

Attachment D
Flawed Contractor Reports from the National Evaluation of Upward Bound
Masked Significant and Substantial Positive Impacts (Draft)

By Margaret Cahalan and David Goodwin

Executive Summary

In January 2009, in the last week of a departing Administration, the U.S. Department of Education (ED) published the final report in a long running National Evaluation of Upward Bound (UB) conducted by the contractor, Mathematica Policy Research. The report was published over objections from the Policy and Program Studies Services (PPSS) ED career technical staff who were assigned to monitor the final Mathematica contract. The report was also published after a “disapproval to publish” rating in the formal review process from the Office of Postsecondary Education (OPE), out of whose program allocation the evaluation was funded. The Mathematica reports from the UB study (Myers et. al. 2004; and Seftor et. al. 2009) have had a large impact on policy development for more than a decade, resulting in an OMB “ineffective rating” and zero budget funding justifications in President Bush’s budget in FY2005 and FY 2006 for all of the federal pre-college programs, UB, Upward Bound Math Science (UBMS), Talent Search and GEAR UP.

Reason for Speaking Out At this Time. As the original (Dr. Goodwin) and final (Dr. Cahalan) Contracting Officers Technical Representatives (COTRs) for the study within the US Department of Education, our official job was to provide Technical Monitoring of the contracts. In the final of three sequential contracts, after concerns about the study were raised, we conducted a Quality Assurance Review (ED-PPSS QA review), and found that the impact estimations from the study being reported by the contractor were seriously flawed so much so that the basic conclusions Mathematica made concerning the efficacy of the Upward Bound program were impacted. While we have spoken out before on this topic, we are speaking out again in 2014, because of the on-going and recent citations of the erroneous findings from the study in Congressional testimony, policy briefs, and public speeches (Whitehurst, 2011, Haskins and Rouse 2013; Decker 2013). These erroneous findings continue to do serious reputational harm to the Upward Bound program.

ED-PPSS QA Review. The ED-PPSS QA review involved an internal review and analysis of all data files from the study, as well as consultation and replication of results by external statistical experts. The data files reviewed included: the initial sampling frame, the baseline survey, five follow-up surveys, student transcripts, 10 years of federal aid files and 10 years of National Student Clearinghouse (NSC) data. The ED-PPSS QA found that the Mathematica reports were seriously flawed, made unwarranted conclusions about the Upward Bound program and were not transparent in reporting. Moreover statistically significant and educationally meaningful positive impacts on the key legislative goals of the Upward Bound program were clearly found when the study errors were addressed using standards based statistical methods. These positive impacts are unacknowledged in the Mathematica reports. Below are highlights from the PPSS QA review and re-analysis.

Major Flaws Identified in the Reports. Major statistical and evaluation research standards violations were found including: 1) A flawed sample design with severe unequal weighting in which the highest weighted students had weights 40 times those of the lowest weighted students and one single project of 67 carried fully 26 percent of the weight; 2) Serious representational errors with one single atypical former 2-year college with an historical focus on certificates selected to represent the largest 4-year and above degree granting stratum; 3) Severe non-equivalency of the treatment and control group on academic risk,

grade at entrance, and educational expectations leading to uncontrolled bias in favor of the control group in all of the impact estimates upon which conclusions were made; 4) Failure to use a common standardized outcome measures for a sample that spanned 5 years of expected high school graduation year; 5) Improper use of National Student Clearinghouse (NSC) data to impute survey non-responders' enrollment and degree attainment status when coverage was far too low and non-existent for 2-year and below degrees, with bias clearly evident; 6) False attribution of large negative impacts in the project with extreme weights to "poor performance" ignoring the extreme bias in favor of the control-group in this project's sample ; 7) Lack of addressing issues of control group receipt of alternative but less intensive federal pre-college services received by the majority (60 percent) of the control group members; and 8) Lack of reporting transparency and failure to acknowledge strong positive impacts of UB on key program goals that are found when these errors are addressed using standards based statistical and evaluation research methods.

PPSS Re-Analysis Found Strong Positive Impacts. Contrary to the Mathematica conclusions that the only overall impact was on certificate attainment, the ED-PPSS QA re-analysis conducted by ED internal monitoring staff found that when NCES and What Works Clearinghouse (WWC) standards were followed to mitigate or correct the errors noted above, there were statistically significant and substantively meaningful positive results for the Upward Bound program. These impacts were on the major legislatively-mandated goals of the program---postsecondary entrance, application for and award of financial aid, and degree attainment (see Figures 5 to 8). The impacts included a 50 percent Treatment on the Treated (TOT) increase in BA degree attainment within six years of expected high school graduation using the balanced treatment and control group (Figure 6). Instrumental variables regression controlling for selection factors revealed that 75 percent of UB/UBMS participants entered postsecondary within one year of high school graduation compared to 62 percent of those who received only a less intensive service such as Talent Search, and 45 percent of those who reported no pre-college service receipt (figure 7). PPSS also found that UB/UBMS participants were 3.3 times more likely to obtain a BA in six years when compared to those reporting no participation in college access supplemental services and 1.4 times as likely when compared to those who reported participating in less intensive supplemental services (Figure 8). For the full re-analysis report detailing issues and full documentation of the re-analysis results, see [http://www.pellinstitute.org/publications-Do the Conclusions Change 2009.shtml](http://www.pellinstitute.org/publications-Do%20the%20Conclusions%20Change%202009.shtml)

Support for "COE 2012 Request for Correction" Submitted to ED in 2012 and for the "2014 Request to Rescind" the WWC UB Study Rating The article concludes that the non-transparent published reports from the National Evaluation of Upward Bound suffer from what is known as a Type II study error, or a failure to detect positive impacts when they are present. Thus the Mathematica conclusions that UB had no impact on postsecondary entrance, financial aid or degree attainment outcomes except for a positive impact on the award of certificates are incorrect. The article expresses support for the Council for Opportunity in Education (COE) formal *Request for Correction* submitted to the Department of Education in 2012 calling for the Mathematica reports to be corrected or withdrawn. The article also supports the 2014 request that the What Works Clearinghouse (WWC) "rescind" the 2009 rating given to the UB study reports of "meets evidence standards without reservations." The 2012 request was accompanied by a Statement of Concern signed by leading researchers in the field, including the sitting presidents of the American Education Research Association (AERA) and the American Evaluation Association (AEA). The complete text of the *Request for Correction* is available at [http://www.coenet.us/files/pubs_reports-COE Request for Correction 011712.pdf](http://www.coenet.us/files/pubs_reports-COE%20Request%20for%20Correction%20011712.pdf), and the *Statement of Concern* signed by leading researchers can be found at [http://www.coenet.us/files/ED-Statement of Concern 011712.pdf](http://www.coenet.us/files/ED-Statement%20of%20Concern%20011712.pdf).

Introduction

In January 2009, in the last week of a departing Administration, the U.S. Department of Education (ED) published the fourth and final report in a long running National Evaluation of Upward Bound (UB) (Myers and Schirm 1996; 1999; Myers et. al. 2004; and Seftor et. al. 2009). The 2009 report was published over the objections of the ED career technical staff assigned to monitor the final contract, and after a “disapproval to publish” rating in the formal review process from the Office of Postsecondary Education (OPE), out of whose program allocation the evaluation was funded.

Upward Bound (UB) is a Federal program, begun in 1965, designed to provide college readiness through supplemental academic services, as well as college awareness, leadership, and counseling services. Congressionally-mandated eligibility requirements specify that two-thirds of the high school participants must be low-income (defined as 150 percent of the poverty level) and students who would potentially be the first person in their family to obtain a bachelor’s (BA) degree (known as “first-generation college” students). The other one-third must be either low-income or first-generation. Upward Bound is one of the first and considered a model flagship Federal program. It is also one of the more intensive low-income and first-generation college access programs with an average cost per student of about \$4,300. There are about 900 Upward Bound (UB) and Upward Bound Math Science (UBMS) programs across the country. Project grantees responsible for implementing UB are 4-year and 2-year postsecondary institution and community organization grantees who together serve about 65,000 high school students yearly. The program has a strong academic focus with an intensive six-week summer traditionally residential program that is held on a college campus followed by weekly academic year sessions throughout high school. As specified in the authorizing legislation, all Upward Bound projects must provide instruction in mathematics through pre-calculus, laboratory science, foreign language, composition and literature through summer programs on a college campus and academic year supplemental services. The goal of Upward Bound is to increase the rate at which low-income and potentially first-generation college participants complete secondary education and enroll in and graduate from institutions of postsecondary education. UB and UBMS grantees hold competitive five-year grants to administer UB services to low-income and first-generation students in high-needs target high schools in their local communities.

Dr. Cahalan is Vice President for Research and Director of the Pell Institute for the Study of Opportunity in Higher Education of the Council on Opportunity in Education (COE). Dr. Cahalan supervised the staff serving as the UB evaluation’s technical monitors and served in this capacity herself in the final few months of the UB evaluation. She is currently the Co-PI of the COE i-3 project “Using Data to Inform College Access Programming.”

Dr. Goodwin is currently an independent consultant for the Gates Foundation. He is the former Director of the unit within the U.S. Department of Education responsible for the UB Evaluation. Dr. Goodwin was Dr. Cahalan’s supervisor at the time of the final Mathematica UB Contract. He was the UB study monitor when the study was first begun in 1992.

The random assignment longitudinal study followed low-income and “potentially first-generation-college” students from middle school or early high school through six to 10 years after their expected high school graduation year (EHSKY). The study was conducted under a

series of three contracts with a baseline and five follow-up student surveys by Mathematica Policy Research (Mathematica).

The results of this seemingly high-quality random assignment study have formed the basis for significant policy justifications—most notably a Bush administration budget request to eliminate funding for Upward Bound and other federal pre-college access programs—Talent Search and GEAR UP, and a decision by the Office of Management and Budget (OMB) to rate the program as “ineffective.” In November 2011, the study report findings were reflected in the testimony to Congress of former Institute for Education Sciences (IES) Director Grover T Whitehurst, asserting that federal programs such as Upward Bound and Head Start had not been shown to be effective. More recently, in May 2013, it has formed the justification for the assertion by a Brookings Policy Brief (Haskins and Rouse, 2013) that in general federal college access programs “show no major effects on college enrollment or completion.” These well-known authors state that their conclusions are based primarily on the Mathematica Upward Bound study. They identify the Mathematica UB study as being the only evaluation of federal college access programs to be given the highest study methods rating by the What Works Clearinghouse (WWC), a clearinghouse, co-incidentally also run under an ED contract to Mathematica.

Ironically, as Technical Monitors for the evaluation while working at ED-PPSS, we found in a Quality Assurance (QA) review of study design and data files that the widely-cited reports from this evaluation were not transparent and made unwarranted conclusions concerning the Upward Bound program. We concluded that the Mathematica reports were seriously flawed in terms of statistical sampling standards violations and importantly had a serious uncontrolled statistical bias in favor of the control group on academic risk factors. These identified biases violate a basic National Center for Education Statistics (NCES) and WWC Standards that the sample be representative of the population of interest and that the treatment and control group be balanced and equivalent on baseline factors related to outcomes. Importantly, we also found, when we conducted a re-analysis based on NCES and WWC standards and the recommendations of independent external statistical reviewers, that there were statistically significant and substantively strong positive results for the Upward Bound program. These impacts were on the major legislatively-mandated goals of the program—postsecondary entrance, application for and award of financial aid, and attainment of bachelors’ (BA) degrees and other postsecondary degrees or credentials. We concluded that the non-transparent published reports from the National Evaluation of Upward Bound suffer from what is known as a Type II study error, or a failure to detect positive impacts when they are present.

We made our concerns and the QA re-analysis positive results well known to Mathematica and the Department of Education at the time (Cahalan 2009). As the ED Technical Monitors for the study, we reiterate our serious concerns publicly now in the light of repeated use of the flawed Mathematica results in Congressional testimony, policy briefs, and public speeches (Whitehurst, 2011, Haskins and Rouse 2013; Decker 2013). We also do so in order to support the formal *COE 2012 Request for Correction* of the Mathematica final report, submitted to ED almost two years ago, by COE and their affiliated regional Educational Opportunity Organizations. These organizations represent TRIO program stakeholders in the evaluation. The COE request for correction was accompanied by a *Statement of Concern* signed by, among others, the Presidents of the American Evaluation Association (AEA) and the American Education Research

Association (AERA). Each of the signers of the Statement of Concern had reviewed the *COE Request for Correction* prior to signing the *Statement of Concern*. We are also writing this report in order to support a formal *Request to Rescind* the rating given by the What Works Clearinghouse (WWC) of “Meets evidence standards without reservations” given to Mathematica Upward Bound reports in the 2009, WWC Practice Guide entitled: [Helping Students Navigate the Path to College: What High Schools Can Do](#).

Before discussing our QA findings in more detail, we wish to make clear that this article is not intended to be a general critique of the random assignment method nor a post-hoc effort to “fish” for positive study findings. Nor is the article intended to discredit the study as a whole. While we object strongly to the failure of Mathematica to address the flaws in their impact estimates or to acknowledge the positive results obtained when these issues are addressed using standards based methods, we also believe that the *National Evaluation of Upward Bound*, when corrected for sampling and non-sampling error, can be a very useful and informative study in the area of pre-college research. The essence of our findings is detailed below.

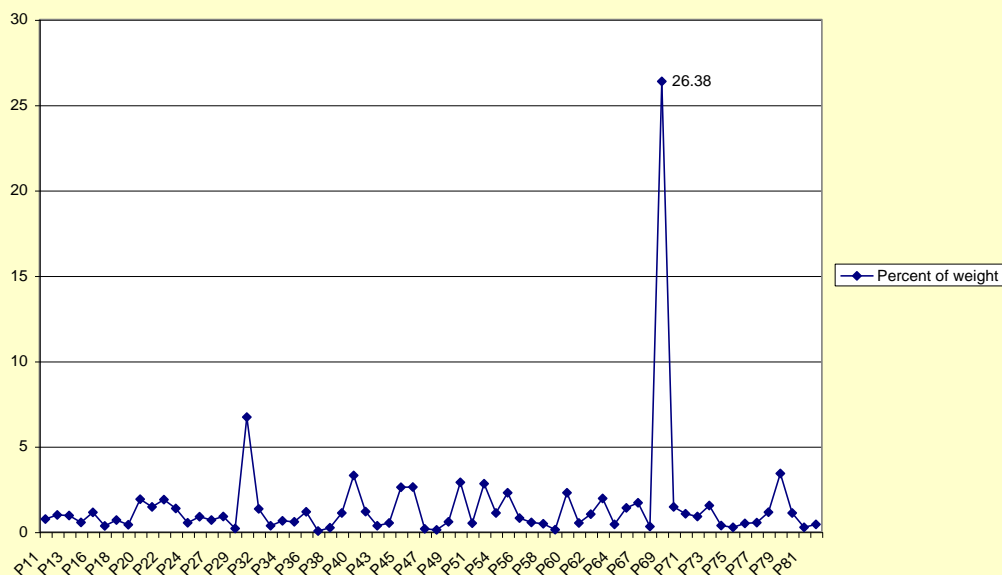
Major Findings from the Technical Monitors Review of Study Issues

Seriously Flawed Sample Design and Severe Unequal Weighting. The design for this study was unusual and overly ambitious and unfortunately resulted in a multi-stage sample with one project carrying 26.4 percent of the final student weights. In what reviewers have called a “seriously flawed sample design” that does not meet NCES standards, only one project in the sample (called project 69) was selected to represent the largest study defined 4-year and above public grantee stratum. Furthermore, because of an unusually large number of “baseline” surveys from interested students submitted by project 69, in the final stage of weighting, project 69 carried fully 26 percent of the weights. Figure 1 shows just how extreme the unequal weighting was from project 69. The method of counting baseline surveys submitted by the sampled projects as “applicants” and constituting a so called “waiting list” and then weighting to the number of baseline surveys (considered applicants) within project defined sub-strata further confounded the already-flawed first stage sample design. In addition, projects used different recruitment methods to obtain the “waiting list” based on returned baseline surveys and were allowed to create project specific sub-strata from which students were randomly selected at differential rates. Subsequently there were large differences among the sampled projects in the ratio of baseline surveys submitted to the number of project openings over the period. The weights were the inverse of the probability of selection at each of the stages (project and student applicant level). Because project 69 was supposedly representing a very large number of both projects and applicants, this flawed design meant that the outcomes of some students from the project 69 “waiting list” carried a weights that were 40 times those of the lowest weighted students (for example, some project 69 sample members had weights of 158 while the lowest

In what reviewers have called a “seriously flawed sample design” that does not meet NCES standards, only one project in the sample (called project 69) was selected to represent the largest study defined 4-year public stratum and carried fully 26.4 percent of the weight.

weighted sample member among all the projects carried a weight of 4). Mathematica reports that were published over almost a 10 year period did not reveal these serious sample design issues.

Figure 1. Percentage distribution of sum of the weights of the 67 projects making up the study sample: National Evaluation of Upward Bound, study conducted 1992-93 to 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), U.S. Department of Education,; study conducted 1992-93 to 2003-04.

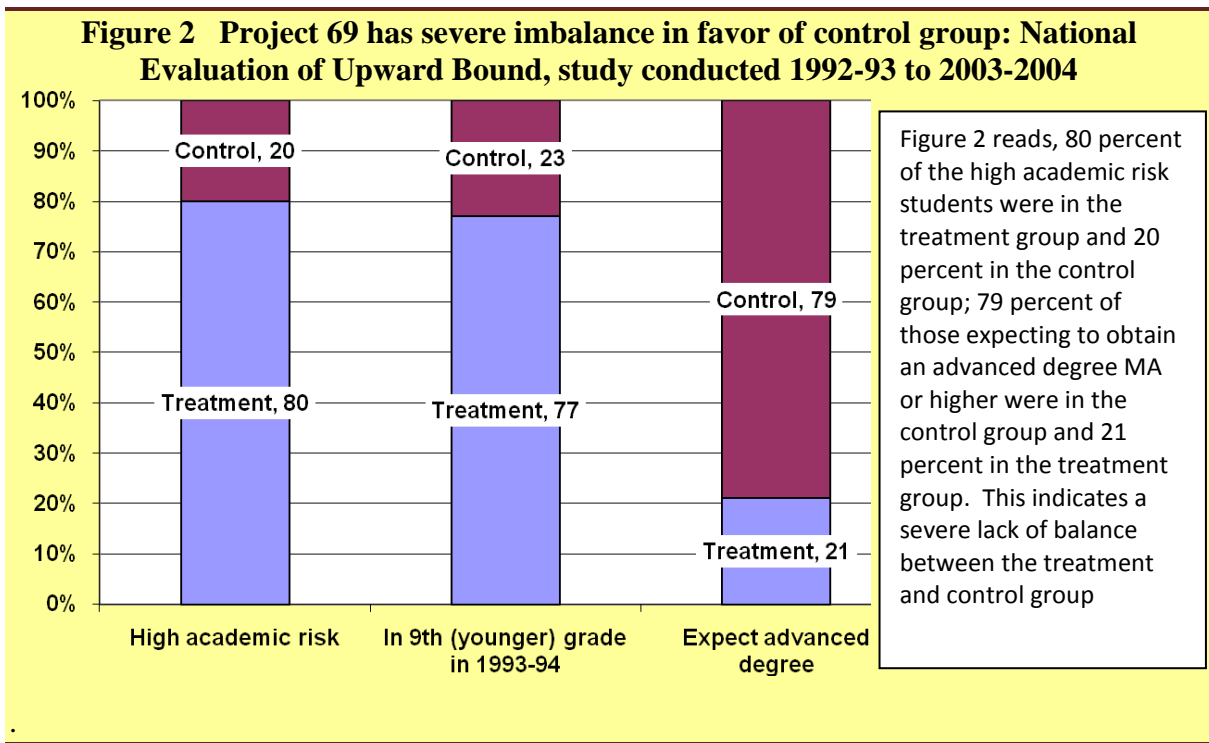
Atypical Project Selected As Sole Representative of Largest Stratum. Unfortunately, project 69, whose students carried 26 percent of the weight, was also found to be atypical. Randomly chosen as the sole representative of the largest study defined 4-year and above grantee stratum, the project 69 grantee institution had historically been a junior college, offering associate and certificate programs taken over to serve as a branch of a nearby 4-year city-wide college system. Project 69’s UB program was non-residential and partnered with a job training program serving Career and Technical Education (CTE) target minority high schools. It thus had a higher-than-average, especially for a 4-year grantee, percentage of its UB participants who were interested in seeking less than 2-year vocational certificates.

The ED staff QA review found that project 69 was “atypical” of the 4-year stratum for which it was the sole representative.

The study reports do not reveal project 69’s representational issues, and indeed Mathematica’s final report specifically asserts that project 69 is an adequate sole representative of the types of projects likely to be present within this, the largest 4-year and above study stratum (Sheftor, et. al. 2009). The stratum project 69 was supposedly representing and that justified its

26 percent weight was a large combined stratum of average sized projects housed at 4-year colleges and universities. It included the major flagship research universities as well as small 4-year liberal arts colleges that had UB grants at the time. Neither of these types of 4-year and above grantees could be adequately represented by project 69.

Serious Lack of Balance between the Treatment and Control Group. A basic standard of the What Works Clearinghouse and random assignment studies generally is that in order to make valid impact estimates, the treatment and control group must be equivalent at baseline on factors related to outcomes. Although the random assignment method is intended to ensure that treatment and control groups are equivalent (and did so quite well for the combined UB sample without project 69), in project 69, the QA review found major differences between the treatment and control groups on factors related to outcomes. The imbalance in project 69 was so large that some external reviewers reported they suspected a failure to implement the random assignment correctly in this project. For example as shown in Figure 2 below, 80 percent of the academically at-risk students from the project 69 sample were in the treatment group (randomly assigned to Upward Bound in middle or early high school), while 20 percent of the academically at-risk students were in the control group (not randomly assigned to UB in middle or early high school).



For project 69, the treatment sample on average resembled the vocational programming emphasis of the project, with a larger than average for a 4-year grantee of participants interested in certificate programs; while the control group on average resembled the typical Upward Bound Math Science (UBMS) applicant with a larger percentage on average interested in obtaining advanced degrees (56 percent). After the identity of project 69 became known to ED at the end of the final contract, in researching the project 69 issue, we found that there was a neighboring newly formed UBMS project operating in the region. As seen in Figure 2, the control group members were in a higher grade, were more academically proficient, and had considerably higher educational expectations at baseline. This suggests that the unusually large number of baseline surveys (n=85) collected by project 69 relative to their actual openings may have been because they included those students who were actually applying for the neighboring UBMS program from a high school science and technology magnet program also located at one of the project 69 target schools along with the Vocational Career and Technical Education program. As Technical Monitors, we discovered these issues only gradually when we did direct QA analysis of the data files to discover why project 69's Upward Bound program had demonstrated such seemingly negative impacts on postsecondary outcomes relative to its control group.

The UB study analyses violate the basic What Works Clearinghouse standard that the treatment and control group be equivalent on baseline factors related to outcomes.

Unfortunately, the severe non-equivalency in project 69 combined with the extremely large weights for the students from this project resulted in an imbalance in the overall sample and an uncontrolled bias in favor of the control group in all of the Mathematica impact estimates (Mathematica had no controls for academic risk factors in their analysis). For example, in the overall sample with project 69 included, 58 percent of the academically at-risk students were in the treatment group and 42 percent in the control group (Figure 3). In contrast, when we did balance checks on the combined sample without project 69, we observed a good balance between the treatment and control group on these same factors, with for example, 51 percent of the academically at-risk students in the treatment group and 49 percent in the control group (Figure 4).

**Figure 3. Imbalance in Overall Upward Bound Sample with Project 69 included:
National Evaluation of Upward Bound, study conducted 1992-93 to 2003-2004**

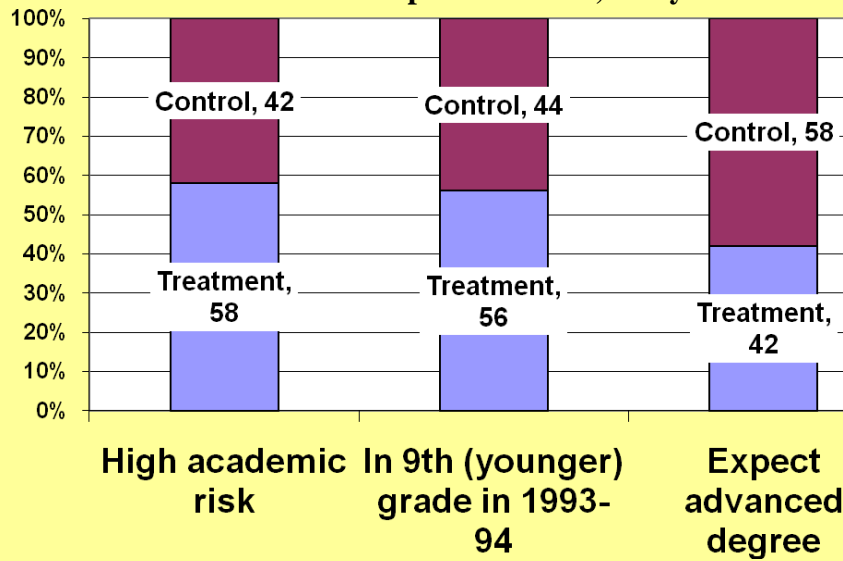


Figure 3 reads, for example: In the overall sample, among the high academic risk students, 58 percent were in the treatment group and 42 percent in the control group

Figure 4. More Balanced Treatment and Control Group for 66 other projects taken together: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-2004

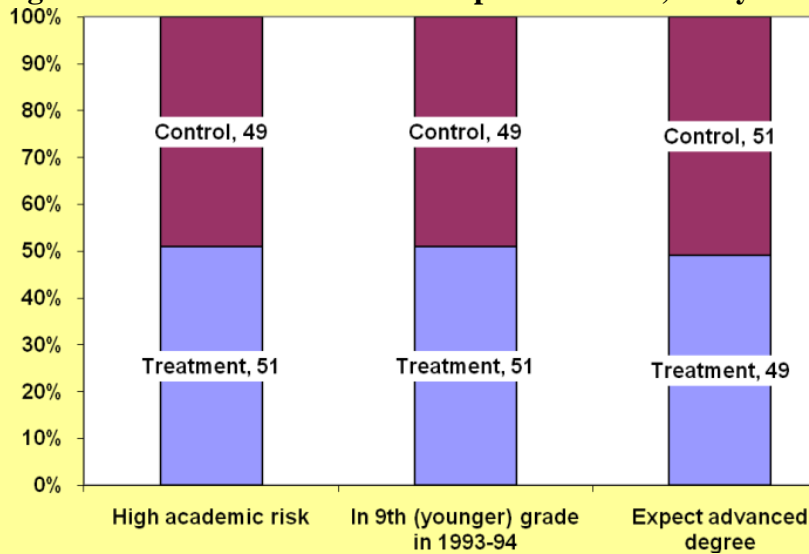


Figure 4 shows the balance between the treatment and control group on key factors when project 69 is excluded

Lack of Standardization of Outcome Measures to Expected High School Graduation for a Sample that Spanned Five years of Expected High School Graduation Year. The issues noted above were aggravated by the fact that Mathematica, in violation of the NCES and What Works Clearinghouse standards, did not standardize the outcome measures for a sample that spanned five years of expected high school graduation years. Mathematica argued that randomization made this unnecessary. However, balance checks done by ED monitoring staff found that on average, the control group was in a higher grade in a fixed academic year than the treatment group (see Figure 3). In addition, to the obvious issues related to differences in levels of potential opportunity to enter postsecondary and complete degrees over five years of expected high school graduation years, this lack of standardization also confounded the ability of the other variables in the regression models to function in a meaningful way to control for baseline differences.

The Mathematica reports, use unstandardized outcome measures for a sample that spanned 5 years of expected high school graduation dates violating NCES and What Works Clearinghouse standards requiring use of common standardized outcome measures.

Improper Use of National Student Clearinghouse (NSC) Data. In violation of NCES standards, the final report of the Mathematica study also makes improper use of NSC data for imputation outcome measures for survey non-responders in a very early period when enrollment coverage was too low and when degree coverage for 2-year and less than 2-year degrees was nonexistent. This improper use of NSC introduced bias into the conclusions Mathematica reported for the study. For example, Mathematica ignored their own significant and substantial positive impact results based on fifth follow-up survey data adjusted for non-response for the award of “any postsecondary degree or credential” (Seftor et.al. 2009, see appendix C), and falsely reported that the study only detected postsecondary credential impacts only for award of certificates.

Mathematica’s own estimate of attainment of “any postsecondary degree or credential” based on responders to the fifth-follow-up survey adjusted for non-response shows a positive substantial and significant Intent To Treat (ITT) impact of UB on award of “Any postsecondary degree or credential” of 13 percentage points (55 percent for UB and 42 percent for the control group) and a Treatment On the Treated (TOT) estimate of a 16 percentage point difference—(Seftor et. al. 2009 Appendix tables C-7 and C14). Against ED Technical Monitors’ recommendation and that of the IES external reviewers to be conservative in use of NSC, Mathematica chose to present in the text tables in the body of the report and base their conclusions only those estimates that used NSC data for non-responders to the fifth follow-up—coding the 25 percent of the sample who were survey non-responders and who were not found in NSC as “not having any degree or certificate.” The significant and large positive results noted above, tabulated by Mathematica itself, are included in Mathematica’s appendix tables C-7 and C-14 but not mentioned in the text body presentation of study conclusions.

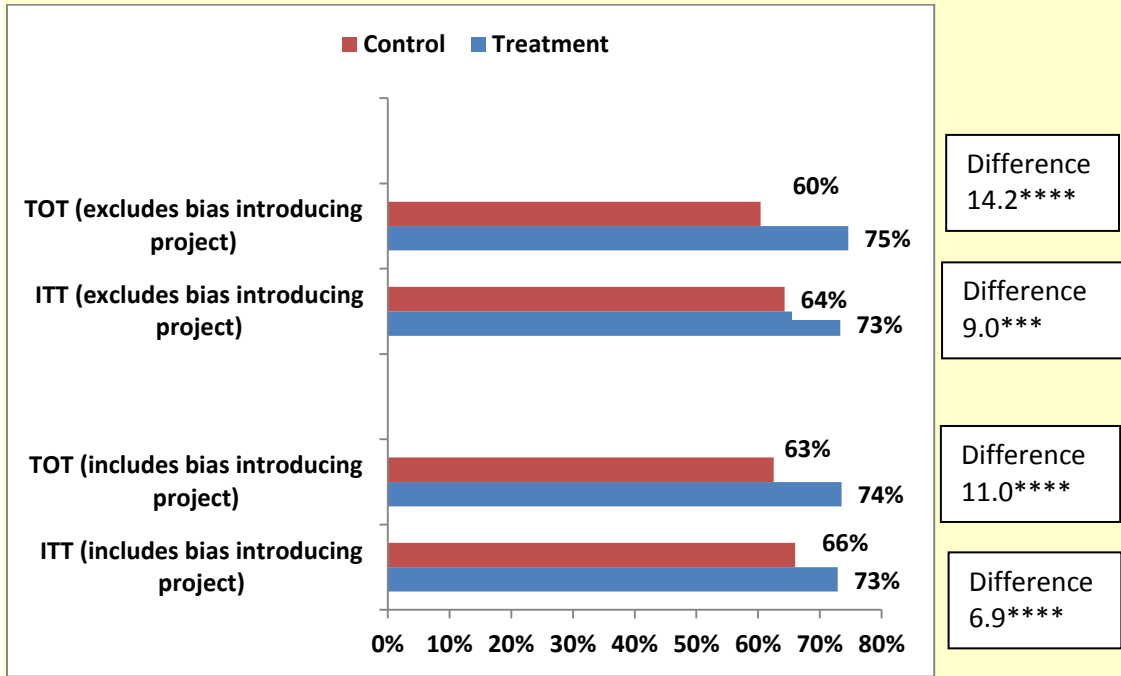
Alternative Re-Analysis Major Findings

As the issues within the Mathematica UB reports became known to ED staff, we began to consult outside experts and to use NCES and WWC Standards as guides to mitigate the issues. We prepared impact estimates that we considered more robust containing less statistical bias. In conducting the re-analysis, we standardized outcome measures to expected high school graduation year. To maximize response, the re-analyses also included information from each of the three applicable follow up surveys (third through fifth), and used 10 years of federal aid and award files to supplement the survey data. However, following NCES standards, we avoided use of the NSC for enrollment and degrees less than the BA due to lack of coverage in this early period in the NSC history. Following expert advice, we prepared and reported all impact estimates with and without project 69 and included impact estimates for the sample, weighted and unweighted. For the full re-analysis report detailing issues and full documentation of the re-analysis results see http://www.coenet.us/files/files-do_the_Conclusions_Change_2009.pdf).

The ED re-analysis standardized outcome measures and found positive outcomes with and without project 69 on enrollment and award of financial aid

Positive Impacts on Postsecondary Entrance and Financial Aid With and Without Project 69. The QA re-analysis of the data standardizing outcome measures to expected high school graduation year (EHSGY) found there were substantial and statistically significant positive impacts on postsecondary entrance, application and award of financial aid, and completion of any postsecondary degree or credential with and without project 69. Figure 5 gives an example of these findings for postsecondary entrance after 1 year. Similar impacts were seen for enrollment four years after expected high school graduation year.

Figure 5. Treated on the Treated (TOT) and Intent to Treat (ITT) estimates of impact of Upward Bound (UB) on postsecondary entrance within +1 year (18 months) of expected high school graduation year (EHSGY) 1992-93 to 2003-04



*/**/****/***** Significant at 0.10/0.05/. 01/00 level.

NOTE. Model based estimates based on STATA logistic and instrumental variables regression and also taking into account the complex sample design. Based on responses to three follow-up surveys and federal student aid files.

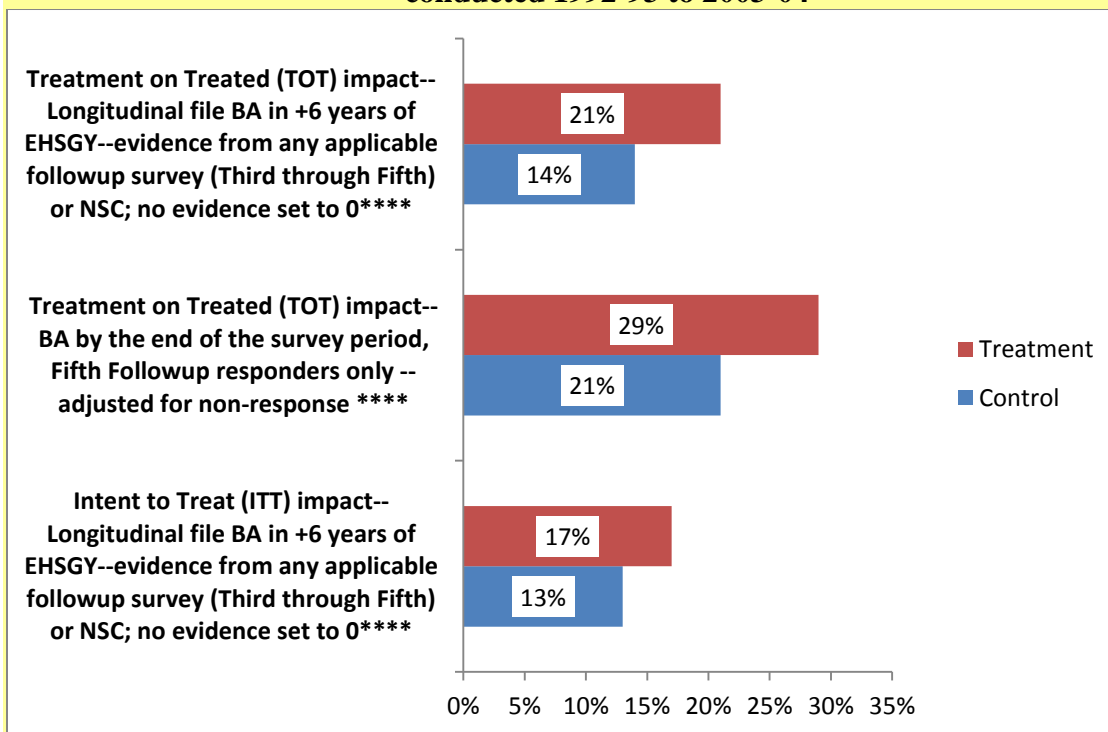
SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, and federal Student Financial Aid (SFA) files 1994-95 to 2003-04. (Excerpted from the *Cahalan Re-Analysis Report*, Figure IV)

BA Attainment Impact Analysis. As noted the representational issues combined with the treatment control group non-equivalency in the heavily weighted project 69 introduced a serious uncontrolled bias into the Mathematica impact estimates. This was especially apparent for BA receipt and could not be addressed adequately by simply standardizing outcomes to expected high school graduation. As noted on average the control group from project 69 resembled Upward Bound Math Science program applicants, being in 10th grade at application, having advanced degree expectations and being more academically proficient. In contrast the treatment group from project 69 on average was comprised of students interested in obtaining certificates,

Among the most impressive of the re-analysis findings was that when the treatment and control group are equivalent, there was a 50 percent increase in BA attainment by 6 years after expected high school graduation date for those students randomly assigned to UB and who participated in the program

more academically at-risk and having lower expectations. In fact, the project 69 treatment group was contributing fully one-third of the weights for those designated as academically at-risk in the overall sample. The PPSS external advisor, Dr. Chromy, recommended basing the BA analysis on the 66 projects that together have a balanced treatment and control group and acknowledging that the study cannot adequately represent the large 4-year and above grantee stratum for which project 69 is the sole representative. The QA re-analysis found that when there is an equivalent baseline treatment and control group as is present when 66 of the 67 projects are taken together, there are also strong positive impacts on BA attainment. As seen in Figure 6, the Treatment on the Treated (TOT) impact analyses revealed that those sampled students randomly assigned to UB and/or who participated in the program had about a 50 percent increase in likelihood of obtaining a BA in six years compared with those not randomly assigned and who did not participate in the program. The Intent to Treat (ITT) estimates found almost a 30 percent increase in BA receipt.

Figure 6. Impact of Upward Bound (UB) on Bachelor’s (BA) degree attainment among low-income and first-generation college applicants to Upward Bound: estimates based on 66 of 67 projects in UB sample: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04



*/**/****/***** Significant at 0.10/0.05/.01/00 level.

NOTE: TOT = Treatment on the Treated; ITT= Intent to Treat; EHSYG = Expected High School Graduation Year; NSC = National Student Clearinghouse; SFA = Student Financial Aid. Estimates based on 66 of 67 projects in sample representing 74 percent of UB at the time of the study. One project removed due to introducing bias into estimates in favor of the control group and representational issues. Model based estimates based on STATA logistic and instrumental variables regression taking into account the complex sample design. We use a 2-stage instrumental variables regression procedure to control for selection effects for the Treatment on the Treated (TOT) impact estimates. ITT estimates include 14 percent of control group who were in Upward Bound Math Science or UB and 20-26 percent of treatment group who did not enter Upward Bound. Calculated January 2010.

Analysis of Control Group Receipt of Alternative Services and Treatment Group Non-Entrance into the Upward Bound Program

Before concluding this report another WWC Standard and key issue needs to be discussed. A major standard of the What Works Clearinghouse and of the random assignment method generally is that the treatment and control group must differ on receipt of the intervention or “the treatment” and that the impact must be attributable to the intervention or no conclusion can be reached. From the beginning of the Upward Bound evaluation, concerns have been raised by participating sites that a large percentage of the control group also had pre-college supplemental services, most frequently other Federal TRIO programs such as Talent Search and even in some cases Upward Bound Math Science—a form of Upward Bound itself. They also reported that often those not randomly selected for the UB treatment group were placed in some other similar service as a substitute for not being randomly selected to be given the UB opportunity.

The majority of the control group also received some form of supplemental pre-college supplemental access services. Most often this was another federal program services such a Talent Search or Upward Bound Math Science

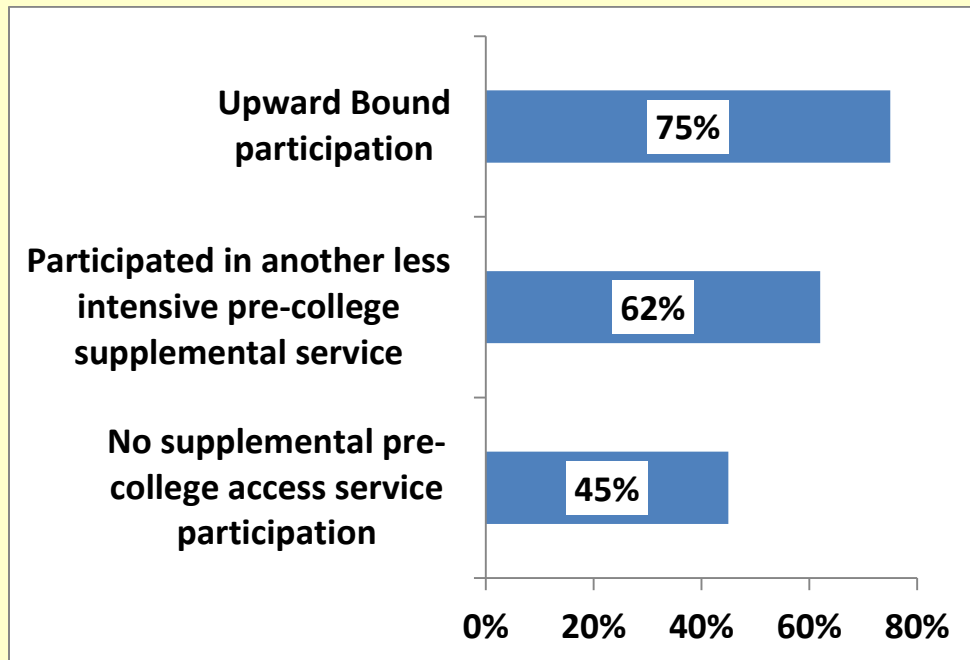
Extent of Receipt of Pre-College Services among the UB Sample. An analysis of the random assignment file, baseline and five follow-up surveys reveals key information about the extent to which the sample members from both the treatment and control group participated in various supplemental pre-college services. The random assignment file reveals that about 26 percent the students randomly assigned to be invited into Upward Bound, were coded as “waiting list dropouts.” All of these cases were kept in the Intent to Treat (ITT) analyses as Treatment cases although it is unclear as to whether most of these students were actually given the “UB opportunity” due to low-income family mobility and other factors. About 20 percent of the Treatment group reported on the First Follow-up Survey that they never entered Upward Bound and a number could not remember being asked to participate. Although about 20-25 percent of the treatment sample did not enter Upward Bound, overall about 92 percent of the treatment group reported receiving some form of supplemental pre-college services (Upward Bound, Upward Bound Math Science, or some other service such as Talent Search). Conversely among the control group about 14 percent reported entering Upward Bound or Upward Bound Math-Science and overall 60 percent of the control group reported some form of supplemental pre-college services in middle or high school by the end of high school. Most frequently for the control group this was reported to be the less intensive federal service, Talent Search. About one-third of both the treatment and control group reported in study surveys that they received supplemental pre-college services such as Talent Search prior to the Random Assignment.

Surprisingly, even well-known scholars such as Haskins and Rouse (2013) misunderstand the information from the Mathematica study, assuming because of its random assignment method that it is a valid indicator of the effectiveness of all college access programs. This conclusion reflects a lack of understanding of the Upward Bound study and is a misuse of the data. As discussed above, the majority of both the treatment and control group in this study had some form of supplemental pre-college services. As noted in most cases the control group had another federal TRIO service such as Talent Search or Upward Bound Math Science. As noted by

Heckman, Hohman, Smith and Khoo (2000), “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.” Considered in this light, some of the internal and external reviewers noted that the Mathematica Upward Bound study might be better analyzed using statistical methods such as two stage instrumental variables regression to observe differences in outcome measures for those who participated in different levels of services.

Below we present results observing differences in outcome variables for three groups: 1) those participating in Upward Bound or Upward Bound Math Science; 2) those participating in some other presumably less intensive pre-college (most frequently the federal Talent Search program); and 3) those reporting not receiving any supplemental pre-college services. A two-stage instrumental variables method was used in which the first stage modeled selection differences between these groups on baseline variables and then these factors were used as control variables in the final models. Figures 7 and 8 respectively present results for postsecondary entrance within one year and for award of BA degree in six years for each of the service groups. Similar impacts were also found for financial aid indicators. As seen in Figure 7, about 75 percent of UB participants entered postsecondary education within one year of expected high school graduation. This compares with 45 percent for students reporting no supplemental service college access services participation and 62 percent for those reporting receiving presumably less- intensive supplemental pre-college services.

Figure 7: Estimates of relative impact of participation in various levels of pre-college access supplemental services on entry into postsecondary education within one year after expected high school graduation: National Evaluation of Upward Bound



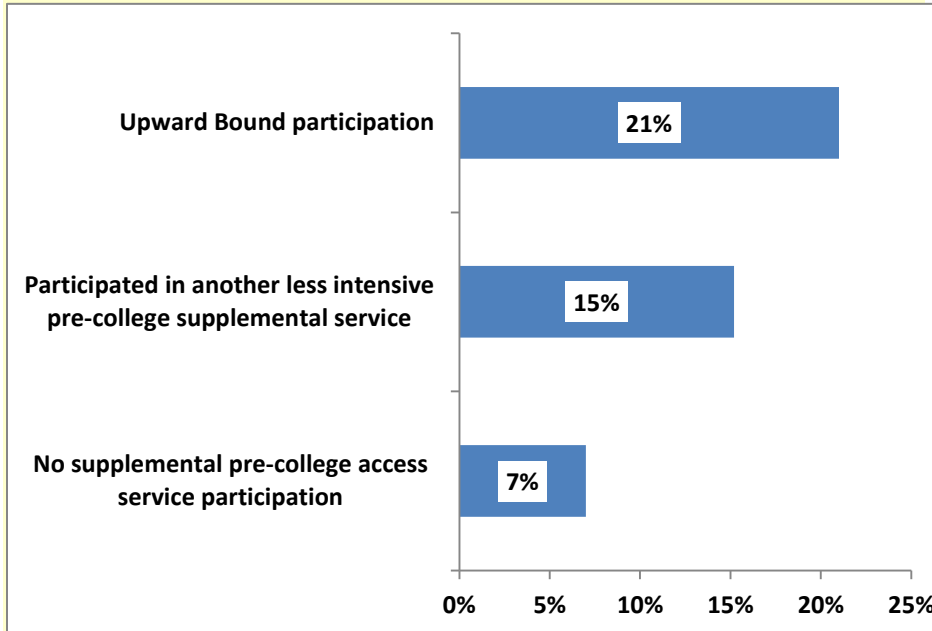
NOTE: Based on data from 66 of 67 projects participating in a Random Assignment Study of about 3,000 middle school and early high school low-income and first-generation UB applicants. The estimates in the figures shown are based on longitudinal data over a 10- year period in an analysis using instrumental two-stage regressions that first model factors related to differences in participation in services and then use these factors in the second stage to control for participation selection bias factors.

SOURCE: Cahalan, Margaret: *Addressing Study Error in the Random Assignment National Evaluation of Upward Bound: Do the Conclusions Change?* The report can be accessed at the following site: [http://www.pellinstitute.org/publications-Do the Conclusions Change 2009.shtml](http://www.pellinstitute.org/publications-Do_the_Conclusions_Change_2009.shtml). The study uses National Evaluation of Upward Bound data files and was sponsored by the Policy and Program Studies Services (PPSS) of the Office of Planning, Evaluation and Policy Development (OPEPD), U.S. Department of Education. Study conducted 1992–99 to 2003–04

As Figure 8 below indicates, among those low-income sample members who reported receiving no pre-college supplemental services, about 7 percent were found to have received a BA degree within six years of their expected high school graduation date. This is very similar to the national data from the National Educational Longitudinal Study (NELS) from the same time period (Ingles et.al. 2002) and also Census Bureau data on the percent of students from families in the lowest income quartile who attain a BA by age 24 (about 7 percent in 2004). Among those sample members not receiving Upward Bound or Upward Bound Math Science (UBMS) but reporting receiving some other type of less intensive services such as Talent Search, about 15 percent had achieved a BA degree by six years after their expected high school graduation. Among those who entered the UB or UBMS program, about 21 percent had attained a BA by six years after the expected high school graduation date (Cahalan, 2009). Thus the instrumental variables regression controlling for selection factors revealed that UB participants were 3.3 times

more likely to obtain a BA in six years when compared to those reporting no participation in college access services and 1.4 times as likely when compared to those who reported participating in other presumable less intensive services.

Figure 8. Estimates of relative impact of participation in various levels of pre-college access supplemental services on BA attainment within 6 years of expected high school graduation: National Evaluation of Upward Bound



UB participants were 3.3 times more likely to obtain a BA in six years when compared to those reporting no participation in college access services and 1.4 times as likely when compared to those who reported receiving less intensive services.

NOTE: Based on data from 66 of 67 projects participating in a Random Assignment Study of about 3,000 middle school and early high school low-income and first-generation UB applicants. The estimates in the figures shown are based on longitudinal data over a 10-year period in an analysis using instrumental two-stage regressions that first model factors related to differences in participation in services and then use these factors in the second stage to control for participation selection bias factors

SOURCE: Cahalan, Margaret: *Addressing Study Error in the Random Assignment National Evaluation of Upward Bound: Do the Conclusions Change?* The report can be accessed at the following site: [http://www.pellinstitute.org/publications-Do the Conclusions Change 2009.shtml](http://www.pellinstitute.org/publications-Do_the_Conclusions_Change_2009.shtml). The study uses National Evaluation of Upward Bound data files and was sponsored by the Policy and Program Studies Services (PPSS) of the Office of Planning, Evaluation and Policy Development (OPEPD), U.S. Department of Education. Study conducted 1992–99 to 2003–04.

Conclusion

Although Mathematica project staff and leadership were sent these fully-documented results in the period of the ED review process of their own final report, and asked to address the concerns raised in the QA review, the results presented above in figures 5 to 8 are not acknowledged in the Mathematica reports. Nor are the seriousness of the representational issues with project 69 or the extent of the treatment control group non-equivalency acknowledged. All impact estimates in the Mathematica reports include project 69, and misleadingly state that the major conclusions do not change substantially because of project 69. Buried in their final report is an admission that results are sensitive to project 69. The report states: “*Because Project 69 had below average impacts, reducing its weight relative to other projects resulted in larger overall impacts for most outcomes compared with the findings from the main impact analysis, which weighted all sample members according to their actual selection probabilities.*” This, however, is also a misleading statement about the effectiveness of project 69. As noted above in Figure 2, a closer look at project 69’s treatment and control group clearly reveals that the so-called “below average impacts” in this project were not due to “project 69’s poor performance” but were due in fact to the extreme differences between the treatment and control group in favor of the control group in this project.

In summary, as Technical Monitors for the study in QA analyses we found that the Mathematica reports are not transparent in reporting study issues and more robust positive results for Upward Bound. Despite being shown “more credible” positive results for Upward Bound that have been replicated, Mathematica continues to report to Congress, the policy research community, and the public unwarranted and non-transparent conclusions concerning the UB program’s effectiveness.¹ This is a very serious matter that needs correcting by Mathematica Policy Research as responsible evaluation contractors and by the US Department of Education.

As noted in 2012, the COE submitted a detailed *Request for Correction* to the US Department of Education. The full text of this request is available at http://www.coenet.us/files/pubs_reports-COE_Request_for_Correction_011712.pdf. As of early 2014, the US Department of Education has refused to consider the COE *Request for Correction* of the Mathematica report, despite the fact that the request was accompanied by an *Statement of Concern* signed by leading researchers that can be found at http://www.coenet.us/files/ED-Statement_of_Concern_011712.pdf. In March of 2014, the co-authors of this paper formally submitted a request to the WWC to rescind its rating of the Mathematica reports as “meets evidence standards without reservations.” We now offer this paper in additional support of these two requests.

¹ In his Nov 19, 2013 Presidential Address to the Association for Public Policy Analysis and Management (APPAM), Mathematica President and CEO, Dr. Paul Decker, presented the flawed data from the 2009 report (Seftor, et. al. 2009) to reaffirm publicly that the UB evaluation study detected no average impacts on UB major legislative goals. He characterized the response of what he called the “Youth Advocacy Community” to the study as constituting “misdemeanors” and “felonies.”

References:

- Cahalan, M. *Addressing Study Error in the Random Assignment National Evaluation of Upward Bound: Do the Conclusions Change?* can be accessed at the following site [http://www.pellinstitute.org/publications-Do the Conclusions Change 2009.shtml](http://www.pellinstitute.org/publications-Do_the_Conclusions_Change_2009.shtml)
- Haskins, R. and Rouse, C. "Time for Change: A New Federal Strategy to Prepare Disadvantaged Students for College," Brookings 2013.
- Heckman, J., Hohmann, N, Smith J., and Khoo, M. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment," *The Quarterly Journal of Economics*, May 2000.
- Horn, L. J., Chen, X., and MPR Associates. "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.
- IES, National Center for Education Statistics (NCES) Statistical Standards---* These may be accessed at the following site url: <http://nces.ed.gov/statprog/>
- IES, What Works Clearinghouse Standards ---*these may be accessed at the following site urls: http://ies.ed.gov/ncee/wwc/pdf/wwc_version1_standards.pdf
<http://ies.ed.gov/ncee/wwc/references/idocviewer/doc.aspx?docid=19&tocid=1/>
- Ingels, S.J., T.R. Curtin, P. Kaufman, M.N. Alt, and Chen, X. *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, 2002.
- Myers, D., and Schirm, A. [*The Short-Term Impacts of Upward Bound: An Interim Report.*](#) Washington, DC: U.S. Department of Education, Planning and Evaluation Service, 1996.
- Myers, D., t Olsen, R., Seftor, N., Young, J., and Tuttle, C. "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., 2004.
- Myers, D., and Schirm, A. "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., 1999.
- Nathan, A.B. *Does Upward Bound Have an Effect on Student Educational Outcomes? A Reanalysis of the Horizons Randomized Controlled Trial Study.* A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Educational Leadership and Policy Analysis) at the University of Wisconsin-Madison 2013 Date of final oral examination: 02/08/2013.

Olsen, R., Seftor, N., Silva, T., Myers, D, DesRoches, D., and Young, J. “Upward Bound Math Science: Program Description and Interim Impact Estimates.” U.S. Department of Education. Washington, DC: Mathematica Policy Research, Inc., 2007.

Seftor, Neil S., Arif, M. and Schirm, A.. “*The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation.*” Report submitted to the U.S. Department of Education. Washington, DC: Mathematica Policy Research, Inc., 2009.

Seastrom, M. *NCES Statistical Standards* (NCES 2003–601). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office, 2002.

Joint Committee on Standards for Educational Evaluation (JCSEE). (widely recognized education evaluation professional standards) (website for which is <http://www.jcsee.org/>)