goals (post-secondary enrollment or completion), target populations (disadvantaged students), key success factors (i.e., small group instruction or events, a multi-year, academically challenging program of study, and cultural and pragmatic events that focused on how a student gets accepted to a post-secondary institution).

A more recent evaluation of college outreach programs finds that interventions targeted at specific student subgroups (i.e. first generation or low-income) do not appear
to increase the educational outcomes of those students (Domina, 2009). One limitation of this analysis is that selection effects on the part of program administrators could not be addressed and effect estimates might be understated (Domina, 2009).

Selection effects may also play a role in determining who applies to college, further complicating the measurement of the effects of the intervention. For example students from low and moderate-income families who were randomly assigned to receive hands-on assistance in completing financial aid application forms as well as general information about financial aid were more likely to turn in the application, enroll in college in the ensuing year, and receive higher levels of financial aid than the control group who received only general information about financial aid (Bettinger, Long, Oreopoulos, and Sanbonmatsu, 2009).

The authors of the earlier comprehensive reports were unanimous in classifying Upward Bound (UB) as an exemplar pre-college program (Gandara and Bial, 2001; Gullat and Jan, 2003; James, et al, 2001; Swail, 2001). In addition to incorporating services that delivered all of the key success factors it was one of the very few programs that had an explicit evaluation component, and it was the only pre-college program to be evaluated using a randomized control trial (RCT).

UB is a longstanding federally funded pre-college initiative designed to increase high school graduation rates, post-secondary enrollment rates and post-secondary completion rates. Recent evaluation reports conducted by Mathematica Policy Research Inc. (MPR) to assess the impacts of UB on educational outcomes find no evidence that the program has been effective in closing achievement or attainment gaps (Myers and Schirm 1997; Myers and Schirm, 1999: Myers, Olsen, Seftor, Young, and Tuttle 2004; Seftor, Mamun, and Schirm 2009). However, research conducted by the Council for Opportunity in Education (COE) reports contradictory findings (Cahalan, 2009). COE finds evidence of a treatment effect on post secondary enrollment and B.A. completion if one site they classified as a statistical outlier was dropped from the sample.

The decade long series of MPR reports, collectively known as the “Horizons Study”, carry considerable weight because the evaluations were conducted using RCT data, in which treatment students were enrolled in the UB program and control students were excluded. In a properly executed RCT, differences in outcome measures between treatment and control groups are attributable to the treatment. Studies of this kind are considered to have high internal validity.

**Purpose**: Motivated by a desire to accurately describe the effects of the UB intervention, I reanalyze the newly released data to answer the following questions: Does treatment assignment have an effect on student education outcomes? Does receipt of the treatment have an effect on student education outcomes? Can the findings from the experiment be generalized to the universe of UB students and projects? Is there evidence of effect heterogeneity?

**Setting**: Research was conducted at 67 nationally representative Upward Bound program sites. The research time frame was 1992-2007.

**Participants**: The study sample contains 2844 observations, 1320 control subjects and 1524 treatment subjects from 67 oversubscribed (i.e. more applicants than openings)
UB project sites. Study participants were middle and high school students in grades 8-11 who applied for a slot at one of these 67 sites between May 1992 and March 1994. This initial sample was 67% female, 85% low income, and predominantly self-identified as minority (44% Black and 18% Hispanic).

**Intervention:** The goals of Upward Bound are to increase the rates at which students from low-income families or who are first-generation, graduate from high school, enroll in and complete post-secondary education. The intervention is described by the existence of, and mandatory enrollment in, the following program components. Students must pursue an academic program of study in high school, attend UB academic year sessions including after school and occasional Saturday classes plus SAT/ACT exam preparation and tutoring, and attend UB administered 6 to 8 week summer sessions. Duration is up to four years. Programs are administered by two or four-year colleges and universities or community organizations.

**Research Design:** The research design is a two-stage randomized control trial. In stage one, 67 nationally representative UB sites were randomly selected into the sample. In stage 2 student applicants to each of the 67 those sites were randomly assigned to treatment or control.

**Data Collection and Analysis:** Data was collected using 7 survey instruments, 2 administered pretest and 5 administered post-test. Surveys spanned the student’s high school and college age years through approximately age 28. Analysis was conducted using OLS and 2SLS regression equations. Reanalysis required the correction of significant design and analysis problems that threaten validity. Problems that threaten internal validity are: 1) the exclusion from the major analyses of valid outcome data for between 13% and 19% of students depending upon the outcome measure, and 2) covariate imbalances between the treatment and control groups with biases that favored the control group. Problems that threaten external validity are: 1) the use of a sample selection process, which bypassed a number of eligibility screening criteria normally employed by UB sites, and therefore generated an RCT sample that is different from the typical set of program eligible students and, 2) the use of arguably incorrect probability weighting schemes.

**Findings / Results:** My findings are different than those reported by MPR. I find evidence that UB assignment has a positive causal effect on high school graduation rates (a 4.6 percentage point increase, results are highly significant), PSE enrollment rates (a 2.9 percentage point increase, results are modestly significant), and PSE completion rates (a 4.7 percentage point increase, results are significant). In addition I find modest evidence of effect heterogeneity for a sub-group of disadvantaged students typically ineligible for UB. I estimate that their PSE completion might increase by 8.57 percentage points, which is almost twice the gain realized by the overall sample. My findings are robust to missing covariate data but sensitive to weighting assumptions used in the OLS equations. My findings are also sensitive to missing outcome data.
Conclusions: There are two important inferences I draw from my reanalysis of the Horizons study data, and both have implications for education policy. First, students who might have typically been ineligible for UB in non-experiment years may experience higher PSE completion rates than the treatment group as a whole, although the estimated effects are only modestly significant. One reason for the minimal statistical significance of this result may be measurement error in my gauge of ineligibility. Applying it would have screened out 34.6% of applicants, versus the 43.6% historically screened out (Moore, Fasciano, Jacobson, Myers, and Waldman, 1997). This difference leaves room for substantial improvement and suggests the possibility of further testing through additional data collection. Nonetheless, this new finding is noteworthy, especially with regard to the ongoing debate about merit-based program eligibility as the use of that criterion might exclude otherwise deserving students from the applicant pool. Second, UB is an effective means of increasing high school graduation rates. To the best of my knowledge, this is the first paper that finds causal evidence of this effect. This finding is important because it suggests UB is an effective way to close the rich-poor achievement gap.

There are three implications of these inferences. First, UB can be used to reduce high school dropout rates. Second, UB can be employed to narrow attainment gaps between students from low and high-income households. Third, eligibility screening processes, such as those that were in place during the time of the Horizons study should be amended to facilitate the participation of typically ineligible students, as even the most skeptical interpretation of the results suggest that these students benefit at least as much as typically eligible students and may even realize greater benefits for particular outcomes.

The findings I have presented in this paper are subject to a number of limitations. My estimates are sensitive to the use of MPR’s post-stratification and non-response weights. If either of these weighting approaches were in fact correct, my effect estimates would be insignificant. In addition I was not able to create a more accurate set of weights to support better generalization of my findings, nor could I accurately recreate MPR’s weights. This analysis highlights the importance of deriving the correct weights, if weights are to be used at all.

I did not have access to the National Student Clearinghouse (NSC) and Student Financial Aid (SFA) data, which impacted my ability to replicate MPR’s findings. Nor was I able to test how sensitive my results were to the inclusion of NSC or SFA data. The PSE data was self-reported and is subject to error. My data sets were missing covariate and outcome data, which leaves my results partially vulnerable to selection effects.
Appendices

Appendix A.

References


COE (2012). Request for Correction for the report: The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation (Referred to as the Mathematica Fifth Follow Up Report), prepared by Mathematica Policy Research. Council for Opportunity in Education


Harris, D. N. (2012). Improving the Productivity of American Higher Education through Cost-Effectiveness Analysis. Wisconsin Center for the Advancement of Postsecondary Education (WISCAPE)


Seftor, N. S., Mamun, A., and Schirm, A. (2009). The impacts of regular upward bound on postsecondary outcomes seven to nine years after scheduled high school graduation. final report. US Department of Education. P.O. Box 1398, Jessup, MD 20794-1398


### Table 1. A Comparison of My ITT Best Estimate Findings with MPR and COE Published Results

<table>
<thead>
<tr>
<th></th>
<th>High School Graduation</th>
<th>12th Grade GPA</th>
<th>PSE Enrollment</th>
<th>PSE Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>My Best ITT Estimate</strong></td>
<td>0.0457***</td>
<td>0.0454</td>
<td>0.0286+</td>
<td>0.0466*</td>
</tr>
<tr>
<td></td>
<td>(0.0121)</td>
<td>(0.0342)</td>
<td>(0.0156)</td>
<td>(0.0230)</td>
</tr>
<tr>
<td><strong>MPR</strong></td>
<td>-0.0100</td>
<td>0.000</td>
<td>0.0156</td>
<td>0.1300*</td>
</tr>
<tr>
<td><strong>COE</strong></td>
<td>Not Applicable</td>
<td>Not Applicable</td>
<td>0.0690***</td>
<td>0.0900</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.1300)</td>
<td>n.p.</td>
</tr>
</tbody>
</table>

Table Notes: The COE measure for PSE completion is B.A. completion. The MPR PSE completion findings appear in the 5th follow up report-Appendix C Tables C.1 case #1 and C.7 case #1. These findings exclude any SFA or NSC data.

n.p. = not published
Standard errors in parentheses
+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$