

A PART Tragedy: The Case of Upward Bound Correcting for Study Error in the Random Assignment 1992-2004 National Evaluation of Upward Bound

Margaret Cahalan

Dr. Cahalan currently serves as the Secondary-Postsecondary Cross Cutting (SPCC) Team Leader within the Policy and Planning Studies Services (PPSS) within the US Department of Education (ED). She has worked in the area of TRIO evaluation since 1991, as a contractor serving as the project director for national evaluations of Student Support Services and Talent Search, and the UB performance reporting contracts. She also served as survey director for the Third Follow-up and initiation of the Fourth Follow-up data collections for the study that is the subject of this article.

Abstract: The paper takes a critical look at study error relative to findings from the National Evaluation of Upward Bound (UB) and explores policy and methodological lessons learned. The nationally representative random assignment study followed a multi-grade cohort from 1992-93 to 2003-04. Following published reports of lack of overall positive effects, OMB gave UB an ineffectual Performance Assessment Rating Tool (PART) rating and Bush Administration budget recommendations in FY05 and FY06 called for zero funding. Efforts were also made to reform the program, by requiring projects to serve more academically at-risk students. Major technical issues discussed include: 1) extreme unequal weighting; 2) treatment-control group non-equivalency; 3) survey non-response bias; 4) lack of standardization for expected high school graduation year (EHSGY) in outcome measurement; and 5) service substitution and dropout issues. Our major finding is that when administrative records are used to supplement survey data and outcomes standardized by EHSGY, contrary to previously published reports, the UB program demonstrated statistically significant positive effects on the major goals of the program; postsecondary entrance, application for financial aid; and attainment of postsecondary credentials.

Key words: evaluation of pre-college programs; random assignment evaluation; sampling and non-sampling error; survey non-response bias

Margaret Cahalan, US Department of Education, Policy and Planning Studies Services (PPSS), 400 Maryland Ave, SW Washington D.C. 20202; e-mail: Margaret.cahalan@ed.gov.

Author' Note: **All tabulations and views reported in this paper are the sole responsibility of the author and do not reflect any review, authorization, or clearance by the United States Department of Education.** Data reported in this paper are based on the 1992-93 to 2003-04 National Evaluation of Upward Bound conducted by Mathematica Policy Research under three contracts to the US Department of Education spanning the period from 1991-2007. The evaluation was conducted under sponsorship of the Policy and Program Studies Services (PPSS), Office of Planning, Evaluation, and Policy Development (OPEPD), US Department of Education. While any errors are the sole responsibility of the author, there are a number of persons who have shared their technical expertise, insight, and criticism with the author who have greatly contributed to this paper. I would especially like to thank, David Goodwin, who has provided constructive criticism, encouragement, and careful review of several drafts of this paper over the course of two years. I would also like to thank David Myers who has critiqued this paper and offered valuable suggestions for its improvement. In addition Lana Muraskin, Jay Noell, James Chromy, Peter Seigel, David Bergeron, Frances Bergeron, John Clement, Sandra Furey, each in different ways have provided insight and critiques that have strengthened the paper.

I. Introduction

The last 4 years have witnessed considerable debate in Congress concerning the Upward Bound (UB) program's continued funding, Absolute Priorities, and evaluation procedures. These debates have culminated in the recent passage of the re-authorization of the Higher Education Opportunity Act (HOEA- HR4137). This act contains a prohibition against TRIO evaluation studies that require projects to recruit more students than they would normally serve for purposes of random assignment evaluation. Much of this debate was sparked by the actions following the program being designated as "ineffective" in the OMB Program Assessment Rating Tool (PART) process. The UB program was rated as "ineffective" on the basis of only one ambitious random assignment study begun in the 1990's that is the focus of this paper, ---the National Evaluation of Upward Bound (designated here as the *1992-2004 UB Evaluation*). Following the "ineffective" rating the Bush Administration budgets in FY05 and FY06 called for "zero funding," and at the same time the Department of Education (ED) developed various initiatives and priorities designed to "reform" the program on the basis of the *1992-2004 UB Evaluation* results. ED also began a new evaluation study to evaluate these reforms. There followed an intensive effort on the part of the TRIO community against the Priority and the new evaluation. In 2007, Congress cancelled funding for the new evaluation, and in 2008 the Absolute Priority itself was cancelled by the HEOA re-authorization.

By way of introduction, the first part of this paper gives an overview of the *1992-2004 UB Evaluation*, and its interaction with the development of ED policy. The second and third parts of the paper focus on a detailed examination of study technical issues and present results of a re-analyses correcting for study error. These results show quite different findings concerning the UB program effectiveness than have been thus far been published. A final section considers lessons learned and recommends that the PART rating for Upward Bound be re-considered in the light of these new analyses. Given that the *1992-2004 UB Evaluation* took on the character of a "high stakes evaluation" with clear consequences for the program in terms of its PART labeling, policy decisions, and funding recommendations; and also now has had serious Congressionally mandated consequences in terms of evaluation method prohibitions, it seems especially important that these findings be made public. The history of this unique study provides an excellent case study in the issues faced by evaluators in designing and implementing a large scale nationally representative random assignment study and in using the results to address questions of national educational program evaluation and policy development.

Program and Study Background

First begun in 1965, Upward Bound is one of the earliest, most intensive (typically involving a 6 to 8 week summer residential stay on a college campus and a non-residential academic year component), and by legislative

requirement, academically focused of the federal high school pre-college supplemental service programs. By 2008, there were over 900 UB projects, operating through competitive grants typically made to colleges and universities. The program was designed to help achieve one of the first articulated major missions of the US Department of Education --equal educational access.¹ Legislatively defined eligibility requirements specify that two thirds of the students served must have family incomes of 150 percent of poverty or less and neither parent can have completed a bachelor's degree (known as "first generation students"). The other one-third of students served must meet one of the two eligibility requirements.

The *1992-2004—UB Evaluation*, had an ambitious rigorous but controversial design employing a random assignment method to assess the impact of the program combined with a national probability sample of projects designed to give the study external validity (something very rare). The study has been conducted very carefully with high levels of resources, following a sample of about 3,000 randomly assigned treatment and control group members from about 70 projects² with six surveys over an 11-year period, and has had very high response rates. However, the subjects of the evaluation, the Upward Bound projects and the regional and national organizations representing the TRIO community (such as WESTOP and the Council on Opportunity in Education (COE)), have questioned the evaluation's feasibility, propriety and accuracy, since its initiation and more so as the findings were reported. The two most frequent criticisms concerned the study requirement that the participating projects deliberately recruit at least twice the number of anticipated openings for random assignment into treatment and control groups; and the fact that the study analyses, did not take into account receipt of equivalent or other similar pre-college services by the control group (see for example critique by Baker 1999). As noted above, the study has also been associated with the development of Department of Education (ED) policy toward the program over the last decade, with budget debates in Congress, and ultimately with significant language incorporated into the 2008 re-authorization of the Higher Education Opportunity Act (HEOA) prohibiting ED from requiring projects to participate in studies that required deliberate over-recruitment of participants for evaluation purposes—the essential method of the study.

Context of Paper

This paper reports the results of quality assurance (QA) review of the study design and past analyses procedures, as well as new analyses, undertaken by the author over a period of two years 2006 to 2008, while serving as the Technical Monitor for the study at ED. It is the thesis of this work that interrelated sources of both sampling and non-sampling error have confounded some of the conclusions thus far published from the study. When these errors are mitigated, by making outcome measures more precise through use of

¹ Upward Bound (begun in 1965), Talent Search (begun in 1971) and Educational Opportunity Centers (EOC) (begun in 1975), and Gear Up (begun in 1998) are four major federal high school programs focused on postsecondary preparation in middle and high school.

² The original sample included 70 projects; due to project level non-response, the participating sample was 67 projects

administrative records in addition to surveys and by standardizing measures by Expected High School Graduation Year (EHSGY); and when survey sampling and weighting issues are addressed, the best estimates for the study indicate that the program had statistically significant and substantial impacts on the major goals of the program.

Policy Background and History

The study was conducted under the general congressional authorization for TRIO evaluations that stated they shall be conducted: “for the purpose of improving the effectiveness of the programs and projects assisted under this chapter.” The study sought to change as little about the normal functioning of the program as possible; however, in order to implement a random assignment, projects were asked to recruit at least twice the number of anticipated openings in the study period and to recruit earlier and with a broader reach. A Baseline Survey completion was a requirement for being put on the “waiting list” to be given the chance for being offered a slot in the program in the study period. Because, Upward Bound applicant eligibility was from grade 8 to 11 and the study spanned over an 18-month period sample members expected high school graduation year (EHSGY) spanned over a 4-5 year range. The *Horizons* study (the name printed on materials given out in recruitment) followed participants with 6 participant surveys, project surveys, target school surveys, case studies, and detailed service records.

Thus far the Department of Education has published three reports containing impact estimates derived from the study (Myers and Schirm 1996; 1999; and Myers et. al. 2004). Mathematica Policy Research, who conducted the study under three contracts to the Department of Education, concluded in the last published Third Follow-up report that: “the Upward Bound Program had no effect on overall enrollment or total credits earned at postsecondary institutions, but it may have increased enrollment in four-year postsecondary institutions ” (Myers et. al. 2004).³ The Third Follow-up report also found that there were significant and substantial effects for the bottom 20 percent of study participants on 9th grade academic indicators, and for students with lower expectations (defined as expecting less than a bachelor’s degree).

Following release of the reports with no overall effects, the program was given an “ineffective rating” in the newly implemented Office of Management and Budget (OMB) Program Assessment Rating Tool (PART) process established by Presidential Order. In FY05 and FY06 these evaluation results were used to justify Bush Administration zero funding budget recommendations, not just for Upward Bound but also for programs with similar goals--Talent Search and GEAR UP. After strong lobbying campaigns by the Council on Opportunity

³ An internal unpublished final fourth Follow-Up report submitted by Mathematica similarly stated. “Upward Bund had no detectable effect on the rate of postsecondary attendance for the average eligible applicant”(Stuart, et. al. 2006).

in Education (COE) and other college access organizations, funding was restored in Congressional budgets and the Administration budgets in FY07 and FY08 dropped zero funding of the programs. Prompted by OMB, efforts were also made to reform the UB program by initiatives and the establishment of the *Absolute Priority for Upward Bound Program Participant Selection and Evaluation* published in the Federal Register in September of 2006. This priority required that one-third of all students served be academically at-risk defined as having low grades or flunking a high school competency test and also specified 9th grade entrance requirements. Although justified by the *1992-2004 UB Evaluation* study results, this approach was also more consistent with the ED focus on raising achievement test scores for those students who were deemed not proficient under *No Child Left Behind*. This was also consistent with the Bush Administration stress on “accountability and achievement gains” rather than the “access and attainment” focus of previous years. At the same time, ED, under the Institute for Education Sciences (IES), launched a new random assignment evaluation contract awarded to ABT, Associates to study the Absolute Priority’s effectiveness.

The *2006 Absolute Priority* and the accompanying evaluation employed much the same methods as the *1992-2004 UB Evaluation*, and coming soon after the zero funding requests in FY05 and FY06, generated a strong protest in the TRIO and the college access practitioner community. COE argued that ED was overstepping its authority in setting Absolute Priorities in the grant competition that changed the legally specified participant eligibility criteria and that only Congress could set TRIO eligibility requirements. Some program defenders argued that, having failed in attempts to kill the program, the Administration was now trying to “reform” the program to reflect its own priorities by turning it into a dropout prevention program rather than a college access program. It was also argued that the methods to be employed in the new UB evaluation were unethical in once again requiring projects to deliberately over-recruit and then deny services to some applicants for study purposes. If services were not denied than the new study would be confounded, as they argued had been the *1992-2004 UB Evaluation* which had been described by some in the TRIO community as “pseudo-science.” Projects argued that their institution’s Internal Review Boards (IRBs) would not grant approval for such procedures. In late 2007, Congress blocked further funding for the new IES evaluation study, and ED formally cancelled the evaluation in early 2008.

As noted, these debates culminated in the insertion into the July 2008 Higher Education Opportunity Act (HEOA: HR4137) of language prohibiting the Department of Education (ED) from using the basic random assignment recruitment method, that was used in the *1992-2004 UB Evaluation* and the cancelled study. Specifically, the new HEOA act prohibits ED from requiring projects to participate in evaluations when the study, “requires the eligible entity to recruit additional students beyond those the program or project would normally recruit; or results in denial of services for an eligible student under the program or project.” The HEOA language, while also specifically mandating rigorous evaluations in Upward Bound by 2010, also

strengthens the focus on evaluations designed to identify those practices most useful to achieving program goals (rather than overall black box evaluations of program effectiveness). The recent HEOA re-authorization also cancelled the Absolute Priority to serve one-third high -risk students and one-third by grade 9.

Paper Technical Approach and Study Questions

This paper draws from methodological work from three intersecting traditions. These are work in experimental design and program evaluation examining the threats to validity (for example, Heckman 2000; and Shadish, Cook, and Campbell 2001); work within survey methods research on “survey error” (for example, Groves, et. al 2004); and, the statistical and program evaluation standards (for example, the *Joint Program Evaluation Standards* 1994, and the *National Center for Education Statistics Standards*, Seastrom 2002). Specifically, the paper includes a critical look at several random assignment experimental design assumption validity issues, and related survey sampling and non-sampling measurement error issues relative the 1992-2004 UB Evaluation. *The approach is to identify the sources of error and then to attempt to mitigate or correct for the error by using multiple ways of viewing the data that reflect a standards based approach to statistical analyses and to evaluation practice.* We present what we consider to be best group of estimates of program impact given the study characteristics.

The study examines the Upward Bound study design and analyses relative to 4 basic assumptions of random assignment studies:

- The sample is representative of the population for which the study is intended to generalize;
- The treatment and control group are equivalent on factors related to outcomes;
- The treatment and control group are treated equally except for the treatment; and
- The treatment and control group are mutually exclusive with regard to the treatment.

This paper includes a critical look at related experimental design validity issues, as well as survey sampling and non-sampling measurement error. We are using the terms sampling and non-sampling error generally to distinguish those errors that come from the sample design from other errors, and not in a precise quantitative sense. By sampling error, we simply mean the error caused by observing a sample instead of the whole population. As we are using the term here –non-sampling error is a catch-all term for the deviations from the true value of the estimate that are not a function of the sample chosen, including various systematic errors. An example might be non-response bias and lack of precision in outcome measurement is also a form of non-sampling error.

Questions Examined in This Paper

Specifically in this paper we address the following questions:

1. **Sample Design, Weighting Issues.** To what extent are the study results sensitive to the unusual sample design that attempted to combine a first stage national probability sample representing many types of grantees with numerous strata, many with only one project member, and corresponding unequal weighting issues, with a second stage highly stratified random assignment of unequal numbers of applicants?
2. **Treatment and Control Group Equivalency.** Given that this was not a simple random assignment study, but a complex highly stratified multi-stage sample to what extent does the sample suffer from treatment and control group non-equivalency?
3. **Standardization of time sensitive outcome variables.** Given that the sample spanned at least four cohorts of expected high school graduation year (EHSGY), what is the result of standardization of outcome measures to years since EHSGY (something not done in published or draft reports to date)?
4. **Survey Non-Response Bias and Attrition from the Study.** Do we obtain the same results as the results published in the Third Follow-up report relying only on survey data when we use a longitudinal file supplementing survey data with administrative record information?
5. **Treatment Substitution and Drop-Out Issues.** What results do we obtain when we model actual participation in the program or Treatment on the Treated (TOT) in addition to the Intent to Treat (ITT) results of being given the opportunity to participate? How do outcomes for those participating in Upward Bound (UB) or Upward Bound Math Science (UBMS) compare with those reporting participating only in some other pre-college service? What results do we obtain when we model the association of receipt of any pre-college support services having some academic focus, with postsecondary outcomes?

We use data from the Baseline Survey conducted in 1992-94 and subsequent follow-up surveys as well as student transcripts, federal Student Financial Aid (SFA) application and award matching information for ten academic years (from 1994-95 to 2003-04) and (for bachelor's degree awards only; due to lack of coverage in the period in which sampled students were graduating high school) National Student Clearinghouse (NSC) data (from 1995-2004). Altogether, about 1,500 students were randomly assigned to the treatment group and about 1,300 were not selected to be offered the "UB opportunity" (were assigned to the control group) from 67 projects out of 395 projects that had been operating 3 years at the time of the study. The surveys collected

information on student background, high school experience, supplemental support services participation, postsecondary participation, and employment activities. The data collections were made more complex by the fact that the study collected information based on applicants who could be in various grades from 8th to 11th over an 18 month period of applicant recruitment.⁴ Hence in different rounds of the follow-up surveys, students were in different grades.

II. A Critical Look at Sources of Study Error

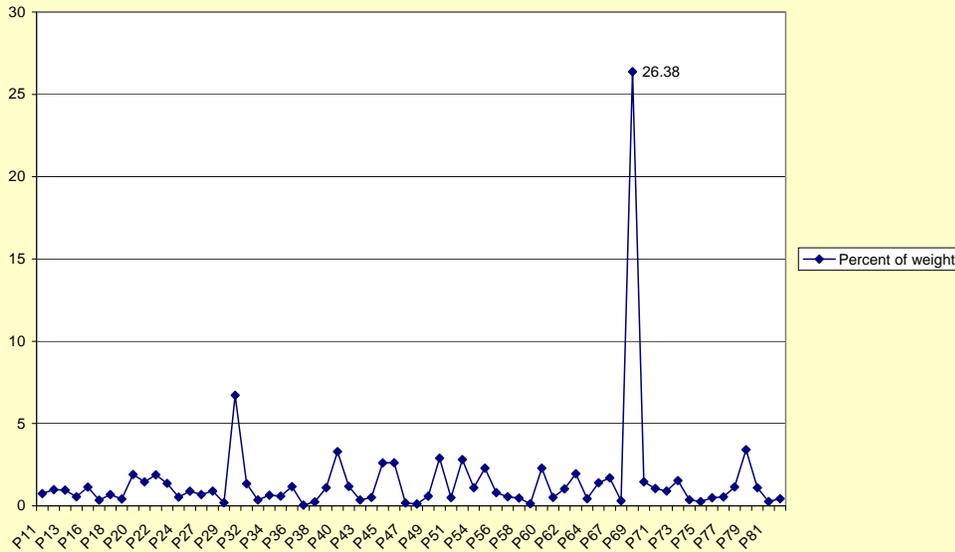
Before presenting results of the re-analyses, the next section, keyed to the study questions noted above, identifies 5 interrelated sources of error in the study published analyses, and also identifies a correction or mitigation to address the issue. This discussion explains the analyses choices made in the subsequent section presenting findings and what we consider to be “best estimates” from the study. Some of these sources of error have been raised by stakeholders, and others have not been known and could not have been known with the documentation that has thus far been published concerning the study sample design and outcome measures. As noted the major thesis of this review and re-analyses is that a number of factors have combined together to confound some of the conclusions that have been thus far published concerning the average effects of the Upward Bound program on the major outcomes of interest.

1. Sample design flaws and unequal weighting issues—Due to a goal of representing and disaggregating results by a wide range of project demographics, the sample is highly stratified with stratum of unequal sizes. The first stage project level sample included 46 strata for 70 projects⁵, and a number of the strata are represented by only one project. Projects were allowed to develop additional strata for their applicants reflecting such things as sex or recruitment round, so that that the study actually had 339 strata with an average of only about 8 sample members per strata. Baseline weighting already reflected large variation in probabilities of project and applicant selection and poststratifications were also done to take into account these individual project level strata and to equalize treatment and control totals per project. As a result there are very serious unequal weighting issues with one project representing 26 percent, 3 projects representing 35 percent, and 8 projects representing 50 percent of the weights (see figure 1). Due to poststratifications there is also unequal weighting within projects.

⁴ As study entrance was over an 18-month period, some students completed the Baseline Surveys after completing the 1991-92 school year instead of the 1992-93 year printed on the survey form adding to the complexity of the time referencing. A few students were in grade 11 in the reference year.

⁵For this study, only UB projects that had been operating for 3 years or more were considered eligible. At the time there were about 395 projects meeting this criteria. Of the 70 projects sampled, the participating sample was 67.

Figure 1. Percentage of sum of the weights by project of the 67 projects making up the study sample: National Evaluation of Upward Bound, study conducted 1992-93-2003-04



NOTE: Of the 67 projects making up the UB sample just over half (54 percent) have less than 1 percent of the weights each and one project (69) accounts for 26.4 percent of the weights.

SOURCE: Data tabulated December 2007 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.

Contrary to probability sampling standards when probability of selection weighting is anticipated, in a flawed sample design, only one project was selected from the largest study defined stratum and this project ended up accounting for fully 26 percent of the weights. Consequently we are unable to tabulate an estimate of sampling variance for the strata. Moreover, examination of the project with the 26 percent of the weight reveals that, although randomly selected to represent the largest 4-year public stratum, it is actually a former 2-year college taken over by a city university system to serve as a branch campus, with largely career and technical less than 4-year programs. It also does not have the hallmark UB summer residential program—as it has no on-campus housing. *So in addition to the serious design flaw of a one project stratum for the largest study defined stratum with an extreme outlier weight, there are serious questions concerning the ability of the selected project to represent the defined stratum.* To mitigate this situation, we conducted weighting sensitivity analyses and results of all models included in this paper are presented both weighted and unweighted and with and without the outlier weight project.

2. Treatment-control group bias in favor of the control group—Examination of the project with the 26 percent of the weight also indicated substantial non-equivalencies in the composition of the treatment and control groups in favor of the control group on key variables associated with outcomes. As can be seen, from table 1 (and highlighted in figure 2) the weighted sample without project 69 is well matched with regard to

treatment and control group attributes on the baseline and academic variables of interest to the study. Using the project poststratified analyses weights; however, project 69 shows indications of non-equivalency between the treatment and control group on several key variables found to be positively associated with differences in the major outcome variables ----- all in favor of the control group. These include:

- Educational expectation ---among controls, 56 percent expected to complete an advanced degree (MA or higher), while among the treatment group, 15 percent so expected—a 40 percent spread.
- Taking Algebra or above in 9th grade - among the controls, 98 percent had Algebra or above in 9th grade, while among the treatment group, the comparable figure was 79 percent —a 19 percentage spread.
- Classified as high academic risk—among the controls 8 percent were classified as high academic risk, while 33 percent of the treatment group was so classified—a 25 percent spread.
- Higher 9th grade GPA ----among the controls 55 percent had a GPA of below 2.5 while among the treatment group 75 percent had a GPA of below 2.5—a 20 percent spread
- Males ---among the controls 14 percent were male while in the treatment —24 percent were male
- Higher grade at baseline—among the controls 42 percent were in 8th grade or below as reported on the Baseline Survey in variable B1 and among the treatment group 60 percent were in 8th grade or below.

The weighted differences between treatment and control group noted above are statistically significant using simple tests, but are not designated as significant here because the large weighted Ns make significance very easy to obtain in simple tests. Tests of significance were also done on the unweighted data (n= 85) and also found to be significant. However, tests for the weighted differences noted above were not done taking into account the individual project level stratifications for which limited information is available.

Table 1. Percent of treatment and control group sample members having various attributes reported on Baseline Survey and 9th grade transcripts: project 69, all other projects, and all projects: National Evaluation of Upward Bound study conducted 1992-93-2003-04

Variable	Project 69			All other projects		All projects	
	Treatment	Control	Diff.	Treatment	Control	Treatment	Control
Expect at least BA	73	84	11	76	81	75	82
Low expectations	27	16	-11	24	19	25	18
Expect MA or higher	15	56	41	37	38	31	43
Expect MA	0	20	20	13	14	10	15
Expect PhD	15	35	20	24	24	23	27
Expect less BA	27	16	-11	24	19	25	18
Expect 2- year	13	10	-3	14	11	13	11
Expect high school only	3	3	-	3	3	3	3
Expect less two year degree	10	2	-8	5	4	7	3
Male	24	14	-10	35	34	32	29
Not native speaker	32	19	-13	10	11	16	13
Participate services before	31	17	-14	29	31	29	30
Took algebra in 9 th grade	79	98	19	58	58	64	70
High academic risk	33	8	-25	23	21	24	19
Lower academic risk	67	92	25	77	79	81	76
GPA below 2.5	76	55	-21	53	54	59	54
Percent of all grade 9 sample members based on 1993-94 grade	74.2	25.8	-48.4	51.3	48.7	55.9	44.1
Percent of all grade 10 sample members based on 1993-94 grade	34.3	65.7	31.4	48.3	51.7	44.7	55.3
White	0	0	-	30	28	20	22
Hispanic	38	44	6	13	17	21	23
Black	60	56	-4	48	48	50	51
Other race	0	2	2	9	8	6	6
Grade 8 or below reported on base	60	42	-18	59	60	59	55
Grade 9 on SS form	63	47	-16	42	44	48	45
Grade 10 on SS	28	45	17	31	29	30	33
First generation college only	27	20	-7	14	15	17	17
Low inc & first generation	73	78	4	81	79	79	79

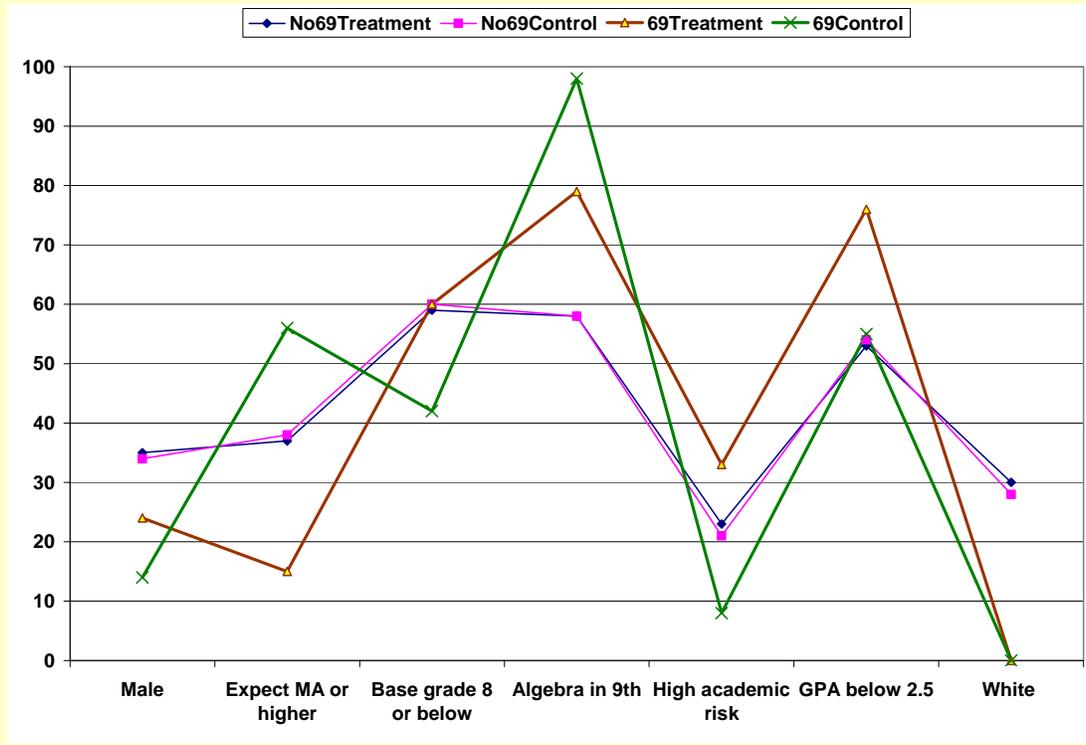
NOTE: Weighted data for project 69 based on 85 sample cases (52 controls and 33 treatment cases poststratified weighted to 11,536 cases-- 5,768 treatment group members and 5768 control group members). Using simple tests, weighted differences over 4 percentage points between treatment and control group for project 69 presented above were statistically significant, but are not designated as significant here because large weighted Ns make significance very easy to obtain in simple tests. Tests of significance were also done on the unweighted data for project 69 (n= 85) and differences 12 percentage points and over were found to be significant. Complex standard errors using the individual project level strata were not done.

SOURCE: Data tabulated December 2007 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.

We do not know how these differences, came into the study from the project, whether by chance through random assignment or some other factors. Whatever the source of the non-equivalencies, they create a source of bias in the impact estimates. Looking at the data one suspects that this configuration may have something to do with the interactions with the large weights related to the adjustments done to equalize the treatment and control group sum of weights per project, the disconnect between the number of openings and the actual number who completed the Baseline Survey to become applicants (there are more control group members than treatment group members from this project), the target school characteristics including CTE and Math and Science programs, the presence of a Talent Search program also operating at the project, recruitment for Upward Bound Math Science, and the role of the “must serves” selected with certainty into the treatment group and eliminated from the study.⁶ Examination of the differences between the treatment and control group (table 1) suggests that the treatment group from this project was more likely to have younger and more at risk students who had received prior services, perhaps recruited from Talent Search and the target school CTE program; and the control group more likely to have a higher proportion of students who appear resemble the Upward Bound Math Science population. This project had an unusually large number of applicants and may have, for example, recruited from a whole math or science class in the Science and Math or Engineering target school magnets, and more from Talent Search and the CTE program vocational school for another round (there are a relatively high proportion of students expecting a less than 2-year degree in treatment group). There were apparently two sub-strata for random assignment of unknown definition in the documentation for the study, and there were at least 6 “must serves” in the treatment group excluded from the study. In one of the project 69 sub-strata the number of treatment and control group members was roughly equal. However in the other sub-strata, it appears there were far fewer openings relative to the numbers on the waiting list and this introduced unequal weighting within this project itself. The weights were 95 for controls in the one sub-stratum and 185 for treatments in that same sub-stratum. The other sub-stratum had weights of 159 for both treatment and controls. There were also UBMS participants among both controls and treatments. These factors combined with the poststratifications and the large base weight may have combined in some way to unbalance the treatment and control groups. (There were 52 control members and 33 treatment members--unweighted). However, these weighted non-equivalencies came into the study, they appear to be introducing some bias in favor of the control group into the overall UB sample, such that the study requirement of equivalency of treatment and control group that the random assignment study is meant to ensure appears to have been compromised.

⁶ In the random assignment process, projects were allowed to designate students on the Horizons waiting list that they considered “must serves”. These students were removed from analyses. There were about 193 such cases or about 6 percent of the 3029 students who completed the baseline survey.

Figure 2. Percentage of project 69 and all other projects having various attributes by treatment and control group status: National Evaluation of Upward Bound, study conducted 1992-93-2003-04



NOTE: Poststratified weighted data. Project 69 tabulation based on 85 sample cases (52 controls and 33 treatment cases-- poststratified weighted to 11,536 cases 5,768 treatment and 5,768 controls). The category "No69treatment" and "No69control" represents all the other projects in the sample excluding project 69; these 66 projects are considered to represent 74 percent of the UB applicants in the study period.

SOURCE: Data tabulated December 2007 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04

One of the more problematic issues concerns the difference in expected high school graduation dates between treatment and control groups discussed in more detail below. In this area, small differences can have a large effect—especially in the early years of the study. Models used in this report attempt to control for these differences. In addition there are apparent non-equivalencies in academic indicators; however, as noted these non-equivalencies are not controlled in the models as the only indicators come from the 9th grade transcripts, after the “treatment period” had begun for some participants

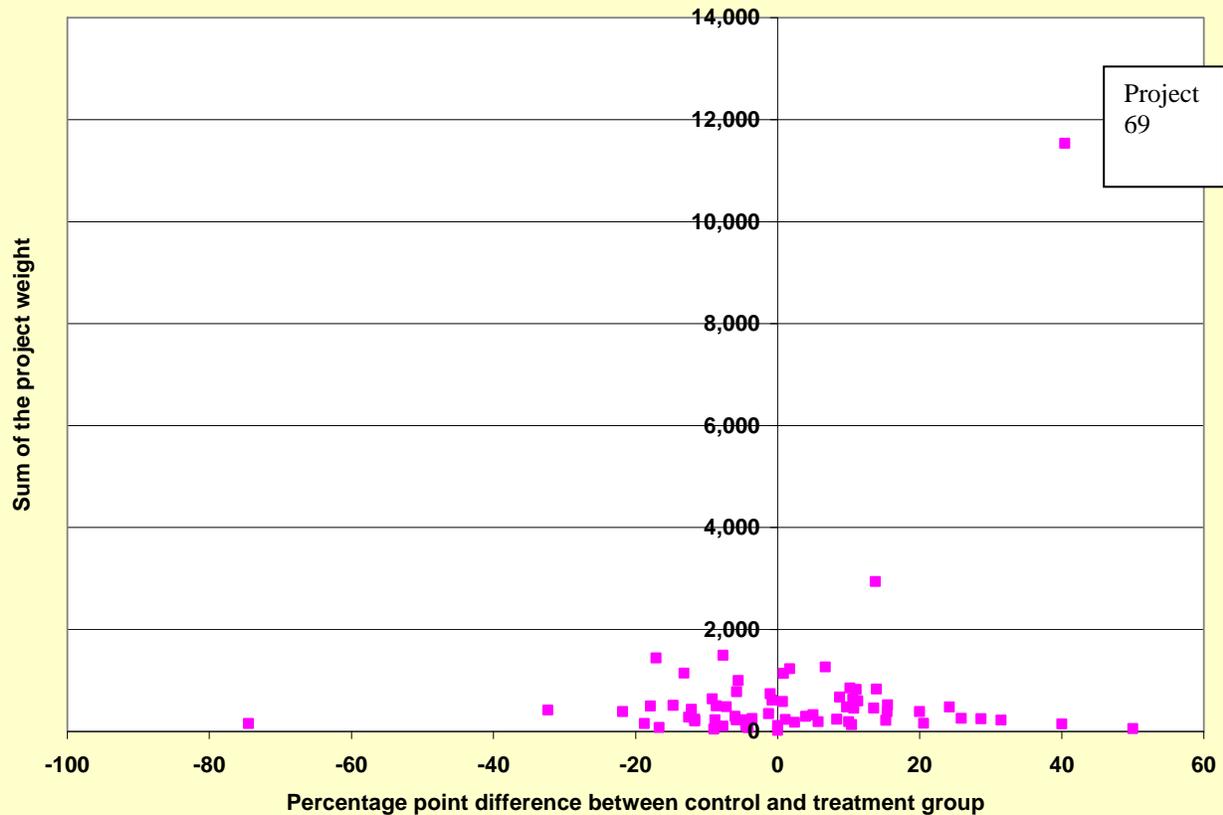
These difficulties may point up why most random assignment studies that are interested primarily in relationships between a treatment and an outcome, do not attempt the complicated dual task of producing precise national point estimates and at the same time obtaining an understanding of the relationship of a particular type of treatment to outcomes. The data are being weighted to application totals that were always an

artifact of this particular study and have always had little external validity in that most projects do not recruit in the manner required by the study in that year.⁷

The primary issue in this case is not the probability that these differences might have happened by chance as tests of statistical significance tell us, or some other more systematic error; but rather the role they may be playing in biasing the impact estimates. For example, there is a 41 point spread between treatment and control in the percent expecting an advanced degree (MA or above) at baseline in project 69; and there is a 12 percent difference in favor of the controls on this variable in the overall sample when project 69 is included. This difference is not statistically significant with a design effect of 18. Given the small sample sizes involved per project (ranging from 4 to 100), the differences noted above are not unique in this sample when individual project level treatment and control differences were considered, and the overall differences between treatment and control group do not usually reach statistical significance. Project 69 is however, unique in having such a large difference among projects having at least 60 cases (30 per cell if evenly divided). The outlier quality of project 69 comes from the combination of the seemingly large differences with the large weight, whether these differences are the result of chance or some other factor. (table 1 and figure 2). These differences, combined with the large weight, compromised the overall equivalency of the treatment and control groups on factors related to outcomes; and were inadequately controlled for in published analyses. While project 69 is not the only project to have differences between treatment and control group, on such variables as educational expectations, as is illustrated in figure 3, presenting the differences in percent expecting an advanced degree at baseline--the large differences combined with the large weight, result in the introduction of bias in favor of the controls into the overall sample. These differences on educational expectations, expected high school graduation year, and 9th grade academic indicators are inadequately controlled for in published analyses. In the re-analyses to be presented below, we attempt to increase the precision of our outcome measures by standardization as to EHSGY and by using administrative records. As did Mathematica, we do not attempt to control by 9th grade academic variables due to the fact that for many sample members 9th grade occurred after the start of the treatment period. To mitigate the role of the outlier project, we present all results both with and without this one outlier project, and also include some additional baseline controls in the models.

⁷ It does not appear that even when one has a reasonable sample design (which this study does not) that there is unanimity in the field about using weights in regression based complex effect estimation with weights that reflect multi-stage samples (McLaughlin, Drori, and Ross 2000; Chambers 2003; Little 2003; Phiffermann et. al 1998 2004). In the context of HLM work, a recent NCES report notes... "there is no unanimity in the field with respect to this question, even as to whether weights should be used at all" (Braun, Jenkins, Grigg 2006).

Figure 3. Plot of the differences between control and treatment group in percent expecting an advanced degree (MA or above) at baseline by sum of the weights by project: National Evaluation of Upward Bound, study conducted 1992-93-2003-04



The Y axis (0-12,000) represents the sum of the weights and the X axis the weighted percentage point difference between treatment and control group. Squares to the right represent positive differences for the control group and to the left of the line negative differences for the control group. This chart shows that project 69 stands alone considering the difference multiplied by the sum of the weights.

SOURCE: Data tabulated May 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.

3. Lack standardization for expected high school graduation dates in outcome variable calculation. As noted, the study applicant cohort was a multi-grade one. The last grade completed grades reported on the Baseline Survey spanned from 7 to 10 and randomization spanned over an 18-month period, adding additional complexity to the study. Published reports and draft reports submitted to ED, assuming that randomization would ensure an equal distribution between treatment and control group have grouped all sample members together looking at results reflecting reports at the time of the various survey rounds. These analyses failed to take into account that the sampling was not a simple random sample with equal assignment by grade into treatment and control group by grade--but involved numerous strata that did not take into account

EHSYG in assignment. As table 2 indicates there are small unbalances between the treatment and control group in expected high school graduation year (EHSYG) with the control group more frequently being in a higher grade at baseline (seemingly largely introduced by one outlier project). These appear to have been biasing the results obtained without standardization. To correct for this we tabulated an expected high school graduation year (EHSYG) and looked at postsecondary entrance within +1 year of the EHSYG (within about 18 months of June graduates) and by within +4 years following the high school graduation year, and looked at bachelor’s receipt within +6 years of EHSYG.⁸

Table 2. Percentage distribution of grade in 1993-94 as reported on the First Follow-up Survey, total sample and project 69: National Evaluation of Upward Bound (UB), study conducted 1992-93-2003-04

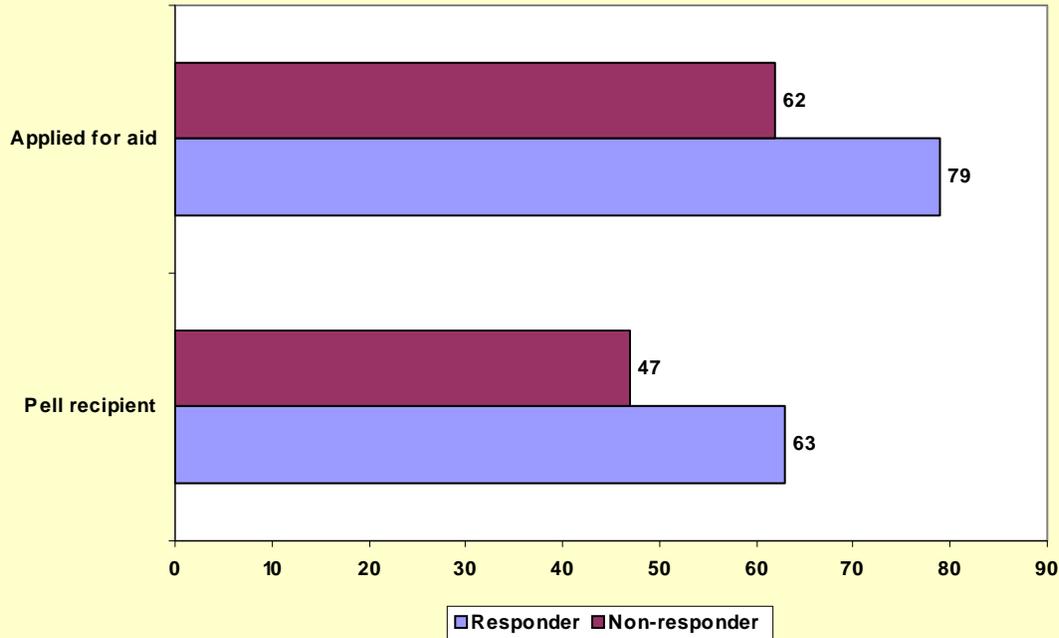
First Follow-up grade in 1993-94	Total sample		Project 69 sample	
	Treatment	Control	Treatment	Control
9	32.4	25.6	33.7	11.4
10	37.7	46.9	29.3	54.6
11	22.2	21.8	20.4	25.5
12	4.1	4.0	6.8	8.5

NOTE: Weights are the poststratified weight used analyses. A portion of those coded as grade 9 (about 9 percent of the total) may have been just entering grade 9 in 1994. Detail does not sum to 100 percent due to missing data
 SOURCE: Data tabulated May 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education.; study conducted 1992-93-2003-04.

4. **Study attrition and non-response bias**—Study attrition, especially differential attrition between treatment and control groups, is a concern in longitudinal studies. The UB evaluation survey response rates have been exceptionally high, but they have declined with each round of data collection. They range from 99 percent on the Baseline (required for entrance into the study “waiting list”), 97 percent on First Follow-up, 85 percent on Second Follow-up, 81 percent on the Third Follow-up, 75 percent on the Fourth follow-up, to 74 percent on the Fifth Follow-up. Previous published reports have been based on only responders to the survey rounds with weights adjusted for non-response. Control group members have consistently had response rate differences of 4-5 percentage points lower than treatment group members. Examination of administrative records from the federal student financial aid applicant and award files indicates large significant differences between survey responders and non-responders in likelihood of being found on the aid files (for example 79 percent of responders to the Fourth Follow-up were found on the aid files, while 62 percent of non-responders were found on the aid files in the period of 1994-95 to 2003-04) (figure 4).

⁸ Because of some inconsistency in high school graduation year, and that fact that a portion of the Baseline Surveys were completed with reference to 1991-92 instead of 1992-93, actual tabulation was for -1 or +1 of the year of expected high school graduation; and + 4 years of expected high school graduation year.

Figure 4. Percent of total study participants found on the federal student financial aid (SFA) files as applicants and as Pell recipients, classified by Fourth Follow-up Survey response status: National Evaluation of Upward Bound, study conducted 1992-93-2003-04



NOTE: Unweighted data based on 2845 Upward Bound sample members from both treatment and control groups

SOURCE: Data tabulated October 2006 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.

These differences are taken as an indication that those who respond to the surveys were also more frequently those who had positive postsecondary outcomes. To mitigate this potential source of bias, we use a longitudinal file of all sample members and use federal student aid (SFA) applicant files,⁹ and where appropriate National Student Clearinghouse (NSC) data to impute outcome measures.

⁹ For postsecondary attendance, the aid applicant status was used as indication of enrollment rather than the Pell Award file due to the income requirement for Pell Award. However, models were also run using Pell Award criteria with similar results.

5. **Equivalent or similar service substitution and study no-show/dropout biases.** Critics of the UB evaluation have been most aware of and vocal about the issues surrounding receipt of equivalent or similar services by the control group. Analyses of the UB study randomization files and the surveys reveal that the task of keeping the treatment and control group mutually exclusive with regard to the intervention was not easy or subject to complete control by the study conductors. About 26 percent of the treatment group were coded as “Dropouts” on the *Horizons Waiting List Randomization File* at the time of being given the “offer” of UB participation or shortly thereafter. All of these cases have been kept in the Intent to Treat (ITT) grouping forming the basis for published study conclusions. At least 12 percent of the control group has evidence of participation in Upward Bound Math Science (UBMS) or UB (this may be an underestimate as the survey included only UBMS participation questions and not regular UB participation questions for the control group). UBMS was an initiative of the UB program begun in 1991, around the same time as the start of the national UB evaluation. Overall 76 percent of the sample reported some form of pre-college support services with some academic component before or after randomization over the course of the applicable surveys (60 percent of the control group and 92 percent of the treatment group) (table 4).

Published analyses to date have carefully maintained fidelity to the Intent to Treat (ITT), original treatment and control group that reflected Mathematica’s random assignment of those completing the Baseline Survey to be given the chance for the “UB opportunity.” *This paper also continues to observe this fidelity to the original random assignment groupings focusing much text discussion on the ITT analyses of those given the “UB opportunity.”* As did the Third Follow-up report, to observe effects of Treatment on the Treated (TOT) or Complier Average Causal Effect (CACE) we also include results from a two stage instrumental variable regression, designed to explicitly model selection effects using actual participation as an instrumented variable (Bloom H. 2005; Angrist, J. Imbens G. Rubin D. 1996). However, unlike the CACE analyses included in the Third Follow-up report, we include UBMS as a form of UB participation.

To help in further understanding of the study results, and the role of UB/UBMS vis-à-vis other pre-college support services, we include two new sets of quasi-experimental observational analyses. Using two-stage instrumental variables regression to help mitigate but not eliminate selection bias, we explore the association of UB/UBMS participation with key outcomes relative to those reporting only some (thought to be less intensive) “other pre-college support service participation.” In another set of models, we compare those with no reported pre-college support service participation with those who reported any type (UB/UBMS or any other) of reported service participation.

Our consideration of these issues was influenced by Heckman, Hohman, Smith, and Khoo (2000) reanalysis of the Job Training Partnership Act (JTPA) studies in which they considered the interpretation of evidence from social experiments when persons randomized out of a program being evaluated have good substitutes for it, and when persons randomized into a program do not enter the program or drop out. They note that “evidence that one program is ineffective relative to close substitutes is not evidence that the type of service provided by all of the programs is ineffective, although that is the way experimental evidence is often interpreted” (Heckman et. al. 2000). Our approach in this paper is to try to present as much observational information on the extent of equivalent or similar services receipt, and the extent of non-participation, as is available and to include additional modeling of the associations between services and outcomes to complement the Intent to Treat (ITT) analyses of those randomly assigned to be given the UB opportunity.

Table 4. Number and percent of study sample participating in UB or UBMS and other pre-college support or supplemental service programs with academic components, by treatment and control group status: National Evaluation of Upward Bound, study conducted 1992-93-2003-04						
	Random Assigned Treatment		Random Assigned Control		Total Horizons Study	
	Unweighted	Poststratified Weighted	Unweighted	Poststratified Weighted	Unweighted	Poststratified Weighted
Total	1,524 (100%)	21,866 (100%)	1,320 (100%)	21,866 (100%)	2,844 (100%)	43,732 (100%)
Reported participated in UB or UBMS service	1,247 (82%)	17,843 (82%)	180 (14%)	2702 (12%)	1,427 (50%)	20,545 (47%)
Reported participated in “another” (not UB and not UBMS) pre-college support or supplemental service program only	128 (8%)	2,332 (11%)	618 (47%)	10,513 (48%)	746(26%)	12,845 (29%)
Did not report participation in any type of (UB or UBMS or any other) pre-college support or supplemental service program	149 (10%)	1690 (8%)	522 (40%)	8651 (40%)	671 (24%)	10,342 (24%)
Reported participated in any type (UB, UBMS or any other) of pre-college support or supplemental service program	1375 (90%)	20,176 (92%)	798 (61%)	13,215 (60%)	2173 (76%)	33,390 (76%)
NOTE: Percents given in parentheses. UB = Upward Bound; UBMS = Upward Bound Math Science. Weighted data use poststratified weights for longitudinal file.						
SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education,; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.						

IV. Method and Results of the Re-Analyses

In this paper, we follow similar statistical procedures as followed in published and draft Mathematica reports.¹⁰ However, instead of using a file made up only of responders to a given round of the survey follow-ups with non-response adjusted weights as was done for the published Third Follow-Up report and the unpublished Fourth Follow-up report, we constructed a longitudinal file containing all cases in the sample.¹¹ As was done in previously published reports, we conducted regression analyses that allow inclusion of some additional variables from the Baseline Survey designed to correct for some of the differences between the treatment and control groups. Standard errors for multi-stage weighted data were tabulated using STATA software taking into account the complex sample design used in the study. For reasons noted above we include analyses of both weighted and unweighted data and results with and without the questionable “outlier” project 69. The tables presenting results give the model estimated expected rates for the treatment and control group for the dependent variables of interest, the difference between treatment and control, and the significance levels. Examples of complete model results are presented appendix tables A and B and additional complete results are available upon request. In text discussion for logistic regression results, we also provide the odds ratio that can be used as a measure of effect size for the dichotomous variables of interest to this study, and this may be calculated with the information given in the tables.¹² Controls used in the models were grade in reference year; low-income status, first generation status; grade on student selection form; baseline educational expectations; race/ethnicity; sex; and past participation in pre-college services. Table 1 gives descriptive statistics on these variables for the sample.

Three types of comparisons are noted below. The Intent to Treat (ITT) estimates use logistic regression to model the impact of being randomly assigned to the treatment or control group independent of whether the sample member actually entered into the treatment. It is most properly thought of as the impact of being given the “Upward Bound Opportunity”. The Treatment on the Treated (TOT) or Complier Average Causal Effect (CACE) analyses, uses instrumental variables regression to control for selection effects, and estimates the impacts of actual participation in the Upward Bound service. Two stage instrumental variables regression first estimates the relationship of the variables in the model to participation in UB/UBMS; and then uses this factor

¹⁰ Mathematica describes their method in a footnote to the unpublished Fourth Follow-up report with regard to the logit models: “In practice, we estimated program effects by estimating a regression model with the outcome of interest as the dependent variable. . . .The independent variables included indicators of students’ random assignment status (treatment status), sex, race, ethnicity, educational expectations at baseline, grade at application, and low-income or first-generation status.”

¹¹ The Third Follow-up report was published in 2004; the Fourth Follow-up report was an unpublished internal draft. The Fifth Follow-up report was under Department of Education review as of November of 2008.

¹² The odds ratio is a measure of effect size. It is defined as the ratio of the odds of an event occurring in one group to the odds of it occurring in another group, or to a sample-based estimate of that ratio. If the probabilities of the event in each of the groups are T (treatment group) and C (control/comparison group), then the odds ratio may be tabulated as follows: $T(1-c)/C(1-T)$. An odds ratio of 1 indicates that the condition or event under study is equally likely in both groups. An odds ratio greater than 1 indicates that the condition or event is more likely in the first group. And an odds ratio less than 1 indicates that the condition or event is less likely in the first group. The odds ratio must be greater than or equal to zero. As the odds of the first group approaches zero, the odds ratio approaches zero.

in the second stage regression to control for but not eliminate the selection effects in the model. In this study, a weighted estimated 18 percent of the treatment group had no evidence of UB or Upward Bound Math Science (UBMS) participation and about 12-14 percent of the control group had evidence of participating in UB or UBMS. As this study reported that more than half of the control group members had participated in non-UB or non-UBMS supplemental pre-college services with an academic component, a third comparison, using instrumental variables regression, was also done. This comparison looked at differences between those who had evidence of UB/UBMS participation compared with those that only had evidence of some other non-UB/non-UBMS service.

Tables 5 to 12 present model based results from the logistic and instrumental variables regressions. Given the historical use of weighted data with the outlier project included to make published estimates for this study, this is the default for results discussed in the text. However, as noted above a careful review of the data suggests that the estimates without project 69 are less subject to the effects of the large relative weights and also to treatment and control group differences and in this sense can be considered to be more robust.

Evidence of Postsecondary Attendance by +1 (18 months) and +4 of Expected High School Graduation Year (EHSGY). Table 5 shows the results of the ITT logistic and TOT instrumental variables regression models, for the outcome variable of having evidence of postsecondary entrance, within +1 (about 18 months) and +4 of expected high school graduation year (EHSGY). For these estimates, we used all applicable follow up surveys and the federal student financial aid (SFA) files. National Student Clearinghouse (NSC) institutional coverage was reported to be at 25 percent in 1996 and for this reason we have not included this data in our estimates of postsecondary entrance evidence. There are undoubtedly some survey non-respondents who did not apply for federal aid, but who did enroll in postsecondary that are non-responders and may be listed in the NSC files who are missed by not using this data. However, evidence from the UB performance reports indicates that about 40 percent of UB participants enroll in the grantee institutions.¹³ Therefore any lack of coverage or sporadic coverage can introduce serious bias into the estimates and this fact outweighs arguments in favor of using this source for postsecondary enrollment in this study. We know, for example, that the project with the largest weight (69) did not participate in NSC at all until after all of its sample members would have graduated. As there are no alternative sources for degrees for non-respondents, we do, however, include NSC data in estimates of BA degree attainment occurring when NSC participation had increased their coverage.

¹³ For example, 45 percent of UB participants who enrolled in postsecondary from 2-year grantee hosts, and 41 percent from 4-year public grantee hosts attended the grantee host grantee institution (Cahalan and Curtin 2004).

Overall about 68 percent of the study sample showed evidence of entering postsecondary within +1 (within two fall starts of the academic year or about 18 months following a June high school graduation) and about 70 percent within +4 of EHSGY. We note for context purposes that the rates of college going among the 1992-2004 UB Evaluation sample are not typical of rates for the US in general and especially for low-income students in the United States in the period. For example, the National Cooperative Higher Education Management System (NCHEMS) national estimate across the total US population (all income groups) of the “chance for entering postsecondary by age 19” was 39.7 percent in 1996 and 38.8 in 1998 and 38.0 in 2004.¹⁴ Census Bureau Current Population Survey (CPS) estimates for the percent of *high school graduates* (excludes high school dropout population) who entered postsecondary by October following high school graduation was 63 percent in 1995 for the total US population and was 41 percent for low-income population (U.S. Department of Commerce, Bureau of the Census, October Current Population Surveys, 1972–2000). The National Educational Longitudinal Study (NELS:88) following a nationally representative cohort of 8th grade students found that by the year 2000, --8 years after expected high school graduation date, about 76 percent of the cohort reported some postsecondary attendance and among low SES students the percentage was 52.1 percent (Ingles et. al. 2002). We note this only to establish that college going rates for those applying to the study waiting list (treatment and control group) were generally higher than those reported for the US population as a whole in the period and were seemingly much higher than for low-income population.¹⁵

ITT Estimates. All model based estimates (weighted and unweighted and with and without the outlier project 69) of differences in the probability of having evidence of postsecondary entrance for those randomly assigned to be given the “UB opportunity” compared to those not given the opportunity (Intent To Treat or ITT estimates) were statistically significant (table 5). The weighted ITT estimated probably (pr) was 6.9 percentage points higher for those given the UB opportunity compared to those not given the UB opportunity ---an increase from 66.0 to 72.9 for having evidence of entering postsecondary within +1 (about 18 months following a June graduate) year of EHSGY. Appendix table A provides a detail of model results for this ITT estimate. Using an odds ratio to convey effect size this indicates that the odds of postsecondary were 1.48 times what they would have been without being given the “UB opportunity.” Removing the bias introducing outlier project 69, there was a 9.1 percentage point difference (64.3 to 73.3) between treatment and control group for the ITT estimates (odds ratio of 1.68 times what they would have been without the “opportunity”). ITT effects for the outcome variable evidence of entering postsecondary within +4 of EHSGY were 6.1 for the entire sample and 8.3 without project 69.

¹⁴ The Third Follow-up UB report using data from survey responders only with a non-response adjustment to the weights estimated that 72 percent of the total sample reported entering postsecondary by 1998-99. Estimates based on “survey responders only” are slightly higher than those based on a longitudinal file of all sample members rather than only survey responders.

¹⁵ It should also be noted that the number served by UB/UBMS per year (about 63,000) is about 1.7 percent of the number of 9-12 graders reported eligible for free lunch per year in the period (3.6 million) (U.S. Department of Education, Federal TRIO Programs, Upward Bound performance reports, 2000–01).

Table 5. Evidence of entering postsecondary within +1 (18 months) and within +4 of expected high school graduation year (EHSGY) for ITT and TOT models: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04

	All sampling strata		One outlier project removed (the remainder represents 74 percent of Horizons waiting list)	
	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)
Evidence of postsecondary entrance within +1 (18 months) of EHSGY	pr-T = 72.9 pr-C = 66.0 Difference = 6.9**** (pr T = 74.7 pr C = 67.9 Difference = 6.8****)	xb T = 73.5 xb C = 62.5 Difference = 10.9**** (xb T = 75.7 xb C = 64.6 Difference = 11.1****)	pr T = 73.3 Pr C = 64.3 Difference = 9.1*** (pr T = 74.8 pr C = 67.6 Difference = 7.2***)	xb T = 74.6 xb C = 60.4 Difference = 14.2**** (xb T = 75.8 xb C = 64.2 Difference = 11.6****)
Evidence of postsecondary entrance within +4 of EHSGY	pr-T = 74.5 pr-C = 68.3 Difference = 6.1*** (prT = 76.8 prC = 70.3 Difference = 6.5****)	xb T = 74.7 xb C = 64.6 Difference = 10.1*** (xb T = 77.6 xb C = 66.7 Difference = 10.8****)	pr T = 75.4 pr C = 67.0 Difference = 8.3*** (pr T = 77.0 pr C = 70.1 Difference = 6.9****)	xb T = 76.2 xb C = 62.7 Difference = 13.5*** (xbT = 77.8 xbC = 66.4 Difference = 11.4****)

*/**/**** Significant at 0.10/0.05/.01/00 level; NS = not significant at the .10 level or below. UB = regular Upward Bound; UBMS = Upward Bound Math Science; ITT = Intent To Treat; TOT = Treated On Treated; CACE Complier Average Causal Effect (CACE); T = Treatment; C = Control or comparison; pr = estimated probability from STATA logit regression; xb = linear prediction from STATA ivreg instrumental variables regression;

NOTE: Unweighted results are given in parentheses. Weights tabulated by Mathematica Policy Research and used in these analyses were poststratified to equalize treatment and control group per project and to reflect different probabilities of selection within projects. Unequal weighting in this study lead to large design effects and reduces the effective sample size. There were 46 first stage strata, several of which were represented by only one project, including one project that represented 26 percent of the total weight. In addition there were substrata within projects so that there were 339 total strata. Sensitivity analysis of the one project stratum with the unusually large weight (project 69) revealed that there were also treatment-control group non-equivalencies in favor of the control group on variables known to be related to the outcomes of interest to the study. For these reasons model results are presented with and without project 69 considered to be an outlier. All models used in these analyses use STATA taking into account the complex sample design. The models also attempt to control for baseline treatment/control group differences by including baseline variables including sex, grade at baseline, educational expectations, race/ethnicity, eligibility status low income or first generation only, participation prior to random assignment, grade at entrance into program. Appendix tables A and B give examples of actual model results for weighted results. **ITT (Intent to Treat)—logistic regression** is based on the original random assignment grouping and represents those randomly assigned to be given or not given the UB opportunity by projects. Note about 26 percent of the treatment sample were coded as having dropped-out of the study waiting list at the time of being given the UB opportunity. Based on First Follow-up Survey data, about 18 percent of the weighted treatment group reported not ever having participated in any UB activity, and about 14 percent of the control group had evidence of UB or UBMS participation. **TOT (Treatment on Treated) sometimes referred to as Complier Average Causal Effect (CACE)** tabulated using STATA, svy: ivreg. To control for participation selection effects, this two stage regression first models participation using the random assignment variable as an instrument and then in the second stage models the effect of participation. **Evidence of postsecondary** calculated based on reported year of first postsecondary entry as reported on any of the UB applicable follow-up surveys or for non-responders to all surveys being found on the student aid files as having applied for aid within +1 year (about 18 months) and within +4 years of expected high school graduation year (EHSGY). Odds ratio for logistic regression results may be tabulated by $\frac{prT(1-prC)}{prC(1-prT)}$.

SOURCE: Data tabulated December 2007 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.

TOT/CACE Estimates. For the instrumental variables regressions, estimating the effect of actual participation in UB or UBMS, the differences (10.9 percentage points) between participants and non-participants for evidence of postsecondary entrance within +1 of EHSGY was significant (62.5 to 73.5) for the complete weighted sample. Differences were 14.2 percentage points for the estimates without project 69 (60.4 to 74.6).

Comparison Relative to Third Follow Up Conclusions

The Third Follow-up report (Myers et. al 2004, table III.1) estimated that enrollment in any postsecondary school based on Third Follow-up (conducted in 1998-99) survey data alone adjusted for non-response was 71 percent for the control group and 74 percent for the treatment group by the time of the third follow up conducted in 1998-99. The difference was not statistically significant. Their analyses differs from the one presented above in using only survey data, in using a file made up only of responders to the survey and not using a longitudinal file, not using administrative records, and not standardizing the outcome variable by expected high school graduation year (EHSGY).

The question can be legitimately raised as to whether had Mathematica included their same analyses based on only survey responders and unstandardized, without the bias introducing project 69 would they have arrived at different conclusions or at least seriously qualified conclusions in that report. The indications are that the answer is yes. Table 6 below presents a number of weighted relevant comparisons for postsecondary entrance with and without project 69 using a similar but not exactly the same model to the one forming the basis of the Third Follow-up report. The first three rows are based on Third Follow-up responders only, with the weights adjusted for non-response as was done in the published report. The first row does not include any standardization for EHSGY and also does not use any other survey rounds of information and does not use administrative records. The second row adds standardization to 18 months but relies only on responders to the Third Follow-up with the non-response adjusted weight. The third row also based on Third Follow-up responders only, adds use of other surveys (not all sample members had 18 months after EHSGY by the time of the Third Follow-up) and aid file information. The last row includes the complete longitudinal sample file of responders and non-responders to the survey, includes information from all applicable surveys, uses the SFA files and also standardizes the outcome measures. It is the same result as in table 5. The first thing to notice is that all models in the table are significant when project 69 is excluded. Effect sizes and significance levels become seemingly larger as more information and precision is added to the tabulation of the outcome measures, and as the estimates are based on both responders and non-responders to the surveys. Other models using data from the Fourth Follow up show similar patterns (data not shown).

Table 6. Various model results using Third Follow Up Survey responders only and using full longitudinal sample for evidence of entering postsecondary for ITT and TOT models: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04

	All sampling strata		One project with bias removed (the remainder represents 74 percent of Horizons waiting list)	
	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)
Third Follow-up survey responders only with no administrative records and no standardization of outcome to expected high school graduation year (EHSGY); uses non-response adjusted weight f3wgtstu (psu3)	pr-T = 76.4 pr-C = 75.4 Difference = 1.0NS	xb T = 75.4 xb C = 71.7 Difference = 3.7 NS	pr T = 77.8 Pr C = 72.2 Difference = 5.7**	xb T = 77.6 xb C = 67.7 Difference = 9.9*
Third Follow-up survey responders only with no administrative records or other applicable surveys, but with standardization to +1 (18 months) of expected high school graduation year uses non-response adjusted weight f3wgtstu ktbpe18	pr-T = 71.2 pr-C = 68.2 Difference = 3.0 NS	xb T = 71.4 xb C = 65.2 Difference = 6.1 NS	pr T = 73.3 Pr C = 65.8 Difference = 7.5***	xb T = 74.0 xb C = 61.9 Difference = 12.1***
Third Follow-up survey responders only – standardized to +1 (18months) of EHSGY and uses all applicable surveys and Student Financial Aid (SFA) records ---keyne2 Uses non-response adjusted weight f3wgtstu	pr-T = 75.9 pr-C = 71.4 Difference = 4.6*	xb T = 76.0 xb C = 67.8 Difference = 8.2 NS.11	pr T = 77.8 Pr C = 70.0 Difference = 7.8****	xb T = 78.2 xb C = 65.6 Difference = 12.6***
Includes all sample members, standardized to +1 (18months) of EHSGY and uses all applicable surveys and SFA records ---keyne2 Keyne2—Uses poststratified weight v5bwgtp1	pr-T = 72.9 pr-C = 66.0 Difference = 6.9****	xb T = 73.5 xb C = 62.5 Difference = 10.9****	pr T = 73.3 Pr C = 64.3 Difference = 9.1***	xb T = 74.6 xb C = 60.4 Difference = 14.2****
<p>*/**/****/***** Significant at 0.10/0.05/.01/00 level; NS = not significant at the .10 level or below. UB = regular Upward Bound; UBMS = Upward Bound Math Science; ITT = Intent To Treat; TOT = Treated On Treated; CACE Complier Average Causal Effect (CACE); T = Treatment; C = Control or comparison; pr = estimated probability from STATA logit regression; xb = linear prediction from STATA ivreg instrumental variables regression.</p> <p>NOTE: Please see table 5 for detailed notes</p> <p>SOURCE: Data tabulated December 2007 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education,; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.</p>				

Observational Models of UB/UBMS Compared to Other Services

Comparing UB/UBMS participants with those sample members receiving another (presumably less intensive) non-UB/non-UBMS service, there is a significant 9.1 percentage point difference for +1 of EHSYG entrance in the weighted instrumental variables regression estimates with project 69, and a 13.3 percentage point difference without project 69 using the same controls as used in the previous models (table 7). Models estimating the association between reporting any pre-college support or supplemental services and entering postsecondary within +1 were highly significant and showed large differences of 25 percentage points for the total sample and 30 percentage points for the sample without the outlier project (table 7). It should be noted that the later estimates would seem to be especially subject to selection bias related to seeking services.

Table 7. Evidence of Postsecondary Entrance within +1 (18 months) and within +4 of expected high school graduation year (EHSYG for observational models comparing types of service receipt: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-2004				
	All sampling strata		One outlier project removed (remainder represents 74 percent of Horizons waiting list)	
	Participated in UB/UBMS compared with participated in other non-UB/non-UBMS pre-college support or supplemental services only (observational – instrumental variables regression)	Any pre-college support or supplemental services reported compared with no services reported (observational – instrumental variables regression)	Participated in UB/UBMS compared with participated in other non-UB/non-UBMS pre-college support or supplemental services only (observational – instrumental variables regression)	Any pre-college support or supplemental services reported compared with no services reported (observational – instrumental variables regression)
Evidence of postsecondary entrance within +1 of EHSYG	xb T = 74.4 xb C = 65.3 Difference = 9.1*** (xb T = 76.2 xb C = 66.8 Difference = 9.3****)	xb-T = 73.5 xbC = 48.6 Difference = 25.0**** (xb T = 75.8 xb C = 51.7 Difference = 24.1****)	xb T = 75.0 xb C = 61.7 Difference = 13.3**** (xb T = 76.3 xb C = 66.3 Difference = 10.1****)	xb T = 74.3 xb C = 44.6 Difference = 29.8**** (xb T = 75.9 xb C = 51.1 Difference = 24.7****)
Evidence of postsecondary entrance within +4 EHSYG	xb T = 75.6 xb C = 67.5 Difference = 8.2*** (xb T = 78.2 xb C = 68.7 Difference = 9.5****)	xb-T = 74.8 xb-C = 51.4 Difference = 23.5*** (xb T = 77.7 xb C = 54.1 Difference = 23.6****)	xb T = 76.5 xb C = 64.4 Difference = 12.1**** (xb T = 78.4 xb C = 68.2 Difference = 10.2****)	xb T = 75.9 xb C = 47.8 Difference = 28.1**** (xb T = 77.8 xb C = 53.7 Difference = 24.1****)

*/**/**** Significant at 0.10/0.05/.01/00 level
 UB = regular Upward Bound; UBMS = Upward Bound Math Science; T = Treatment; C = Control or comparison; xb = linear prediction from STATA ivreg instrumental variables regression. Odds ratio = $\frac{prT(1-prC)}{prC(1-prT)}$.
NOTE: Unweighted data given in parentheses. Please see table 5 for detailed notes.
SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.

Presence on the Federal Aid Application Files

Overall 57 percent of the entire sample (treatment and control) was found on the federal student financial aid (SFA) files as applicants within +1 of the EHSGY and 63 percent was found on the file within +4 of the EHSGY (data not shown in tables). Table 8 gives model results using only presence on the SFA files as the outcome variable. The “applied for aid” outcome variable, while not a measure of postsecondary entrance itself, in addition to being an important outcome for the program, is of considerable value to the study. Among the study measures we have available to us, it is the least subject to the non-response bias that is a factor with regard to the survey responder data; or to the non-coverage bias that is a serious concern with using the National Student Clearinghouse (NSC) data especially for these years. As such it provides some validation for the results presented above in table 5, which remain subject to response bias. This measure is also of interest as an outcome measure in itself, as it pertains to one of the major legislative goals of the TRIO programs. TRIO programs are authorized under the same legislation as the federal aid provisions, to assist in making low-income students aware of and prepared to use the federal aid program.

Table 8. Appearance on the federal student financial aid (SFA) files within +1 (18 months) and within +4 of expected high school graduation year (EHSGY) for ITT and TOT/CACE models: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04

	All sampling strata		One outlier project removed (remainder represents 74 percent of Horizons waiting list)	
Outcome variable	Given UB Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)	Given UB Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)
On federal student financial aid (SFA) file as applicant within +1 of EHSGY	pr-T = 61.6 pr-C = 55.9 Difference = 5.7*** (pr T = 65.6 pr C = 59.6 Difference = 6.1****)	xb T = 62.9 xb C = 53.6 Difference = 9.3**** (xb T = 67.0 xb C = 57.2 Difference = 9.8****)	pr T = 64.5 pr C = 56.4 Difference = 8.1**** (pr T = 66.0 pr C = 59.7 Difference = 6.3****)	xb T = 66.2 xb C = 53.6 Difference = 12.6**** (xb T = 67.3 xb C = 57.2 Difference = 10.1****)
On federal student financial aid (SFA) file as applicant within +4 of EHSGY	pr-T = 65.4 pr-C = 58.7 Difference = 6.7**** (prT = 69.1 pr C = 63.3 Difference = 5.8****)	xb T = 66.7 xb C = 56.1 Difference = 10.6**** (xb T = 70.2 xb C = 60.5 Difference = 9.7****)	pr T = 67.7 pr C = 60.4 Difference = 7.3*** (pr T = 69.4 pr C = 63.6 Difference = 5.8****)	xb T = 69.1 xb C = 57.1 Difference = 11.9*** (xb T = 70.4 xb C = 60.7 Difference = 9.7****)

*/**/**** Significant at 0.10/0.05/.01/00 level; NS = not significant at the .10 level or below.
NOTE: Unweighted data given in parentheses. Please see table 5 for detailed notes. UB = regular Upward Bound; UBMS = Upward Bound Math Science; ITT = Intent To Treat; TOT = Treated On Treated; CACE Complier Average Causal Effect (CACE); T = Treatment; C = Control or comparison; pr = estimated probability from STATA logit regression; xb = linear prediction from STATA ivreg instrumental variables regression.
SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education.; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.

ITT and TOT/CACE Estimates. Evidence in table 8 indicates that the UB program was effective in increasing the rates at which low-income students applied for federal aid. Being given the UB opportunity (ITT estimate) resulted in about a 5.7 percentage point increase in likelihood of being found on the aid files within +1 of EHSYG and a 6.7 percentage point increase in being found within +4 of EHSYG. When project 69 is removed the effects were 8.1 and 7.3 percentage points respectively. The effects of TOT UB/UBMS participation estimates were 9.3 and 10.6 percentage points for being found on the aid files within +1 and +4 of EHSYG respectively. Without the outlier project, differences were 12.6 and 11.9 percentage points respectively (table 8).

Degree or Credential Attainment

While about 70 percent of the *Horizons* sample reported evidence of postsecondary attendance within +4 of EHSYG, by the Fourth Follow Up (2001-02), about 30 percent of the entire *Horizons* sample reported or had evidence of some type of postsecondary credential or degree. Data from the Baseline Survey indicates that 72 percent of all sample members reported they expected to attain at least a bachelor's degree. Descriptively (not comparing treatment and control groups), we found that about 20 percent of the entire *Horizons* sample had evidence of a BA in +6 of EHSYG. We note for context purposes that the National Cooperative Higher Education Management System (NCHEMS) 1998 national estimate across the total (all income groups) US population of 9th graders was 17 percent for degree completion within 150 percent of program time for those going directly into college from high school.¹⁶ The National Educational Longitudinal Study (NELS: 88) following a nationally representative cohort of 8th grade students found that by the year 2000, --8 years after expected high school graduation date-- about 26 percent of the entire cohort reported obtaining a bachelor's degree and among lowest SES quartile the percentage was estimated to be 6.9 percent. Among the NELS:88 students, whose parent's had no college, the percentage was 11 percent (Ingles et.al. 2002).

Table 9 presents model results containing three postsecondary degree or certificate related measures: 1) any postsecondary degree or credential by the end of the Fourth Follow Up (not standardized by EHSYG) based on Fourth Follow Up Survey responders only, with non-response adjusted weights; 2) a bachelors degree by the end of the Fourth Follow-Up study period based on survey responses only adjusted for non-response but not standardized noting that most sample participants ranged from 4 to 6 years out of expected high school; 3)

¹⁶ The National Cooperative Higher Education Management System (NCHEMS) Information Center using NCES Common Core Data, NESC IPEDS Residency and Migration Survey, and NCES IPEDS Graduation Rate Survey estimates that the percentage of 9th graders who graduate from high school on time, go directly to college, return for second year and graduates within 150 percent of program time was 17 percent in 1998; and 18.4 in 2004.

evidence of a bachelor’s degree award within +6 of EHSYG based on responses to all applicable follow-up surveys, SFA file information, and NSC information.

Table 9. Evidence of attainment of any postsecondary degree or certificate and attainment of a BA within +6 of expected high school graduation year (EHSYG) for ITT and TOT models: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04				
	All study sampling strata		One outlier project removed (remainder represents 74 percent of Horizons waiting list)	
	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)
Evidence of attainment of any postsecondary degree or credential by end of study period—Fourth Follow-up Survey data only with non-response adjustment				
Fourth follow-up survey responders only—evidence of any degree; weighted data uses non-response adjusted weight	pr-T = 31.2 prC = 26.9 Difference = 4.3* (pr T 33.0 prC = 28.3 Difference = 4.7**)	prT = 33.1 pr C = 26.4 Difference = 6.7* (pr T = 35.1 pr C = 27.7 Difference = 7.4*)	pr T = 30.3 pr C = 25.9 Difference = 4.4 NS .12 (pr T = 32.9 pr C = 27.9 Difference = 4.0 **)	Xb T = 32.8 xb C = 25.7 Difference = 7.1 NS .14 (xbT = 35.1 xb C = 27.3 Difference = 7.8**)
Fourth follow-up responders only evidence of a BA degree---not standardized---responders only using non-response adjusted weight				
Fourth Follow-up survey responders only—evidence BA degree, not standardized by EHSYG; weighted data uses non-response adjusted weight	pr-T = 13.7 prC = 11.2 Difference = 2.5 NS .14 (pr T 15.3 prC = 12.8 Difference = 2.5**)	prT = 17.8 pr C = 13.2 Difference = 4.6 . NS .14 (pr T = 18.8 pr C = 13.8 Difference = 5.0**)	pr T = 12.8 pr C = 8.9 Difference = 3.9**** (pr T = 15.1 pr C = 12.2 Difference = 2.9**)	Xb T = 17.7 xb C = 10.4 Difference = 7.3*** (xbT = 18.8 xb C = 13.2 Difference = 5.6**)
Evidence of BA in +6 of EHSYG--- All applicable follow-up surveys, Pell Award Files, NSC—responders and non-responders included –longitudinal file				
Uses all applicable follow-up surveys, NSC, and Pell graduation variable; standardized to EHSYG ;longitudinal file poststratified weight.	pr-T = 16.9 pr-C = 16.0 Difference = .9 NS (pr T = 18.4 pr C = 16.1 Difference = 2.3**)	xb T = 19.7 xb C = 17.4 Difference = 1.7 NS (xb T = 21.4 xb C = 16.6 Difference = 4.8**)	pr T = 17.0 pr C = 13.3 Difference = 3.7**** (pr T = 18.3 pr C = 15.6 Difference = 2.7***)	Xb T = 21.1 xb C = 14.1 Difference = 7.0**** (xb T = 21.6 xb C = 16.1 Difference = 5.5***)
<p>*/**/***/**** Significant at 0.10/0.05/.01/00 level; NS = not significant at the .10 level or below.</p> <p>UB = regular Upward Bound; UBMS = Upward Bound Math Science; BA = bachelor’s degree; ITT = Intent To Treat; TOT = Treated On Treated; CACE = Complier Average Causal Effect (CACE); T = Treatment; C = Control or comparison; NSC = National Student Clearinghouse; pr = estimated probability from STATA logit regression; xb = linear prediction from STATA ivreg instrumental variables regression.</p> <p>NOTE: Unweighted data in parentheses. Please see table 5 for detailed notes. Unweighted estimates for survey only estimates do not have a non-response adjustment. Fourth Follow-up Survey conducted in 2001.</p> <p>SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; federal Student Financial Aid (SFA) files 1994-95 to 2003-04; and National Student Clearinghouse Data 1995-2004.</p>				

As can be observed in table 9, the differences between treatment and control group on educational expectations and academic risk introduced by project 69 and other unequal weighting are manifest in these results. They also reflect the fact that those who have chosen to work on a bachelor’s degree, taking at least 4-years, would not have completed this degree as quickly as those seeking a lesser degree or certificate. The fact that project 69 had

a focus on less than four year degrees, is manifest in that there is little difference in the impact estimates with and without 69 for award of any degree; while for award of the bachelor's degree weighted results with 69 are insignificant, while those without 69 show significant and substantial differences in favor of the treatment group. It should be kept in mind that the project 69 control group had substantially higher educational expectations than did the treatment group (56 percent expected an advanced degree compared with 15 percent of the treatment group); and that overall there was a 12 percentage point difference between the treatment and control group in favor of the controls in percent expecting an advanced degree at baseline.

For evidence of bachelor's degree attainment in +6 years of EHSGY we used longitudinal file with data from all of the applicable follow-up surveys (Second through Fifth), a Pell award file expected bachelor's degree graduation within the year variable, and the NSC data. Those having evidence of BA attainment from any one of these sources were considered to have evidence and those with no evidence were considered "no's." Descriptively (not comparing treatment and control groups), as noted above, we found that about 20 percent of the entire *Horizons* sample had evidence of a BA in +6 of EHSGY.

ITT and TOT/CACE. Considering the BA within +6 of EHSGY outcome measure, we see the role that the non-equivalencies in educational expectations and academic preparation and large weights introduced by project 69 may be playing. Weighted estimates with project 69 are not significant. All unweighted comparisons are significant and weighted estimates without project 69 are significant. For the 74 percent of the sample weight not represented by project 69, the estimated probability was 17.0 percent for the treatment group and 13.3 percent for the control group for the ITT comparison (a 28 percent increase), and for the TOT the estimated probability was 21 for the treatment and 14 for the control group (a 50 percent increase) (table 9).

Selected Sub-Group Results

One of the key findings discussed in the Third Follow-up Report, was that the program demonstrated significant and substantial effects for those participants who were in the bottom 20 percent of the sample on 9th grade academic achievement; and also for those reporting on the Baseline Survey that they did not expect to obtain a bachelor's degree. In this paper, we also discuss these same groupings as they were used for the analyses presented in previously published reports, and have influenced ED policy development.

Some Special Limitations with Regard to the Sub-Group Analyses Based on Academic Risk and Educational Expectations. There are a number of additional cautions that should be mentioned in reviewing results with regard to these subgroups that are more subject than the total sample to the uneven weights and the design effects of the sample. Sample sizes for the sub-groups are substantially smaller. The estimates for those

in the bottom 20 percent on academic risk are based on unweighted n's of 521 cases and those not expecting at least a BA degree are based on 718 unweighted cases (overall of the total sample, 28 percent weighted cases did not report they expected at least a BA at baseline). We also note that the ITT treatment and control groups no longer have equal weighted totals per project, but reflect differences (some introduced by the poststratification weight adjustments) between treatment and control on these variables. Using the study weights 24 percent of the treatment group and 19 percent of the control group were in the high-academic-risk group; and 32 percent of the treatment group and 25 percent of the controls were in the lower education expectation sub-group).

Another factor to keep in mind in looking at the data on sub-groups based on expectations is that, as has been noted, the students were in different grades in the “last academic year” (7 to 10) before they completed the Baseline Survey—so the models suffer from the fact that these expectations were expressed at different grades for different students. It should also be noted that almost all of the *Horizons* applicants expected some type of postsecondary credential at the time of completion of the baseline. Overall about 72 percent of the sample reported they intended to obtain at least a bachelor's (BA) degree on the Baseline Survey, and most of the 28 percent with below BA expectations, expected some form of less than BA postsecondary degree. On the Baseline Survey, 3 percent of the entire sample indicated that high school completion was as far as they expected to go in school. The academic risk measure from student transcripts is keyed to a specific 9th grade, but students differed in whether they had been randomized before or after the 9th grade. This measure also suffers from missing data, as it is available only on those students for whom high school transcripts were obtained.

Subgroup Results for Evidence of Postsecondary Entrance. Table 10 gives significance and difference levels for ITT and TOT predicted rates for evidence of postsecondary within +1 of EHSGY for those deemed to be at higher academic risk (in the bottom 20 percent of the sample in 9th grade on academic indicators) and those of lower risk (in the top 80 percent of the sample in 9th grade). Keeping in mind the limitations noted above, the results in table 10 suggest confirmation of the published results in the Third Follow-up Report that there were significant and substantial effects on the postsecondary entrance outcomes for those in the bottom 20 percent of the academic achievement indicators in the 9th grade (high academic risk students). Significant but moderate effects were found in this re-analysis for those in the top 80 percent. While not shown here, similar results were obtained comparing subgroups based on higher and lower expectations with regard to expectation of a bachelor's degree.

Table 10. Evidence of entering postsecondary within +1 (18 months) of expected high school graduation year (EHSGY) for sample members with higher academic risk (bottom 20 percent) and lower academic risk (top 80 percent) for ITT and TOT models: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04				
	All sampling strata		One outlier project removed (remainder represents 74 percent of Horizons waiting list)	
	Given UB Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)	Given UB Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)
Evidence of postsecondary entrance within +1 of EHSGY				
Among students with higher risk (bottom 20 percent of academic achievement in 9 th grade)	pr-T = 60.1 pr-C = 41.0 Difference = 19.1*** (pr T = 62.1 pr C = 46.5 Difference = 15.6. ****)	xb T = 65.8 xb C = 39.7 Difference = 26.3*** (xb T = 65.8 xb C = 44.3 Difference = 21.5****)	pr T = 58.0 Pr C = 44.1 Difference = 13.8**** (pr T = 61.8 pr C = 46.7 Difference = 15.1****)	xb T = 60.6 xb C = 43.0 Difference = 17.6*** (xb T = 65.4 xb C = 44.6 Difference = 20.9****)
Among students with lower risk (top 80 percent of academic achievement in 9 th grade)	pr-T = 80.1 pr-C = 73.9 Difference = 6.2**** (prT = 80.1 prC = 75.2 Difference = 5.6****)	xb T = 79.9 xb C = 70.3 Difference = 9.5*** (xb T = 81.1 xb C = 72.1 Difference = 9.0****)	pr T = 80.5 pr C = 71.9 Difference = 8.6**** (pr T = 80.9 pr C = 74.8 Difference = 6.1****)	xb T = 80.1 xb C = 67.7 Difference = 13.2**** (xbT = 77.3 xbC = 66.5 Difference = 10.8****)
<p>*/**/**** Significant at 0.10/0.05/.01/00 level; NS = not significant at the .10 level or below. UB = regular Upward Bound; UBMS = Upward Bound Math Science; BA = bachelor's degree; ITT = Intent To Treat; TOT = Treated On Treated; CACE = Complier Average Causal Effect (CACE); T = Treatment; C = Control or comparison; pr = estimated probability from STATA logit regression; xb = linear prediction from STATA ivreg instrumental variables regression. NOTE: Students with higher risk were in the bottom 20 percent of academic achievement in 9th grade; Students with lower risk were in the top 80 percent of academic achievement in 9th grade based on student transcript information. See table 5 for detailed general notes. Appendix B tables give examples of actual model results. SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; federal Student Financial Aid (SFA) files 1994-95 to 2003-04; and National Student Clearinghouse Data 1995-2004.</p>				

Sub-Group Results for Evidence for a Bachelor's (BA) in +6 of Expected High School Graduation Year (EHSGY).

Table 11 gives results for the sub-groups for evidence of bachelor's degree. It should be noted that too few, three percent of the total (25 cases unweighted) of the high academic risk students in the sample obtained a BA within +6 of EHSGY to make comparisons between treatment and control group. However, for the 80 percent of the sample classified as in the lower academic risk group, significant and substantial positive effects were found for the models when run weighted and unweighted, and with and without project 69 (table 11). These results, reflect the fact that project 69 contributed a high proportion of the high academic risk weight for the treatment group (one third of the treatment group in project 69 were classified as high risk, while

only 8 percent of the control group in project 69 were so classified. By removing the high risk sub-group from the overall sample models, one is also removing some of the project 69 introduced non-equivalencies. Rates of BA in +6 for those with lower educational expectations on the Baseline Survey were also significant for Upward Bound. Rates for those expecting at least a BA degree, (higher expectations) were also significant without the inclusion of project 69, due to the presence of the larger proportion of control group members expecting an advance degree at baseline.

Table 11. Evidence of attainment of a BA within +6 of expected high school graduation year (EHSGY) for academic risk and educational expectation sub-groups for ITT and TOT models: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04				
	All sampling strata		One outlier project removed (remainder represents 74 percent of Horizons waiting list)	
	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)
Academic Risk Sub-Groups --Evidence of BA in +6 of EHSGY--- All applicable follow-up surveys, Pell Award Files, NSC				
Among students with higher risk (bottom 20 percent of academic achievement in 9 th grade)	Insufficient cases to tabulate—25 cases unweighted in this subgroup had BA in +6 evidence. Note 4 percent of the BA's in +6 years were in this subgroup.			
Among students with lower risk (top 80 percent of academic achievement in 9 th grade)	pr-T = 23.7 pr-C = 20.4 Effect = 3.3** (pr T = 24.2 pr C = 20.6 Effect = 3.6***)	xb T = 25.8 xb C = 20.7 Effect = 5.1** (xb T = 26.4 xb C = 20.4 Effect = 6.1**)	pr T = 22.6 pr C = 17.8 Effect = 4.8**** (pr T = 24.0 pr C = 20.0 Effect = 4.0***)	xb T = 25.8 xb C = 18.0 Effect = 7.8**** (xb T = 25.1 xb C = 21.1 Effect = 3.9***)
Educational Expectation Sub-Groups --Evidence of BA in +6 of EHSGY--- All applicable follow-up surveys, Pell Award Files, NSC data				
Among students with lower expectations --(expected less than a BA at baseline)	pr-T = 11.7 pr-C = 6.0 Effect = 5.7*** (pr T = 9.0 pr C = 7.4 Effect = 1.6 NS)	xb T = 16.5 xb C = 9.9 Effect = 6.6** (xb T = 11.2 xb C = 8.3 Effect = 2.8 NS)	pr T = 7.1 Pr C = 3.7 Effect = 3.3*** (pr T = 8.5 Pr C = 6.7 Effect = 1.8 NS)	xb T = 10.3 xb C = 5.1 Effect = 5.3*** (xb T = 10.7 xb C = 7.6 Effect = 3.1 NS)
Among students with higher expectations --(expected at least a BA at baseline)	pr-T = 19.0 pr-C = 19.4 Effect = -4 NS (pr T = 22.3 pr C = 18.9 Effect = 3.3**)	xb T = 20.5 xb C = 20.7 Effect = -.2 NS) (xb T = 24.9 xb C = 19.4 Effect = 5.5**)	pr T = 19.6 Pr C = 15.7 Effect = 4.0*** (pr T = 20.2 pr C = 16.6 Effect 3.6***)	xb T = 24.9 xb C = 17.4 Effect = 7.4*** (xb T = 25.3 xb C = 19.0 Effect = 6.3***)
<p>*/**/****/***** Significant at 0.10/0.05/.01/00 level; NS = not significant at the .10 level or below. UB = regular Upward Bound; UBMS = Upward Bound Math Science; BA = bachelor's degree; ITT = Intent To Treat; TOT = Treated On Treated; CACE Complier Average Causal Effect (CACE); T = Treatment; C = Control or comparison; NSC = National Student Clearinghouse; pr = estimated probability from STATA logit regression; xb = linear prediction from STATA ivreg instrumental variables regression. NOTE: Unweighted data are in parentheses. See table 5 for detailed notes. Estimates for award of the bachelor's degree in +6 are standardized to EHSGY. SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education.; study conducted 1992-93-2003-04; federal Student Financial Aid (SFA) files 1994-95 to 2003-04; and National Student Clearinghouse Data 1995-2004.</p>				

V. Conclusions and Lessons Learned

This paper has demonstrated the potential importance of study design, implementation, and analyses choices relative to study conclusions and identified several sources of error relative to the major assumptions upon which any experimental design rests—sample representation, treatment and control group baseline equivalency, equal treatment of control and treatment groups except for the treatment of interest, and mutually exclusive treatment and control groups vis-à-vis the treatment. Below we summarize the major conclusions, and discuss some lessons learned in the context of national evaluation of Upward Bound analyses, and in the context of future study designs.

1. The key conclusion that the Upward Bound program has “no detectable effect on postsecondary enrollment” should be reconsidered in the context of the PART process. The results in this paper indicate that the program demonstrated statistically significant effects on the key goals of the program: postsecondary entrance; application for financial aid; and the attainment of postsecondary credentials.
2. The experience of this UB evaluation suggests that even in rigorous random assignment studies attention must be paid to study design and implementation error issues and small differences between treatment and control groups that may bias the conclusions. In this study there should have been more transparency concerning the sample design issues, and the representation, unequal weighting and treatment-control group non-equivalency issues introduced by project 69. Reports should have included estimates presenting results with and without the problematic project 69, especially given that the conclusions of no detectable differences were known to be sensitive to this one project. Use of administrative records earlier in the study history, and more precision in tabulating outcome measures with regard to time since expected high school graduation should also have been done at an earlier stage in the analyses and reporting phase for the study. The study designers, analysts and report preparers while demonstrating a formulaic correctness of complex procedures, manifest a lack of sensitivity to the complex sampling and measurement issues faced and the threats to the validity of experimental design issues.
3. Additional analysis is possible and needed making use of a rich data set that contains detailed information collected over 6 surveys as well as project and target schools surveys. ED expects to release the data files under restricted license to interested researchers in the near future. In future longitudinal studies effort should be made to release and disseminate the data to the interested research community as each round of follow-ups is completed.

4. This study confirms that additional policy emphasis needs to be placed on increasing student postsecondary retention and completion. This study has shown, that even among low income and first generation college students who have high expectations and who are highly served (76 percent had pre-college services), few were able to attain their goal of a BA degree as measured by 6 years after high school. While over 70 percent expected to attain a BA when sampled in middle or early high school, and more than 70 percent had evidence of entering postsecondary, only about 20 percent had attained such a degree by 6 years out of high school.
5. The experience of this UB evaluation suggests that it is very difficult within the resources generally available for national studies to attempt national probability of selection estimation with planned disaggregating and coverage of special sub-groups of interest in the context of a random assignment study. Additional work is also needed to understand the implications of using complex weighted data in the context of random assignment studies.
6. The 4-year applicant grade/age-group-span (7-10), combined with the probability of selection weights, greatly increased the complexity of this study; and makes it very important to use adequate controls for EHSGY in modeling results. The issues this introduced into the study suggests that multi-grade cohorts should be avoided in any new study design.
7. The study was very carefully implemented and is an exceptionally rich source of information on the paths of a certain segment of low income and first generation college students spanning --for over half of the students-- from middle school to young adulthood. However, it needs to be kept in mind that the study is not representative of the low income and first generation college population in general. The students were already college bound (most of the students expected to go to college at the time of application) and relatively highly motivated to seek pre-college supplemental services (76 percent of the total sample reported they had some form of pre-college supplemental services either before or after random assignment).
8. The study analyses would have benefited from some inclusion of school and project variables in the analyses. These data were collected and might help inform understanding of the results.
9. While the re-analyses demonstrates the importance of attention to study sampling and non-sampling errors even with gold standard random assignment designs, this complex study was conducted with great care under difficult circumstances; and is unique in the information it provides.

The length and seriousness of these UB study evaluation methods and policy debates is a testimony to the complexity of the issues. As we look to the future in the light of the new HEOA evaluation language, it is clear that a new generation of TRIO evaluations must be designed that will involve working in partnership with stakeholders in developing designs and procedures that are feasible, useful, accurate and ethical. Caution needs to be taken making sure the conclusions are warranted and that the results reported are transparent. The concept of having clusters of projects working together utilizing their institutional research resources to engage in rigorous self and external evaluation work focused on program improvement seems to have the most promise for the future. Challenges will be to develop and implement a rigorous protocol that will in the end answer questions of use to practitioners concerning how best to use resources; how best to serve different types of students; and how to adapt programs to the ever changing secondary-postsecondary landscape.

References

- Angrist, J. Imbens G. Rubin D. (1996) *Identification of Causal Effects Using Instrumental Variables*. Journal of the American Statistical Association.. Vo. 91. No. 434,
- Baker H., (1999) *The MPR Upward Bound Report: An Equity Researcher's Response*, Norcal WESTOP, Professional Development Seminar, October 1999.
<http://westop.csuchico.edu/resources/matheval.html#anchor2716138>
- Bloom, Howard (ed.), (2005) *Learning More from Social Experiments*, New York: Russell Sage Foundation.
- Brick Michael, Cahalan, Margaret, Gray Lucinda, Severynse Jacqueline. (1995) *A Study of Selected Nonsampling Errors in the 1991 Survey of Recent College Graduates*, National Center for Education Statistics, U.S. Department of Education. NCES 95-640.
- Burkheimer G., Riccobono J. Wisenbaker G., (1979) *Final Report: Evaluation Study of the Upward Bound Program—A Second Follow-Up*, (Research Triangle Park, NC: Research Triangle Institute.
- Cahalan Margaret W and Curtin Thomas R, (2004) *A Profile of the Upward Bound Program 2000-2001*, Washington D. C. US Department of Education, Office of Postsecondary Education, Washington, D.C.
- Curtin Thomas R and Cahalan Margaret W (2004) *A Profile of the Upward Bound Math-Science Program 2000-2001*, Washington D. C. US Department of Education, Office of Postsecondary Education, Washington, D.C.
- Campbell Donald and Stanley Julian, (1969) *Experimental and Quasi-Experimental Designs for Research*, 1963, Reprinted from Handbook of Research on Teaching, American Educational Research Association.
- Alisa Cunningham, Sarah Krichels, Jamie Merisotis. (2004) *Persistence and Attainment Measures for Federal Student Aid Programs*. Postsecondary Education Statistical Analyses and Technical Development. Contract Number: 99-CO-0109. Task Order 38. October 5, 2004. Institute for Higher Education Policy.
- Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: Wiley
- Groves, Fowler, Couper, Lepkowski, Singer, and Tourangeau (2004). *Survey Methodology*. University of Chicago Press.
- Heckman, J. Hohmann, N., Smith J., Khoo M., (2000) Substitution and Dropout Bias in Social Experiments: A study of an Influential Social Experiment, *The Quarterly Journal of Economics*, May 2000.
- Horn Laura J., Xianglei Chen, and MPR Associates. (1998) "Toward Resiliency: At-Risk Students Who Make It to College." U.S. Department of Education, Office of Educational Research and Improvement. Washington, DC: U.S. Government Printing Office, May 1998.
- Ingels, S.J., Curtin, T.R., Kaufman, P., Alt, M.N., and Chen, X. (2002). *Coming of Age in the 1990s: The Eighth-Grade Class of 1988 12 Years Later*. (NCES 2002-321). Washington, DC: U.S. Department of Education, National Center for Education Statistics
- Jung S. and Applied Systems Institute, (1984) *Reanalysis of High School and Beyond Data to Estimate the Impact of Upward Bound*. Washington, DC: Applied Systems Institute.

-
- McLaughlin D, Drori G. and Ross M. (2000) U.S. Department of Education. National Center for Education Statistics. *School-level Correlates of Academic Achievement: Student Assessment Scores in SASS Public Schools*, NCES 2000-303
- The Joint Committee on Standards for Educational Evaluation. (1994) *The Program Evaluation Standards* The Joint Committee is accredited by the American National Standards Institute ([ANSI](#)). Standards approved by ANSI become American National Standards. Created in 1975, the Joint Committee is a coalition of major professional associations concerned with the quality of evaluation. The Joint Committee is housed at The Evaluation Center, Western Michigan University
- Moore, M. *A 1990's View of Upward Bound: Programs Offered, Students Served, and Operational Issues*. Washington, DC: U.S. Department of Education, Planning and Evaluation Service, 1996.
- Myers D., (1991) "The Effects of Upward Bound and Supplemental Service Programs: Findings from Extant Data" Rockville, MD: Westat, Inc.
- Myers, D., Schirm A., (1996) *The Short-Term Impacts of Upward Bound: An Interim Report*. Washington, DC: U.S. Department of Education, Planning and Evaluation Service.
- Myers, David, Robert Olsen, Neil Seftor, Julie Young, and Christina Tuttle. (2004) "The Impacts of Regular Upward Bound: Results from the Third Follow-Up Data Collection." Report submitted to the U.S. Department of Education. Washington, DC: Mathematica Policy Research, Inc..
- Myers, D and Schirm A. (1999) "The Impacts of Upward Bound: Final Report on Phase I of the National Evaluation." Report submitted to the U.S. Department of Education. Washington, DC: Mathematica Policy Research, Inc.. (<http://www.ed.gov/offices/OUS/PES/higher/upward.pdf>).
- Olsen, Robert, Neil Seftor, Tim Silva, David Myers, David DesRoches, and Julie Young. (2007) "Upward Bound Math Science: Program Description and Interim Impact Estimates." U.S. Department of Education. Washington, DC: Mathematica Policy Research, Inc.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and Quasi-Experimental Designs For Generalized Causal Inference*. Houghton Mifflin.
- Seastrom, M. (2002). *NCES Statistical Standards* (NCES 2003-601). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Standing K., Judkins D., Keller B., Shimshak A., *Early Effects of the GEAR UP Program*, draft report submitted to US Department of Education, Policy and Program Studies Service. 2008.
- Elizabeth Stuart, Robert Olsen, Allen Schirm, Julie Young, David DesRoches (unpublished internal report 2006) *The Impacts of Regular Upward Bound on Postsecondary Activities Within 4-6 Years After Scheduled High School Graduation*, Unpublished final report, Mathematica Policy Research.
- U.S. Department of Commerce, Bureau of the Census, October Current Population Surveys, 1972-2000 (<http://www.census.gov/population/www/socdemo/school.html>).
- Weisberg, Herbert F. (2005). *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. University of Chicago Press.

Table A. Intent To Treat (ITT) logistic regression results models for dependent variable of having evidence of postsecondary from any applicable survey or from SFA files by +1 (18 months) of expected high school graduation year (EHSGY): National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04 (Uses alternative grade variable for standardization of EHSGY)

pr-T = 72.7 pr-C = 66.0 Difference = 6.7**** npse18	Variable name	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]
FFUTC (random assigned to treatment)	ffutc	0.371381	0.092769	4	0	0.183739 0.559023
Ffgr9 (grade 10 ref)	ffgr9	0.00818	0.329167	0.02	0.98	-0.65762 0.673983
Ffgr11 (grade 10 ref)	ffgr11	-0.16084	0.164541	-0.98	0.334	-0.49366 0.17197
Ffgr12 (grade 10 ref)	ffgr12	-1.62816	0.395105	-4.12	0	-2.42734 -0.82898
Clowoy (Low income only)	clowoy	0.274296	0.247483	1.11	0.275	-0.22629 0.774878
Cfgenoy (First generation only)	cfgenoy	0.375038	0.193827	1.93	0.06	-0.01701 0.76709
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	0.783521	0.301535	2.6	0.013	0.17361 1.393432
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	0.427427	0.226795	1.88	0.067	-0.03131 0.886164
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-0.31887	0.207559	-1.54	0.133	-0.7387 0.100957
Cexdk (Baseline educational expectation was don't know---ref BA)	cexdk	-0.79806	0.145596	-5.48	0	-1.09256 -0.50357
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	-1.28867	0.286644	-4.5	0	-1.86846 -0.70888
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-0.96014	0.163404	-5.88	0	-1.29065 -0.62962
cex14aa (Baseline educational expectation was 2-year---ref BA)	cex14aa	-0.62401	0.119286	-5.23	0	-0.86529 -0.38273
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	0.035613	0.11215	0.32	0.753	-0.19123 0.262458
Cexphd (Baseline educational expectation was PhD---ref BA)	cexphd	0.223459	0.101548	2.2	0.034	0.01806 0.428859
Cothrac (Race was not Hispanic or Black or White—ref Black)	cothrac	-0.0687	0.285566	-0.24	0.811	-0.64631 0.50891
Chisp (Hispanic—ref Black)	chisp	-0.26647	0.196386	-1.36	0.183	-0.6637 0.130757
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.5774	0.210339	-2.75	0.009	-1.00285 -0.15195
Cfemale (Female)	cfemale	0.578733	0.077784	7.44	0	0.421399 0.736067
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.400858	0.131542	3.05	0.004	0.134789 0.666926
_cons	_cons					

NOTE: Model uses an alternative variable from the First Follow-up (A3) instead of variable B1 from the Baseline Survey on which to standardize EHSGY (See table B-1). SFA = Student Financial Aid files. Ref = left out reference in dummy variable sequence. See table 5 in body of text for additional note information. Number of strata (wprstco)= 28; Number of PSU (wprojid) = 67; uses postratified longitudinal baseline weight (v5bwgtp1).

SOURCE: Data tabulated June 2008 using; National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.

Table B. Instrumental variables regression for Treated On Treated (TOT) modeling dependent variable of appearing on the federal SFA files by +1 (18 months) of expected high school graduation year (EHSGY): National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04 (estimate reported in table 8)

Variable name	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	xb T = 62.9 xb C = 53.6 Difference = 9.3**** (Kaidhs)	
xnewgp (Participated in UB/UBMS)	0.123149	0.037028	3.33	0.002	0.048253	0.198045	
Gr79293 (Grade 7 in 1992-93 ref grade 9)	-0.01293	0.132541	-0.1	0.923	-0.28102	0.255162	
Gr89293 (Grade 8 in 1992-93 ref grade 9)	-0.12538	0.048361	-2.59	0.013	-0.2232	-0.02756	
Gr109293 (Grade 10 in 1992-93 ref grade 9)	0.008279	0.125262	0.07	0.948	-0.24509	0.261646	
Gr119293 (Grade 11 in 1992-93 ref grade 9)	-0.24429	0.168436	-1.45	0.155	-0.58498	0.096407	
Clowoy (Low income only)	0.017819	0.049742	0.36	0.722	-0.08279	0.118432	
Cfgenoy (First generation only)	0.051589	0.051158	1.01	0.319	-0.05189	0.155066	
C11gssf (Grade was 11 on student selection form—ref grade 9)	-0.02485	0.141647	-0.18	0.862	-0.31135	0.261663	
C10gssf (Grade was 10 on student selection form—ref grade 9)	-0.04074	0.055938	-0.73	0.471	-0.15389	0.072402	
C8gssfm (Grade was 8 on student selection form—ref grade 9)	-0.14285	0.118341	-1.21	0.235	-0.38222	0.096515	
Cexdk (Baseline educational expectation was don't know---ref BA)	-0.18389	0.040305	-4.56	0	-0.26541	-0.10236	
Cexhs (Baseline educational expectation was high school only---ref BA)	-0.26295	0.0628	-4.19	0	-0.38997	-0.13592	
Cex13v (Baseline educational expectation was vocational---ref BA)	-0.18834	0.035934	-5.24	0	-0.26103	-0.11566	
cex14aa (Baseline educational expectation was 2-year---ref BA)	-0.17929	0.024134	-7.43	0	-0.22811	-0.13047	
Cexma (Baseline educational expectation was Masters Degree---refer BA)	0.022336	0.025866	0.86	0.393	-0.02998	0.074655	
Cexphd (Baseline educational expectation was PhD---ref BA)	0.042772	0.018322	2.33	0.025	0.005713	0.079831	
Cothrac (Race was not Hispanic or Black or White---ref Black)	0.018877	0.046879	0.4	0.689	-0.07595	0.113699	
Chisp (Hispanic---ref Black)	-0.06829	0.06105	-1.12	0.27	-0.19177	0.055199	
Cwhite (Race was White, not Hispanic---ref Black)	-0.10059	0.038615	-2.6	0.013	-0.17869	-0.02248	
Cfemale (Female)	0.139323	0.02847	4.89	0	0.081736	0.196909	
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	0.032356	0.0231	1.4	0.169	-0.01437	0.07908	
_cons	0.586899	0.075209	7.8	0	0.434774	0.739023	

NOTE: Results of this table appear in table 8. Standardized based on Baseline Survey question B1 with correction for 1991-92 responders. SFA = Student Financial Aid files. Ref = left out reference in dummy variable sequence. See table 5 for additional note information. Number of strata (wprstco) = 28; Number of PSU (wprojid) = 67; uses poststratified longitudinal baseline weight (v5bwgtp1). Instrumented: xnewgp; Instruments: gr79293 gr89293 gr109293 gr119293 clowoy cfgenoy c11gssf c10gssf c8gssfm cexdk cex13v cexhs cex14aa cexma cexphd cothrac chisp cwhite cfemale parbefor ffutec.

SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Planning Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education.; study conducted 1992-93-2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.