

### **Abstract of the COE *Request for Correction* for the report:**

*The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation (Referred to as the Mathematica Fifth Follow Up Report)*, prepared by Mathematica Policy Research

COE is submitting this *Request for Correction*, to ask that the above cited report be removed from the ED websites until such time as it can be corrected to meet ED's *Information Quality Guidelines*. The violations noted below and in detail in our *Request for Correction* are serious and have had and continue to have serious adverse consequences for our collective understanding of college access programming and how best to serve the needs of low income and first-generation high school students. They also have adverse effects on the credibility of the field of program evaluation and data driven policy development. The following is a summary of the major ED *Information Quality Guidelines*, *Joint Committee Program Evaluation Standards*, *IES/NCES Statistical Standards*, and *IES What Works Clearinghouse Standards* violations detailed in the COE *Request for Correction*.

1. **Seriously flawed sample design** with only one project of 67 (known as project 69) representing the largest 4-year and above public grantee stratum and given 26 percent of the weight
2. **Misrepresentation of this same largest 4-year and above degree granting institution stratum by a former junior college.** This grantee which awarded largely 2 year and below vocational certificates and which partnered with a job training program was formally classified in the IPEDS sample frame listing as a 4-year institution because it had been taken over by a city college system to serve as a branch campus. This grantee should have been declared ineligible to represent the 4-year stratum due to its lack of 4-year and above degrees. Moreover this lack of adequate representation is never acknowledged in the published reports.
3. **Non-equivalent treatment control group with an uncontrolled bias in favor of the control group** on academic risk factors. This bias is also unacknowledged in the reports. The non-equivalence in the overall sample was largely due to the one project carrying 26 percent of the weight having extreme differences between the treatment and control group on academic risk factors, educational expectations at baseline, and grade at application indicators. Without this project (for the 66 other projects taken together) the treatment and control group are well matched as one would expect in a random assignment study.
4. **Using non-equivalent outcomes measures that were unstandardized as to expected high school graduation year** for a sample that spanned 5 years of expected high school graduation years.
5. **Improper use of the National Student Clearinghouse data when coverage was too low** (at about 25 percent in the most applicable period) and there is evidence of bias due to clustering of non-coverage in the UB grantees in the sample.
6. **Failure to adequately address the issues of control group substitution of services and treatment applicant non-participation issues** with 60 percent of the control group reporting some form of alternative pre-college supplemental services and 26 percent of the treatment group being waiting list drop-outs.
7. **Lack of transparency and misleading statements concerning the major study flaws** and issues of concern such that reviewers and readers of the report are misled into thinking that correct evaluation and statistical procedures were followed.
8. **Failure to acknowledge that when the above issues are addressed using standards based statistical methods that there are strong significant and substantial positive impacts on key outcomes of interest for the Upward Bound program.** These include positive and substantial impacts on postsecondary enrollment and award of federal aid when outcome results are standardized to expected high school graduation year for the entire sample with and without project 69. Importantly they also include a 30 percent Intent to Treat (ITT) and a 50 percent Treatment on the Treated (TOT) increase in BA attainment in +6 years after expected high school graduation for that portion of the sample not represented by project 69.

## Memorandum

**To:** Director, Information Management,  
Office of the Chief Information Officer, U.S. Department of Education  
**RE:** Information Quality Request  
Room 4060, ROB-3, 400 Maryland Avenue, SW  
Washington, DC 20202  
Email: [ocio.infoqualityrequest@ed.gov](mailto:ocio.infoqualityrequest@ed.gov)

**From:** Council for Opportunity in Education

**Date:** 1/17/2012

**Re:** Information Quality Request for the US Department of Education (ED) Office of Planning Evaluation and Policy Development (OPEPD), Policy and Planning Studies Services (PPSS) report entitled: *The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation (Referred to as the Mathematica Fifth Follow Up Report)*, prepared by Mathematica Policy Research

The *Department of Education Information Quality Guidelines*<sup>1</sup> provides a provision for requesting correction of information that does not meet the information quality guidelines. Following the procedure outlined, in that document on page 10, this ***is a formal Information Correction Request made with regard to information contained in the 2009 report The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation (Referred to as the Fifth Follow Up Report), prepared by Mathematica Policy Research.*** (see *Appendix A—copy of conclusions from Mathematica Fifth Follow-up Report*, quoted from the Executive Summary). COE and the partners to this submission, are requesting that the *Mathematica Fifth Follow Up Report* **be withdrawn from the PPSS and OPE linked ED website in a timely manner until such time as it is adequately corrected to meet the Information Quality Guidelines.**

---

<sup>1</sup>Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554) directed the U.S. Office of Management and Budget (OMB) to issue government-wide guidelines that “provide policy and procedural guidance to Federal agencies for ensuring and maximizing the *quality, objectivity, utility, and integrity* of information (including statistical information) disseminated by Federal agencies.” (OMB, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*, February 22, 2002, 67 FR 8452-8460).

## Guidelines Used in Assessing Quality and Making Recommendations for Correction

The guidelines applicable to this report and that have been used in making this judgment are noted below and are referenced throughout this document.

- *U.S. Department of Education Information Quality Guidelines*
- *Joint Committee on Standards for Educational Evaluation (JCSEE)*. (widely recognized education evaluation professional standards) (*Appendix B*—gives a summary of these standards the web site for which is <http://www.jcsee.org/>)
- *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/pdf/wwc_version1_standards.pdf)  
<http://ies.ed.gov/ncee/wwc/references/idocviewer/doc.aspx?docid=19&tocid=1/>

## Sources of Information Other Than the *Mathematica Fifth Follow-up Report* That Are Used in Preparing this Request for Correction

This request relies on the following sources:

1. Independent analysis by Margaret Cahalan<sup>2</sup> published by COE in September 2009. Dr. Cahalan served as Secondary-Postsecondary Team Leader (SPCC) within the PPSS office responsible for Technical Monitoring of the study. Dr. Cahalan's report is entitled: *Addressing Study Error in the Random Assignment National Evaluation of Upward Bound: Do the Conclusions Change?* and can be accessed at the following site [http://www.coenet.us/files/files-do\\_the\\_Conclusions\\_Change\\_2009.pdf](http://www.coenet.us/files/files-do_the_Conclusions_Change_2009.pdf). Dr. Cahalan, joined ED in 2004 and retired from ED in 2011. From October 2004 until January 2007 she supervised the PPSS staff who served as the CORs/Technical Monitors for the final Mathematica UB evaluation contract (known as the Fifth Follow-up contract). In the last 10 months of the contract, from February 2007 to the end of the contract in November of 2007 she also served as the COR/Technical Monitor for the study. (A copy of the major impact findings as presented in the Executive Summary of Dr. Cahalan's report is included

---

<sup>2</sup> Dr. Cahalan while employed as research contractor served as Project Director for the National Evaluation of Student Support Services and the National Evaluation of Talent Search and the NCES and NSF National Surveys of Recent College Graduates. *She also served as Survey Director for the data collection for the Third Follow-up and beginning of the Fourth Follow-up of the National Evaluation of Upward Bound while employed as the Associate Director of the DC Survey and Information Services Division at Mathematica Policy Research from 1996 to 2001.* From 2002 until fall of 2004, Dr. Cahalan served as the RTI Project Director of the NCES National Survey of Postsecondary Faculty (NSOPF). From 2004 to 2011, while employed at ED, Dr. Cahalan served as the Team Leader for the Secondary-Postsecondary Cross-Cutting (SPCC) Team within PPSS, the unit at ED responsible for the National Evaluation of Upward Bound. In the final months of the Fifth Follow-up UB contract she also served as the UB evaluation's COR/Technical Monitor. In 2011, Dr. Cahalan joined the Pell Institute at COE where she currently serves as the Co-Principal Investigator on the ED I-3 grant "Using Data to Inform College Access Programing."

in this request as *Appendix C*. The report is referred to here as the “*Cahalan Re-analysis Report*”).

2. The official final *Performance Report* on the Contractor Mathematic Policy Research prepared by Dr. Cahalan (This is included as *Appendix D*).
3. A personal article/statement by Dr. Cahalan concerning the study posted to the web in 2010 entitled “*Re-analysis finds Strong Positive Results from the Random Assignment National Evaluation of Upward Bound (UB)*.” Washington, DC, June 2010. This article can be accessed at [http://chears.org/publications/fulltext/UBarticlemarch2010\\_4.pdf](http://chears.org/publications/fulltext/UBarticlemarch2010_4.pdf) (This is included as *Appendix E*)
4. Email from Dr. Goodwin to IES review office staff in July 2009 summarizing his views on the Upward Bound report. Dr. Goodwin, recently retired from the Gates foundation was the original study Technical Monitor and Dr. Cahalan’s supervisor within ED at the time of the publishing of the *Mathematica Fifth Follow-up Report*. (The email is included as *Appendix F*)
5. A chronology of events from Spring 2005 including a description of the ED final review process for the *Mathematica Fifth Follow-up Report* which ended with the publication of the report on January 15 2009 (*Appendix G*).
6. Email from Mathematica Project Director and table prepared by Mathematica and sent to PPSS in 2005 concerning project 69 (*Appendix H*)
7. Documentation emails and memo’s prepared by PPSS staff concerning final ED review process for the *Mathematica Fifth Follow-up Report* (*Appendices I to O*)
8. Documentation output from STATA regression models taken from *Appendix B* in the *Cahalan Re-analysis Report*. This output provides credible evidence that contradicts the key conclusions made by Mathematica in the Fifth Follow-up Report. These impacts are found when the study error issues are addressed using *What Works Clearinghouse and NCES standards based statistical procedures* and show results that seriously call into question the Mathematica conclusions put forth in the *Mathematica Fifth Follow-Up* report concerning the Upward Bound program impacts on postsecondary entrance and postsecondary degree attainment. (*Appendix P*)
9. Graph showing interaction effects of project 69 on a key outcome measure (*Appendix Q*)
10. Documentation output from STATA regression models on BA receipt without project 69 (*Appendix R*)

## **Information Requested by ED and OMB**

The specifications on page 10 of the ED Information Quality Guidelines call for provision of the following information.

1. Identification of the requester (i.e., name, mailing address, telephone number, and organizational affiliation, if any);
2. A detailed description of the information that the requester believes does not comply with the Department’s or OMB’s guidelines, including the exact name of the data collection or

report, the disseminating office and author, if known, and a description of the specific item in question;

3. Potential impacts on the requester from the information identified for correction (i.e., describe the requestor's interest in the information and how the requestor is affected by the information in question); and
4. An explanation of the reason(s) that the information should be corrected (i.e., describe clearly and specifically the elements of the information quality guidelines that were not followed).

Below this information is presented organized under these four headings.

***1. Identification of the requester (i.e., name, mailing address, telephone number, and organizational affiliation, if any);***

This request is on behalf of the following requesters:

**Council for Opportunity in Education (COE) represented by Maureen Hoyer**, Executive Vice President of the Council for Opportunity in Education, 1025 Vermont Avenue, N.W., Suite 900 <http://www.coenet.us/> Washington, D.C. 20005-3516, Phone: (202) 347-7430, Fax: (202) 347-0786

**AEEE ---Association for Equality and Excellence in Education, Inc.--Link: <http://www.aeee.org/>** -- AEEE is a professional association comprised of administrators, counselors, teachers, and other staff members in **New York** and **New Jersey** who are involved in programs which promote access, retention and opportunity for disadvantaged and minority youth in high schools and colleges.

**ASPIRE, Inc.--Link: <http://www.aspire-online.org/>**--ASPIRE, Inc. is a professional organization of persons engaged in the administration, recruitment, counseling, instruction, tutoring, or other support services for disadvantaged students. ASPIRE works to advance and defend the ideal of equal educational opportunity. States which form ASPIRE include **Colorado, Montana, Wyoming, North Dakota, South Dakota, and Utah.**

**CATP--Caribbean Association of TRIO Programs---**CATP represents Puerto Rico and the U.S. Virgin Islands. CATP is committed to the development and excellence of the TRIO programs. CATP promotes access, retention and opportunity for disadvantaged and minority youth in Puerto Rico and the U.S. Virgin Islands.

**MAEOPP--Mid-America Association of Educational Opportunity Program Personnel--Link: <http://www.maeopp.org/>** --MAEOPP represents 10 states: **Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, Ohio, and Wisconsin.** MAEOPP members include state colleges and universities, independent colleges, vocational-technical schools, state and local educational agencies, community colleges, public and private high schools, and community agencies. MAEOPP strives to promote

secondary and postsecondary support programs to meet the academic, financial, and socio-cultural needs of minority and/or disadvantaged students.

**MEAEOPP--Midwestern Association of Educational Opportunity Program Personnel--Link:**

<http://www.meaeopp.org/> --MEAEOPP is a regional association composed of educational opportunity program personnel in the District of Columbia, Delaware, Maryland, Pennsylvania, Virginia, and West Virginia. The mission of the Association is to promote equal educational opportunity and greater accessibility to higher education for non-traditional students, who, by reason of socio-economic status, ethnic definition, physical handicap, and/or restricted cultural-educational experience, find themselves in a disadvantaged position with traditional or conventional students.

**NASP--Northwest Association of Special Programs--Link:** <http://www.nasp-trio.org/> --

NASP is an organization representing professional educators who work with low-income, first-generation, and disabled students throughout the Pacific Northwest. NASP promotes the development, improvement, and extension of education to nontraditional students living in Oregon, Washington, Alaska, and Idaho.

**NEOA--New England Educational Opportunity Association--Link:** <http://www.neoonline.org/> --

NEAO serves as a powerful and respected voice for equal educational opportunity. The Association is committed to playing a leading role in advocating for non-traditional, low-income, minority, and physically disabled students. NEAO represents and serves institutions and students in the states of Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.

**SAEOPP--Southeastern Association of Educational Opportunity Program Personnel--Link:**

<http://www.saeopp.org/> --SAEOPP is composed of educational opportunity program personnel in Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee. SAEOPP seeks to bring together those persons who have an active interest in becoming professionally involved in broadening access to and retention in postsecondary education.

**SWASAP--Southwest Association of Student Assistance Programs--Link:** <http://www.swasaptrio.org/>

--SWASAP is an association whose purpose is to advance and defend the ideal of equal educational opportunity for all TRIO eligible individuals by providing its members the guidance and support needed to foster those activities necessary to cause the target population to persist in achieving their educational goals. SWASAP represents the states of Louisiana, Arkansas, New Mexico, Oklahoma, and Texas.

**WESTOP--Western Association of Educational Opportunity Personnel--Link:** --

<http://westop.csuchico.edu/> -- WESTOP is an association whose purpose is to bring together individuals interested in furthering educational opportunities for disadvantaged students. WESTOP represents the states of Arizona, Hawaii and the Pacific Territories, Nevada, and California (Northern California Chapter, Central California Chapter, Southern California Chapter).

2. ***A detailed description of the information that the requester believes does not comply with the Department's or OMB's guidelines, including the exact name of the data collection or report, the disseminating office and author, if known, and a description of the specific item in question;***

***Exact Name of the Report:*** *The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation (Referred to as the Mathematica Fifth Follow Up Report)*<sup>3</sup>, This report was placed on the ED website in January of 2009 and is accessed at the following address: (<http://www2.ed.gov/about/offices/list/oepd/ppss/reports.html#higher>). *Appendix A* is a copy of the conclusions from this report as published in the Executive Summary). This request for correction specifically focuses on the *Mathematica Fifth Follow-up Report*. However, similar issues pertain to the *Mathematica Third Follow-up* report published by ED in 2004 and also currently posted to the PPSS website.

**Disseminating Office:** Office of Planning, Evaluation and Policy Development (OPEPD), Policy and Program Studies (PPSS), Program and Analytic Studies (PAS) Division

**Authors:** Neil Seftor, Arif Mamum and Alan Schirm of Mathematica Policy Research contractor conducting the study under three contracts to the Department the last of which was Number ED-04-CO-0152.

***Description of the specific items in question:***

**Major Impact Estimates Reported in Report are “Seriously Flawed”**

This request for correction is being made because of serious technical flaws in the Mathematica Upward Bound evaluation analyses and reports. *These flaws are significant technically and are serious enough to affect the major conclusions that are made from the study concerning the Upward Bound program's effectiveness.* Therefore we believe it is imperative that they be addressed. Those PPSS staff most knowledgeable and responsible for the Technical Monitoring of the study, have stated prior to and after the report was published in 2009 that the report is flawed. As Dr. Cahalan wrote explaining her reason for posting a statement of concern about the report (see *Appendix E*).

*Put simply, as the Department of Education (ED) Technical Monitor for the final period of the study, in a QA examination of the data, confirmed by consultation with independent statistical experts, I found that the postsecondary results in the Mathematica report suffer from inadequately controlled for bias in favor of the*

---

<sup>3</sup> The National Evaluation of Upward Bound was conducted under three contracts to Mathematica Policy Research over the period of 1992-2007. The last and final report (the *Mathematica Fifth Follow-up Report*) was published by ED in January 2009 in the last week of the Bush Administration after 18 months in the ED review process. It was published over the objections of Dr. Cahalan, as the PPSS Technical Monitor, and over a formal disapproval by the Office of Postsecondary Education (OPE) in the ED Ex Sec Review Process.

*control group, and a sample with serious representational flaws and unequal weighting issues. In the interest of full government transparency this article is also written as an effort to comply with professional evaluation standards that state:*

*The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results, (Standard P6 in Joint Committee on Education Evaluation Standards).*

*I also wish to share some alternative results that I believe, give a more credible representation of the findings from this evaluation. These re-analyses show that when identified study flaws are addressed using standard statistical procedures there are positive impacts for the UB program on the goals of the program—postsecondary entrance, award of financial aid, and degree or credential attainment (Appendix E).*

**A description of the major issues with the Mathematica Fifth Follow-up Report is also provided in the final Performance Report for the Contract (see *Appendix D*)** and in numerous emails and memo's written by Drs. Cahalan and Goodwin at the time of the final review of the report (*Appendices I to O*). Dr. David Goodwin who was Dr. Cahalan's supervisor in the review period and the former Director of the Policy Analysis Services (PAS) of PPSS who was the original UB study COR/Technical Monitor,<sup>4</sup> also wrote some months after the *Mathematica Fifth Follow-up Report* was published: "I believe the *Mathematica report (Fifth Follow-up)* currently published by ED is seriously flawed." (July 2009 email memo to IES—see *Appendix F*).

This request relates to the major impact estimates reported in the tables and the accompanying text discussion of these results in the *Mathematica Fifth Follow-Up Report*. However, we wish to emphasize that while the study design and implementation flaws we discuss in detail in Section 4 of this request do not allow for robust and unbiased estimates for the entire applicant population, as the IES external reviewer C stated, reasonably robust and unbiased estimates are possible for that share of the applicant population constituting about 74 percent of Upward Bound applicants at the time. The identified issues are correctable using standards based, clear and transparent statistical methods. Once these limitations are openly acknowledged and addressed following statistical and evaluation research standards we believe that the Upward Bound study can meet the Department's information quality guidelines with regard to quality, objectivity, utility, and integrity.

Specifically, the impact estimates in the following tables and accompanying text have been found in QA examination to be in need of correction.

---

<sup>4</sup> Dr. Goodwin retired from the Department of Education in September of 2009 to join the Gates Foundation



### **Specific Items/Sections of *Mathematica Fifth Follow-up Report* Identified as in Need of Correction**

- a. Executive Summary (Conclusions and accompanying discussion of study methods (*Appendix A* contains the conclusions as presented in the Executive Summary))
- b. Chapter I (Discussion)
- c. Chapter II (Discussion)
- d. Chapter III (Tables III.1; III.2; III.3) and accompanying discussion
- e. Chapter IV (Tables IV.2; IV.3; IV.4) and accompanying discussion
- f. Appendix C (Tables C-1 to C-14)
- g. Appendix E (Tables E-1 to E-8)
- h. Appendix G—Discussion and tables concerning project 69
- i. Appendix I (Tables I-1 to I-84) (subgroup impacts)

In Section 4, (following Section 3 below on potential impacts should the errors not be addressed) we present details and recommendations for correction related to the major items in need of correction in the *Mathematica Fifth Follow-up Report*.

3. **Potential impacts on the requester from the information identified for correction (*i.e.*, describe the requestor's interest in the information and how the requestor is affected by the information in question);**

The Department's *Information Quality Guidelines* indicate that government information that is particularly influential needs to meet higher quality standards. Per the OMB guidelines, information is designated as influential if the Department determines that the information is reasonably likely to have a clear and substantial impact on public policies or private sector decisions if disseminated. Scientific, financial, and statistical information all may be considered influential.

The correction of the Mathematica UB report is critical to three communities most affected by ED publishing a report that does not meet information quality guidelines. These three communities are:

1. **The Congress of the United States**, responsible for legislation providing for the TRIO programs and on-going yearly funding authorizations and appropriations and the **Office of Postsecondary Education (OPE)** responsible for administration and regulatory supervision of the grant program; and
2. The immediate stakeholders in the UB evaluation (**potential low income and first generation college students and their UB grantee service providers**); and
3. **The evaluation research community** (including government offices responsible for evaluation such as IES and PPSS, research contractors and consultants, and the academic community)

The Council for Opportunity in Education (COE) and its member organizations are major representatives of the interests of low income and first generation students and college access service providers whose interests have been and are continuing to be seriously adversely affected by the information reported in the *Mathematica Fifth Follow-up Report* that does not meet EDs Information Quality Standards.

The Department of Education has published and Mathematica Policy Research has authored four contractor reports containing estimates derived from the study (Myers and Schirm 1996; 1999; and Myers et. al. 2004; Seftor et. al. 2009). The first of the reports focused on high school outcomes and the second two on college entrance and completion. Mathematica concluded in the *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). The *Mathematica Third Follow-up Report* also reported that there were significant and large effects for the bottom 20 percent of study participants on 9<sup>th</sup> grade academic indicators and for students with lower expectations (defined as expecting less than a bachelor’s degree at baseline). The unpublished *Mathematica Fourth Follow-up Report* and the *Mathematica Fifth Follow-up Report*, conclude similar lack of impact for postsecondary entrance. In the *Mathematica Fifth Follow-up Report*, significant positive effects for the award of certificates were reported, but were not reported for the award of the bachelor’s degree or award of any degree or certificate (see *Appendix A--Conclusions from Mathematica Fifth Follow-up Report*, Executive Summary).

***The conclusions reported by Mathematica have been extensively used by ED and OMB in decision making concerning the program over a period of more than 10 years.*** In the context of the published reports noted above reporting lack of overall effects, but findings of large significant effects for sub-groups of students determined to be at a “higher academic risk” and reporting “lower baseline college expectations,” the Office of Management and Budget (OMB) urged the program to enact improvements increasing the targeting of the program to students more at academic risk. In response, in 2003 the Department of ED developed the “Upward Bound Participant Expansion Initiative” designed to provide additional funding to projects to serve students deemed to be more “at risk.”

***On the basis of the Mathematica study reports, the program was given an “ineffective” rating by OMB in the Program Assessment Rating Tool (PART). Justified on the basis of the PART rating, the FY2005 and FY2006 federal budgets called for zero funding of Upward Bound, Talent Search (TS), and GEAR UP.*** The recommendation for zero funding was dropped in the FY07 and FY08 budgets. However, the misinformation contained in the study continued to be used to justify FY09 and FY10 level funding in a period when there were large increases in other federal education programs and when increasing college enrollment and completion was a stated to be a national priority.

In 2006, ED began work on a new random assignment study described in the ***Absolute Priority for Upward Bound Program Participant Selection and Evaluation*** published by the Department of Education in the Federal Register on September 22, 2006. Based on the Mathematica study findings, the Absolute Priority would require that one-third of the students be

defined as academically at risk as evidenced by low GPA or not passing a high school competency test and placed requirements for 9<sup>th</sup> grade entrance into the program. The new study planned to use much the same methodology to evaluate the priorities that had been used in the first evaluation. This study was cancelled by ED in February 2008 following Congress prohibiting further funding of the study. Subsequently a new IES study was begun and is now on-going.

**This request for correction is being made precisely because of the influence this \$14 million study has had and continues to have on TRIO and the evaluation research communities.** As the UB evaluation has been portrayed as meeting the highest quality random assignment study standards its findings have had considerable influence. Results of the study continue to be widely quoted in professional literature without the authors' knowledge of the very serious issues concerning the validity of the Mathematica reports. For example, in 2009 the American Youth Policy Forum (AYPF) published an informative report entitled *Success at Every Step*, reviewing research results concerning 23 pre-college programs (<http://www.aypf.org/publications/SuccessAtEveryStep.htm>) in which the Mathematica study is quoted extensively and the lack of overall positive results for Upward Bound is noted. Another example is the April 2011 request for OMB review of the new IES evaluation of Upward Bound being conducted by the contractor DIR, Inc. The background information in the OMB Justification Section, cites the findings from the Mathematica report of no overall impacts on postsecondary outcomes with no recognition of the serious validity issues concerning the study that have been raised or the alternative positive impact analysis findings (see COE's response in the Public Comment submitted May 20, 2011).

*COE and its member organizations believe it is irresponsible on the part of the US Department of Education and a disservice to the field of evaluation research for ED to publish reports sent to Congress which violate the basic accuracy, utility, and proprietary standards of the Joint Committee on Program Evaluation Standards, NCES Statistical Standards, and the What Works Clearinghouse Standards for causal inference studies.* We describe in detail, the reasons the Mathematica reports from the study do not meet the Information Quality Guidelines or commonly accepted standards for evaluation research.

4. **An explanation of the reason(s) that the information should be corrected (i.e., describe clearly and specifically the elements of the information quality guidelines that were not followed).**

**Weight of Evidence Approach.** The U.S. Department of Education *Information Quality Guidelines* state that *"In the Department of Education's correction request process, the burden of proof rests with the requester."* In the document below we present a "weight of evidence" case that includes two parts: 1.) Presenting evidence for why the report needs correcting and observation of quality guidelines or standards not followed, and 2.) Presenting "standards based" technical information as to how the report should be corrected to meet program evaluation and statistical standards.

## Major Applicable Quality Guidelines and Standards

The Upward Bound Evaluation employed an unusual and difficult design. The UB evaluation combined a random assignment experimental design with what was intended to be a nationally representative two-stage sample capable of making design based inferences (using inverse probability of selection weights) to the national Upward Bound program and also to key sub-groups of applicants. Presumably this design was attempted because there was a desire to report to Congress and other interested parties on the effectiveness of the national federal program. The four sets of guidelines and standards, identified in Section 1 of this request as most applicable to the issues of concern (*ED Quality and Information Guidelines, Joint Program Evaluation Standards, NCES Standards, and What Works Clearinghouse Standards*) have much overlap in the requirements they specify concerning objectivity, accuracy, utility, and integrity or propriety. Exhibit 1 identifies some key information quality guidelines and specific standards from these sources especially relevant to the *Mathematica Fifth Follow-up Report* issues.

The quality requirements and standards most relevant to this application for correction can be summarized as follows.

- A. *The sampling follows correct procedures, is checked and is found to be representative of the population for which the study is intended to generalize;*
- B. *The treatment and control group are equivalent on factors related to outcomes at the baseline and throughout the study (equivalence);*
- C. *The outcome measures are valid and provide the same standard for all sample members*
- D. *The study must achieve adequate overall and differential coverage and response, and must concern itself with differential attrition, non-response and non-coverage bias issues of all the data sources used relative to the population of interest*
- E. *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;*
- F. *The study design, data collection, analyses and reporting have integrity and the study is transparent, replicable and provides full documentation of procedures and pertinent limitations and sources of error*

**Exhibit 1**  
**Key Information Quality Guidelines and Standards that are Applicable to the Concerns with**  
**Regard to the *Mathematica Fifth Follow-up Report***

***Department of Education Quality Information Guidelines***

**Research and Evaluation information products should, at a minimum: ...**

- Pose the research or evaluation question in a balanced and unbiased manner;
- Provide an unbiased test of the question; ...
- Present conclusions that are strongly supported by the data; ....
- Confirm and document the reliability of the data, and acknowledge any shortcomings or explicit errors in any data that is included;

**Statistical Information Guidelines**

- The source of data should be reliable. The sample should be drawn from a complete list of items to be tested or evaluated, and the appropriate respondents should be identified, *correctly sampled*, and queried with survey instruments that have been properly developed and tested
- .....Appropriate steps should be taken to *ensure that the respondents are a representative sample*;
- Data should be capable of being reproduced or replicated based on information included in the documentation,...such as identification of other sources of potentially corroborating or conflicting information.

**What Works Clearinghouse Handbook of Procedures and Standards**

A study may fail to meet WWC evidence standards if .....

- It does not include a valid or reliable outcome measure, or does not provide adequate information to determine whether it uses an outcome that is valid or reliable. ....
- The intervention and comparison groups are not shown to be equivalent at baseline
- The overall attrition rate exceeds WWC standards for an area.
- The differential attrition rate exceeds WWC standards for an area.
- The measures of effect cannot be attributed solely to the intervention.....

**NCES Statistical Standards Concerning Non-Response and Coverage**

- **STANDARD 2-2-4:** A nonresponse bias analysis is *required* at any stage of a data collection with a unit response rate less than 85 percent. The extent of the analysis must reflect the magnitude of the nonresponse (see Standard 4-4).
- **STANDARD 3-1-2:** NCES data collections that are used as sampling frames for other NCES surveys must strive for coverage rates in excess of 95 percent overall and for each major stratum. **STANDARD 3-1-3:** .... If there is not evidence of a coverage rate of at least 85 percent of the target population, then frame enhancements such as frame supplementation or dual frame estimation must be incorporated into the survey study design.

**Joint Program Evaluation Standards:** The Joint Committee Program Evaluation Standards (see *Appendix B* for a summary) address Integrity under the heading of Propriety. Standard P6 noted below discusses the full disclosure of findings

- **P6 Disclosure of Findings** The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results

In the section below we discuss the Mathematica report with reference to each of these requirements and standards.

#### **4-A The sampling follows correct procedures, is checked and is found to be representative of the population for which the study is intended to generalize**

##### **4-A-1 Reasons Report Does Not Meet Guidelines and Standards in the Area**

**Sample Design Flaws and Incorrect Sampling Procedures.** The sample design has been consistently described by PPSS internal and external reviewers as seriously flawed with only one single project (known as project 69) selected to represent the largest study defined 4-year stratum, thus violating basic standards for probability sample design. This project's sample members had extreme weights (for example the sum of the weights for this project was 11,536 compared with a sum of the weights of 20 for the smallest sum of the weights project—see *Exhibit 2*) and carried fully 26 percent of the total weight for the study. As one IES external reviewer noted with regard to the flawed sample design:

*“The decision made in 1992-94 to select only a single project at random from this 26% share of the applicant population created a design in which design-based estimation and inferences for the full population could not be robust for the true population values. Simply applying a population weight to an inadequate sample of one cluster from a 26% share stratum will not correct this.*

*What to do? With respect to design-based inferences for all other strata, the baseline sample of programs should enable robust inferences for that share of the UB survey population not included in the Project 69 stratum.” (IES external statistical reviewer C July 2008)*

**Failure to acknowledge the incapacity of project 69 carrying 26 percent of the weight to adequately represent it's 4-year grantee stratum.** In addition to the incorrect sampling procedures cited above, notably, the project 69 with 26 percent of the weight was found (at the end of the Mathematica study contract when the frame list by stratum was delivered to ED<sup>5</sup>) to be atypical of the large diverse 4-year public stratum for which it is the sole representative—and should have been excluded as ineligible to represent its stratum. The contractor, Mathematica failed to report this fact in any of the reports from the study and to the contrary reported that the project was “typical” of its stratum. This may be part of the reason these representational issues were not observed by the IES external reviewers. In the *Fifth Follow-up Report* Executive Summary Mathematica states:

*“Project 69 was similar to other projects in this stratum on a broad range of characteristics. Similarly data from the student surveys and NSC and FSA records indicated that the students from project 69 did not have unusual characteristics” (Executive Summary Mathematica Fifth Follow-up Report, Page xvii- xviii).*

---

<sup>5</sup> Dr. Cahalan has stated that she requested this frame from Mathematica 10 months before it was finally delivered to PPSS in December 2007 after the contract ended. She requested the frame for Dr. James Chromy, a Fellow of the American Statistical Society and the PPSS Statistical Technical Assistance contractor from RTI, to provide QA and external expert statistical consultation to PPSS on the sample design and project 69 problem.

**Exhibit 2. Percentage distribution of sum of the weights by project for the 67 projects making up the study sample: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04**

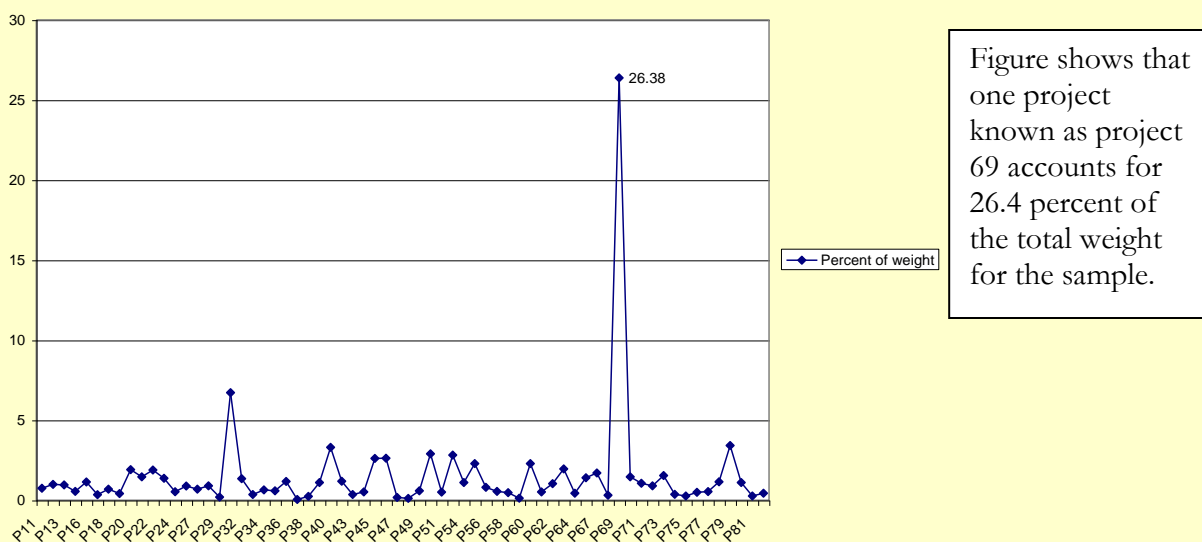


Figure shows that one project known as project 69 accounts for 26.4 percent of the total weight for the sample.

**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 Program Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), U.S. Department of Education; study conducted 1992-9 to -2003-04. (Taken from Cahalan, “Re-Analysis Report”, 2009, Figure 1 )

The Mathematica statement above and the one below gave the false impression to external IES reviewers and readers that project 69 was problematic only for its large weight and carefully do not reveal the full representational issues or the extreme differences between the treatment and control group in this project (to be discussed below). The Mathematica report goes on to state that analyses that omit project 69:

*Do not appropriately represent the most common stratum of Upward Bound projects. Thus such analyses do not answer the evaluation’s research questions about the impacts of the national Upward Bound program. Moreover the estimates for such analyses do not generalize to urban projects, large projects or any other well-defined subset of projects for which the findings might have policy implications. In contrast the findings from the main impact analyses, which include all projects weighted based on their selection probabilities are intended to generalize to the national Upward Bound program” (Executive Summary, Mathematica Fifth Follow-up Report, page xviii)*

The atypical representation of project 69 as the sole representative of the large and diverse 4-year stratum is reported in Dr. Cahalan’s COE publication; however, the serious mis-

**representational issues with project 69 are masked in the Mathematica report.** Concerns with project 69 representational issues are also noted in the email from Dr. Goodwin of July 2009 to IES in which he summarizes his reasons for stating the report is “seriously flawed.” Dr. Cahalan and Dr. Goodwin found when Mathematica finally sent the sampling frame list of eligible grantees to ED and the identity of project 69 became known to them at the end of the contract in December of 2007, that the project with 26 percent of the weight that was supposedly the sole representative of the largest 4-year stratum was in fact a former junior college that had been taken over by a city college system. Although officially classified as a 4-year institution it had largely 2-year and below certificate offerings. The UB program partnered with a job-training program and the grantee institution awarded certificates in among other things construction trades. The large diverse frame stratum for which project 69 is the sole representative includes the major research universities that had UB grants at the time as well as 4-year majority white and majority black liberal arts colleges. It also did not have the hallmark 4-year grantees’ UB summer residential program present in virtually all projects within the 4-year stratum it was representing—as it has no on-campus housing.

**Correct sampling procedures require a test of a sample to ensure that the sample drawn can represent the stratum on known characteristics likely to be related to the estimates of interest.** If the sample is found not to be representing the stratum on known characteristics defining the stratum and likely to produce biased estimates for the total stratum then, best practice calls for revising the sample design, or if that is not possible making sure that the cases selected are “eligible” to represent the population of interest. Dr. Chromy, *A Fellow of the American Statistical Society* who provided external statistical expertise consulting with Dr. Cahalan and Dr. Goodwin concerning the sampling issues in 2006-07, and who more recently (fall 2011) has reviewed this COE request for correction has noted that in cases in which the sampling frame is imperfect, cases selected found to not meet the eligibility requirements for the stratum of being a grantee that was a primarily a 4-year and above granting institution should be excluded as ineligible to make valid estimates.

Dr. Chromy states:

*“ If representing the stratum of four-year institutions was a clearly stated objective of the study, than a sample of size 1 for this group is clearly inadequate..... With an imperfect sampling frame, it would be an accepted practice to check each project drawn and drop it from the sample if it does not meet the study population definition; this is a form of screening for eligibility. (James Chromy comments on the COE request for Correction, October 2011).”*

**It’s probable that these eligibility checks were never done relative to the most important sampling factor defining the stratum (that it was actually a UB grantee with mostly 4-year programs).** In addition the diverse non-Hispanic majority stratum was defined to include UB grantees that were majority white and grantees that were majority black. It is logically not possible for one project to represent projects that have both majority white and majority black participants. In this case there are no white sample members in project 69 as it is 60 percent black and 40 percent Hispanic. **It’s clear the sample as designed and implemented cannot support estimates for this 4-year non-Hispanic stratum that was supposedly the sole representative of 26 percent of UB at the time.** Unfortunately contrary to these expert reviewers’ advice, Mathematica based



all its conclusions concerning the program on the basis of this flawed design with an “atypical for its 4-year stratum” project carrying 26 percent of the total weight.

**As noted above, not all the facts concerning project 69’s representational issues are acknowledged in the Mathematica reports.** This misrepresentation of a largely 2-year and less than 2-year grantee as the sole representative of the largest 4-year public stratum, combined with the extreme large weight and the uncontrolled for academic factor bias in favor of the control group in this project’s sample (to be discussed below) contributes to a type I error of over-estimating the impact of UB on the attainment of certificates of the type awarded by the project 69 grantee institution and a type II error of failure to detect effects for the attainment of Bachelor’s degrees. Very strong positive effects on BA attainment were found for 66 of the 67 sampled projects taken together, which were found to meet WWC standards for baseline equivalence and were well matched when taken together on relevant attributes (see Exhibits below), but not when project 69 is included in the impact estimates.

#### **4-A-2 Correction Required to Meet Standards and Information Quality Guidelines in Area**

Following the PPSS Technical Monitor and an IES external reviewer recommendation, the report needs to acknowledge that due to sample design flaws and given the fact that there was also a “bad draw” of a project atypical of its stratum as the sole representative of the stratum that the Mathematica UB evaluation study is not capable of providing robust unbiased estimates for the entire population of UB applicants. Impact estimates presented and discussed in the text body should be re-done to exclude project 69 and should be clearly labeled as not representing the entire population of applicants. *The study needs to acknowledge that it cannot represent this 4-year grantee stratum and that impact estimates reported are only for the 74 percent of the population for which (as noted by the IES reviewer) there can be made reasonably robust impact estimates.* Estimates including project 69 can be placed in an appendix for methodological interest, but should not be used to judge the effectiveness of the UB program. The discussion presented below provides more evidence of why estimates with project 69 should not be used to make effectiveness judgments about the UB program.

#### **4-B. The treatment and control group are equivalent on factors related to outcomes at the baseline and throughout the study (equivalence)**

The major strength of the random assignment design is that it is supposed to ensure that the treatment and control group are equivalent on factors related to outcomes. As the What Works Clearinghouse standards note:

*“In an RCT, researchers use random assignment to form two groups of study participants. Carried out correctly, random assignment results in groups that are similar on average in both observable and unobservable characteristics and any differences in outcomes between the two groups are due to the intervention alone, within a known degree of statistical precision.” (What Works Clearinghouse Handbook of Standards and Procedures, version 2 page 13)*

If the treatment and control group are not equivalent there will be a bias (a factor taking the estimate away from the true estimate in a systematic or non-random manner) in the estimates of impact. **This equivalence of the treatment and control group is a required standard for quality random assignment studies.**

#### **4-B-1 Reasons Report Does Not Meet Guidelines and Standards in the Area**

**The overall national estimates contained in the Mathematica Fifth Follow-up report contain uncontrolled for bias in favor of the control group and do not meet the *What Works Clearinghouse* and basic random assignment study requirements that the treatment and control groups be equivalent on factors related to the outcomes being assessed.** Cahalan reports that in her QA examination she found that with project 69 included (carrying 26 percent of the weight), there is evidence that there is an uncontrolled for or inadequately controlled bias in favor of the control group in the overall national estimates on academic risk factors, grade at entrance into the UB program and educational expectations. The source of this bias in the overall estimates is the large non-equivalencies of project 69's treatment and control group. The other 66 projects in the sample when taken together have equivalence on these factors.

For unknown reasons, the treatment group in project 69 resembled on average a vocational technical oriented group consistent with the project 69 programming and the control group resembled more the Upward Bound Math Science (UBMS) population. A closer look at project 69 indicates that there are large differences between treatment and control group members on characteristics likely to affect postsecondary outcomes, which cannot be adequately controlled in the statistical analysis. These variables include the control group having: higher educational expectations, a higher grade at baseline, to be more likely to have algebra or above in 9<sup>th</sup> grade, and to be less likely to be classified as a high academic risk, (*Exhibit 3*). **This source of bias is unacknowledged and masked in the Mathematica report and contributes to a type II error of failure to detect impacts on key outcomes.**

These differences between the treatment and control group profiles in project 69 explain the seemingly negative impacts in this site that Cahalan reports were observed and communicated to PPSS by a former Mathematica lead analyst, Dr. Elizabeth Stuart, in 2005 with regard to the unpublished *Fourth Follow-up Report*. Dr. Stuart noted that project 69 had negative impacts and that the conclusions for the study changed depending on this case (see Appendix H). Cahalan reports that PPSS asked Mathematica to qualitatively and quantitatively look into this grantee to help explain why this project would have had such negative impacts. Not receiving a response from Mathematica, Cahalan requested copies of the data files and conducted her own QA analysis in the period after 2006. ***Cahalan found that differences between the treatment and control group in project 69, combined with the large weight, compromised the overall equivalency of the treatment and control groups on factors related to outcomes.*** Moreover these were inadequately controlled for in published analyses. Without project 69 the treatment and control groups are seemingly well matched in terms of these variables (*Exhibit 3*).

Another way of demonstrating the non-equivalency leading to bias is presented below in Exhibits 3 to 5. As the *What Works Clearinghouse Standards* specify in a random assignment study there should be about a 50-50 balance between treatment and control group in the percent of sample members with attributes related to study outcomes. As can be seen in Exhibits 4 and 6, this balance was compromised by the large imbalances in project 69. Given that project 69 carries 26 percent of the weight, the overall sample is unbalanced with 58 percent of the high-risk students in the treatment group and 42 percent in the control group (Exhibit 4). The extreme imbalances in project 69 are seen in Exhibit 5. *For example, in project 69, 80 percent of the high academic risk students were in the treatment group and 20 percent in the control group.* The sample without project 69 when taken together is more well balanced with for example 51 percent of the high risk students being in the treatment group and 49 percent in the control group (Exhibit 6).

The extreme differences observed in project 69 might have occurred by chance within a small sample; however, project 69 has a sample of 85 members. One of the IES external reviewers noted that, while it cannot be proven, it may be that the random assignment was not implemented correctly or broken in this site for whatever reason. The project has two sub-stratums of unknown defining characteristics and hence has large unequal weighting within the stratum itself. It also has more control group members (55 controls and 30 treatments) than treatment members as the number of applicants was larger than twice the number of openings—the weights were adjusted to make the treatment and control total weights equal (see detailed discussion in *Cabalan Re-analysis Report* page 20). There were also several sample members from project 69 who were selected with certainty into the UB program (so called “must serves”). These cases were removed from the study and their weights re-distributed among those randomly selected adding more complexity and perhaps contributing to some of the treatment/control imbalance observed in project 69.

**Exhibit 3. 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 to 2003-04**

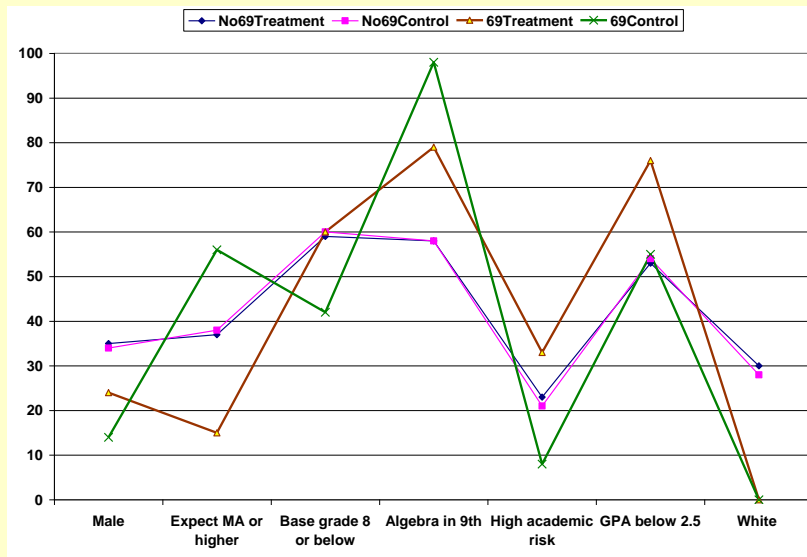
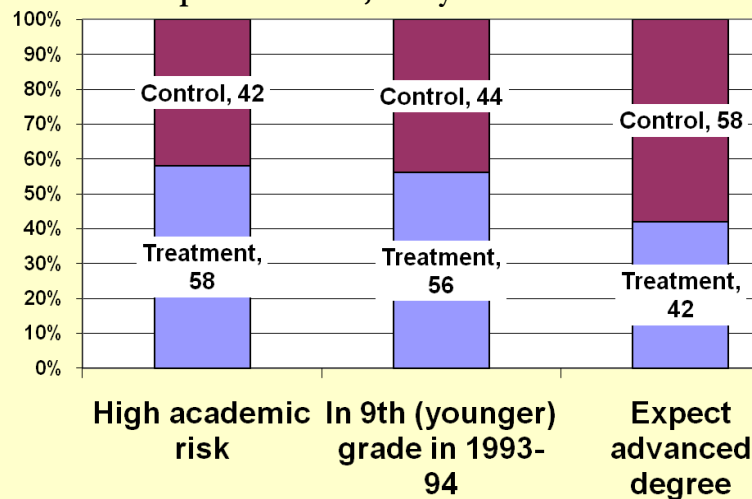


Figure shows that the UB treatment and control group are well matched without project 69 on the variables in the chart; however, in project 69 the treatment and control group manifest substantial differences. For example, 56 percent of the control group in project 69 expected an MA or higher at baseline compared with 15 percent of the treatment group. In contrast, among the other 66 projects in the sample, 38 percent of the control group and 37 percent of the treatment group expected an MA or higher.

**NOTE:** Project 69 tabulation based on the 85 sample cases from project 69 (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.

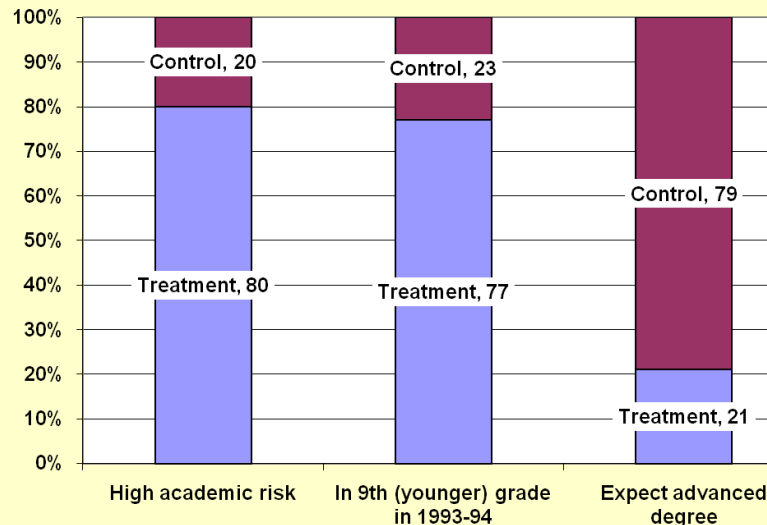
**SOURCE:** Data tabulated December 2007 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Program Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), U.S. Department of Education; study conducted 1992-93 to 2003-04.

**Exhibit 4. Percentage distributions in all 67 sampled projects (including project 69) between treatment and control groups on various attributes: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04**



Note with project 69 included there is not a balance or equivalence between treatment and control group as standards require in a random assignment study. Figure is read as follows: For example, among those who were classified as higher academic risk, 58 percent were in the treatment group and 42 percent in the control group. In a random assignment study distribution should be about 50-50 between treatment and control group; figure shows imbalance in overall sample with project 69 included.

**Exhibit 5. Percentage distributions for project 69 between treatment and control groups among those sample members who were a higher academic risk, in 9<sup>th</sup> (earlier) grade in 1993-94, and who expected an advanced degree at baseline: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04 (N= 85)**

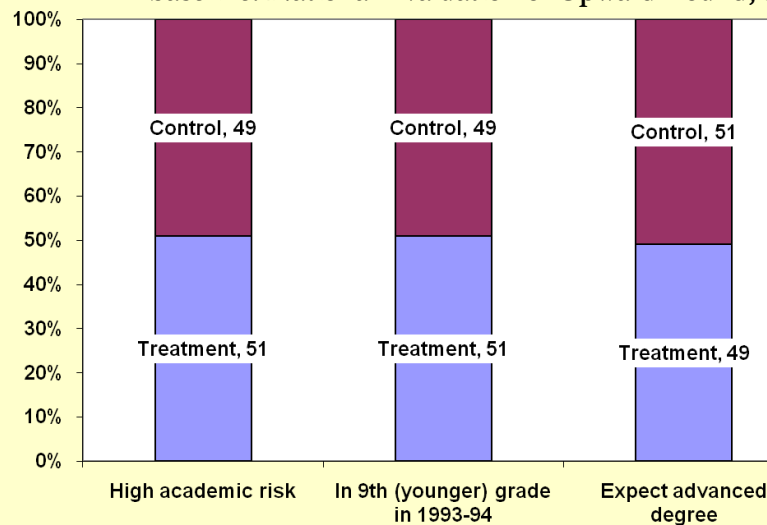


Note large imbalance in project 69 distribution. Figure is read as follows: For example, among those who were classified as higher academic risk, 80 percent were in the treatment group and 20 percent in the control group. In a random assignment study distribution should be 50-50 between treatment and control group; figure shows imbalance in project 69.

**NOTE:** High academic risk includes those sample members in the bottom 20 percent of the sample on 9<sup>th</sup> grade GPA and other academic indicators.

**SOURCE:** Data tabulated April 2009 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04.

**Exhibit 6. Percentage distributions for 66 of 67 sampled projects (excluding project 69) between treatment and control groups among those sample members who were a higher academic risk, in 9<sup>th</sup> (earlier) grade in 1993-94, and who expected an advanced degree at baseline: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04**



Note without project 69 there is a balance between treatment and control group as expected in random assignment study. Figure is read as follows: For example, among those who were classified as higher academic risk, 51 percent were in the treatment group and 49 percent in the control group. In a random assignment study distribution should be about 50-50 between treatment and control group.

**NOTE:** High academic risk includes those sample members in the bottom 20 percent of the sample on 9<sup>th</sup> grade GPA and other academic indicators.

**SOURCE:** Data tabulated April 2009 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04.

However the differences between the treatment and control group in project 69 occurred, Cahalan reports she found that the treatment sample from project 69 has a profile more like that of the Vocational/Career and Technical students of one of the project 69 grantee target high schools and the control group has a profile more like that of a Science and Engineering magnet component of another project 69 target high school. The control group in project 69 resembles more the Upward Bound Math Science (UBMS) population being in a higher grade at baseline and more academically proficient. It may be that project 69 was helping recruit for participation in a UBMS program known to be conducted by another nearby neighboring UB grantee. This would explain why project 69 had such a large number of applicants (completed baseline surveys) relative to its own openings which (given that the study was being weighted to the number of applicants not the number of openings) contributed to making the unequal weighting issues even more problematic at the second stage for project 69.

Sensitivity analyses revealed that even with inclusion of baseline controls for some of the non-equivalencies used in the models published in previous reports, results are sensitive to the inclusion or exclusion of project 69. None of the analyses in the Mathematica or the Cahalan reports use academic risk variables as controls, as they are from 9<sup>th</sup> grade transcripts that for some students occurred after they had begun participation in Upward Bound. The grade variable used by Mathematica in their analyses was that from the Student Selection Form that was grade at entrance into UB and not linked to a fixed time point and hence did not provide an adequate control for this non-equivalency.

#### **4-B-2 Correction Required to Meet Standards and Information Quality Guidelines in Area**

The uncontrollable or inadequately controlled for bias introduced by the extreme differences between the treatment and control group in project 69 on the characteristics most strongly related to outcomes leads to the same recommendation as already was presented above with regard to the sampling design and representational flaws observed. *The most transparent and straightforward correction is to present analyses only for that portion of the sample not represented by project 69--and acknowledge that the study cannot make robust estimates of impact for the project 69 stratum or for the entire population of interest to the study. This includes acknowledging that some statistical reviewers suspected that the randomization implementation may have been compromised in this case, or that the complicated multi-stage and sub-stratum interaction with the base weights and post-stratification adjustments given the extreme size of the weights created large non-equivalence in the treatment--control group for the project 69 sample.* Given the large weights for project 69 the normal protection from imbalances provided by the randomization process in studies of this size (about 2844-- about 1500 treatments and 1300 controls) appears to have been compromised. When project 69 is removed there is a balance between treatment and control for the remaining 66 projects when taken together. As the IES external reviewer notes, this balance allows for reasonably robust estimates for that portion of the population of interest not represented by project 69.

#### **4-C. The outcome measures are valid and provide the same standard for all sample members**

As the *What Works Clearinghouse Standards* note the estimates of outcome measures must be valid and must be based on common standards. This section focuses on the issues related to having a common standard of outcome measurement. It is best practice in estimating rates of college going and completion to make sure that the point of reference is the same (for example, Census reports “Percent Entering into Postsecondary by the *Fall after High School Graduation*”; NCES reports a “6 year graduation rate for institutions”). Standardizing by expected high school graduation date as noted by Dr. Chromy “is clearly necessary for proper interpretation of program effects.” If we do not have a common metric, we do not know whether observed differences or lack of differences in rates are related to the intervention or to differences in years of opportunity to complete.

##### **4-C-1 Reasons Report Does Not Meet Guidelines and Standards in the Area**

**Failure to standardize outcomes by expected high school graduation year (EHSGY) dates in outcome variable calculation.** The UB study applicant cohort was a multi-grade one. The “last grade completed” reported on the Baseline Survey (Question B1) ranged from grade 7 to a few in grade 11, with 57 percent being coded in 8<sup>th</sup> grade or below with reference to last completed grade on the baseline. During the Fifth Follow-up period of data collection 2003-04, the distribution of years since high school graduation date was as follows: approximately 10 percent of the sample had reached 6 years since EHSGY, 30 percent 7 years, 34 percent 8 years, 19 percent 9 years and 5 percent had reached 10 years. Adding to the complexity the Baseline Survey completion period spanned over an 18-month period. Examination of the distribution of expected high school graduation years (see Exhibits 7 to 9 below) taken from the baseline and other survey questions (not

used in the Mathematica analyses) shows that there is an imbalance between the treatment and control group in expected high school graduation year (EHSGY) with the control group more frequently having earlier completion dates (seemingly largely introduced by project 69).

These inadequately controlled for differences appear to have been biasing the results obtained without standardization especially when project 69 is included as is the case in all the Mathematica impact estimates. Mathematica analysts were apparently unaware of this imbalance. Working on the assumption that their random assignment design with a large sample would assure the treatment and control group were about equal in grade distribution, the Mathematica analysts for the study prepared impact estimates that included some controls for grade listed on the *Student Selection Form*,<sup>6</sup> but did not standardize outcomes to take into account differences in student grades at fixed points in time. The Student Selection Forms, completed by the project directors, were not well-linked to specific dates and the recruitment period went over two spring recruitments and entry summers. There is also some confusion as the same grade classifications included both those who were rising grade entrants and those already in the specific grade---(e.g. 8<sup>th</sup> grade rising 9<sup>th</sup> graders and those already in the 9<sup>th</sup> grade are coded as 9<sup>th</sup> graders). Therefore it is not a good source for estimating EHSGY or adequate to control for differences in EHSGY. Upward Bound Eligibility requirements are that students must be at least a rising 9<sup>th</sup> grader to go into Upward Bound, but recruitment and baseline completion for more than half of the sample took place when the 8<sup>th</sup> grade was the last grade completed.

As can be seen in Exhibit 7, reflecting the grade distribution at the time of the first follow-up, (1993-94) a larger proportion of the control group was in grade 10 than was true for the treatment group. As was presented in Exhibit 5, this difference in age distribution is pronounced in project 69. Among project 69's grade 9 sample (younger portion), 77 percent were in the treatment group and 23 percent were in the controls at baseline. In the overall sample, 56 percent of the grade 9 (younger portion) of the sample were in the treatment group and 44 percent in the control group (Exhibit 4).

To address the range of grades present in the sample and some observed treatment-control non-equivalencies in grades in favor of the control group it is necessary to standardize outcome measures to a fixed time point. Exhibit 8 shows the distribution of expected high school graduation year based on survey responses to relevant questions on the baseline survey, the first follow-up survey, and the third follow-up survey. *We also note that the students were in different grades at the time they completed the baseline survey, which means that some of the variables reflect different grade measures. For example, for the question on educational expectations, some students answered before they entered high school and others after they had completed one or two years of high school.*

It should also be noted that overall those who completed the baseline survey in earlier grades were less likely to be found on the Federal Student Aid (SFA) files. Cahalan examined the SFA files for the UB sample from 1994-95 to 2003-04. She reports that there is a 9 point spread between grade 8 and grade 9 sample members at baseline in the proportion being found on the SFA applicant file between 1994 and 2003 (66 percent for grade 8 and 75 percent for grade 9 reported on the baseline). This may be due to the older students having more years of opportunity, or it may also be due to the

---

<sup>6</sup> This form was completed by the UB project Director for each *Horizons* study applicant who completed the baseline survey and entered the sample.



fact that those applying for the program at later grades had already made it through the transition to high school, were closer to the event of interest, and hence more likely to enter postsecondary.

**In addition and importantly, the fact that the outcome measures were not standardized also confounds the variables used in the Mathematica models to control for differences between the treatment and control group.** For example, educational expectations at baseline is one of the controls used in the Mathematica models; however, two students who have the same postsecondary degree expectations from the baseline survey when surveyed again in the follow ups at fixed points in time may have had up to 5 years difference (most had 1 to 3 years) in the years of opportunity they have had to obtain the outcome degrees or credentials.

#### **4-C-2 Correction Required to Meet Standards and Information Quality Guidelines in Area**

In order to meet standards for a common metric to measure outcomes and impact, the outcome measures for the multi-grade cohort spanning 5 years of expected high school graduation years needs to be keyed to specific amounts of time from the expected high school graduation year (EHSGY). This requires two sets of data—the estimation of EHSGY for all sample members and also observing from the relevant sources of information (Third to Fifth Follow-up surveys, 10 years of *Federal Student Aid* (FSA) files, and for bachelor's degree the National Student Clearinghouse (NSC) data. These sources provide self-reports or administrative record evidence of the dates of first entrance into postsecondary and the dates of awards of degrees or certificates. Mathematica did not derive these variables and include them on their reports or the files delivered to ED. PPSS derived these variables and assisted by RTI International, have included them on the UB restricted use file. These derived variables can be used to correct the report to use a common metric for outcome measures.

Exhibit 7. 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 to 2003-04					
	Total sample		Project 69 sample		
First Follow-up grade in 1993-94	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	
Missing	3.6	1.7	9.8	--	
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. See table 3, data from the third follow-up question B1YY. SOURCE: Data tabulated May 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Program Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education: study conducted 1992-93 to 2003-04.					
Exhibit 8. Percentage distributions of expected high school graduation year (EHSGY) as tabulated from the baseline survey, the first follow-up survey, and the third follow-up survey: National Evaluation of Upward Bound (UB), study conducted 1992-93 to 2003-04					
EHSGY	Baseline survey question B1 (form references 1992-93 grade; some completed with 1991-92 grade reference; 100 percent response)	First follow-up survey question A1 (form references 1993-94 grade; did not allow for grade 8—97 percent response)	Third follow-up survey question B1YY (student report of year of high school entrance— 80 percent response)		
1994		4	5		
1995	10	22	19		
1996	33	42	34		
1997	44	29	30		
1998 and after	13	0	10		
NOTE: Note detail may not sum to 100 percent due to missing data. Expected High School Graduation Year (EHSGY) as tabulated based on the following questions from the various surveys Column 2: Baseline survey question B1: What grade were you in during the LAST SCHOOL YEAR (1992-93 school year)? (note some students reportedly answered for the 1991-92 school year) Column 3: First follow-up survey question A1: What grade (are you in/were you in during the 1993-94 school year) or (are/were) you not attending junior high or high school (now/then)? Column 4: Third follow-up question B1YY: What month and year did you first enter high school? SOURCE: Data tabulated May 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04.					
Exhibit 9. Percentage distribution of grade listed on the <i>Student Selection Form</i> , total sample and project 69: National Evaluation of Upward Bound (UB), study conducted 1992-93 to 2003-04					
	Total Sample			Project 69	
Student Selection Form Grade Reported	All	Treatment	Control	Treatment	Control
8	13	13	13	0	0
9	46	48	45	63	48
10	31	30	34	28	45
11	9	9	10	10	8
NOTE: Student Selection Form is not keyed to a specific academic year. Recruitment spanned over 2-academic years and distribution reflects the grade reported by projects as the grade the student was classified as when the forms were completed. A portion of those coded as grade 9 may have been “rising 9 <sup>th</sup> graders” entering UB in summer before high school. Others may have been already in the 9 <sup>th</sup> grade or in spring of the 9 <sup>th</sup> grade. Weights are the poststratified weight used analyses. SOURCE: Data tabulated May 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Program Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education: study conducted 1992-93 to 2003-04.					

- 4-D. The study must achieve adequate overall coverage and response rates and must concern itself with differential attrition, non-response and non-coverage bias issues of all the data sources used relative to the population of interest

#### 4-D-1 Reasons Report Does Not Meet Guidelines and Standards in the Area

**Study attrition and non-response bias**—Study attrition, especially differential attrition, is a concern in longitudinal studies. Both the What Works Clearinghouse (WWC) and the NCES Statistical Standards deal extensively with these issues. WWC discusses issues in terms of attrition and differential attrition and NCES Standards discusses the issues in terms of non-response bias and coverage issues. As the WWC Standards and Procedures Handbook notes:

Randomization, in principle, should result in similar groups, but attrition from these groups may create dissimilarities. Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC is concerned about overall attrition as well as differences in the rates of attrition for the intervention

NCES standards require a non-response or non-coverage bias analysis if any stage of the data collection the unit falls below 85 percent.

**STANDARD 2-2-4:** A nonresponse bias analysis is *required* at any stage of a data collection with a unit response rate less than 85 percent. If the item response rate is below 85 percent for any items used in a report, a nonresponse bias analysis is also *required* for each of those items (this does not include individual test items). The extent of the analysis must reflect the magnitude of the nonresponse (see Standard 4-4). **STANDARD 2-2-5:** In cases where prior experience suggests the potential for an overall unit response rate of less than 50 percent, the decision to proceed with data collection must be made in consultation with the Associate Commissioner, Chief Statistician, and Commissioner.

The UB evaluation survey response rates have been exceptionally high, although they declined with each round of data collection. They range from 99 percent on the baseline (required for entrance into the study “waiting list”), to 80 percent on the Third Follow-up to 74 percent on the Fifth Follow-up. Reports through the unpublished Fourth Follow-up have been based on only responders to the survey rounds with probability of selection weights adjusted for non-response. The Third and Fourth Follow-ups reported a response differential between the treatment and control group of about 4 to 5 percentage points higher for the treatment group. However, the Fifth Follow-up reported less of a differential. Despite high response rates, the unequal weighting issues with the study, make it especially vulnerable to variation in individual response creating instability of estimates especially for sub-groups. In project 69, for example, in one of the sub-stratum the treatment cases carried weights of 185 and the control group cases in the same stratum had weights of 95.

Mathematica does not include a non-response bias analysis in their reports for the survey data. Cahalan did a limited non-response bias study using the federal student aid (SFA) files and found 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). This is taken as an indication that those who respond to the surveys were also those who more frequently had positive postsecondary outcomes.

**Among the most serious attrition and coverage issue for the impact estimates concerns the improper use of National Student Clearinghouse (NSC) data for enrollment and less than 4-year awards.** These violate basic NCES and other statistical standards as to the extent of coverage and biased coverage. As defined in the NCES standards

***Coverage error** refers to the discrepancy between statistics calculated on the frame population and the same statistics calculated on the target population. **Undercoverage** errors occur when target population units are missed during frame construction, and **overcoverage** errors occur when units are duplicated or enumerated in error. **STANDARD 3-1-3:** .... If there is not evidence of a coverage rate of at least 85 percent of the target population, then frame enhancements such as frame supplementation or dual frame estimation must be incorporated into the survey study design.*

The *Mathematic Fifth-Follow-up report* at PPSS insistence for the first time in the Mathematica UB evaluation study uses administrative records from the Student Financial Aid files (SFA) and the National Student Clearinghouse (NSC) along with the Survey follow-up data in estimating outcomes (only follow-up surveys were used in the previous Third and Fourth Follow-up reports). PPSS reportedly encouraged Mathematica to obtain the NSC data; however, Cahalan reports that when she reviewed the NSC data and the coverage reported on the NSC website she found the coverage was too low to recommend using the information for postsecondary enrollment or degrees below the BA degree. NSC below BA degree coverage was almost non-existent for 2-year and less than 2-year certificates in the most applicable period.

As the NSC only began operations one or two years before the first students in this sample were graduating from high school (1994-95) and their website reports only having achieved about 25 percent coverage by 1996, NSC data can only be used with some confidence for estimates of bachelor's degrees earned which would have occurred later when coverage at least for 4-year institutions would have increased. In addition to poor coverage there is evidence of bias due to clustering of UB participants in grantee institutions who were not participating in NSC at the time (An estimated 30-40 percent of UB participants who attend postsecondary enroll in the grantee institution). Notably project 69 with 26 percent of the total study weight was not participating until after all of its sample had graduated high school (1996 and 1997). NSC did not begin coverage of degrees earned until well after it began coverage of enrollment so this data, would not have been reported for the period most applicable to 2-year and less than 2-year awards. Coverage by NSC for two-year and less than two-year enrollment and degrees or certificates remains problematic. It is reasonable to cautiously use NSC data only for the bachelor's degree outcome measure, as there is not evidence of bias and more time would have elapsed.

In addition Mathematica Fifth Follow-up Report is misleading in its discussion on page 26-28 of the extent of coverage of the NSC in the period of applicability implying that it was about 57 percent when the NSC web site notes it was about 25 percent in 1996. There is also no acknowledgment that the 2-year and less than 2-year institutions would have had much less enrollment coverage, or the fact that degree coverage of any type was not collected by NSC until several years after the most applicable period. Nor is it noted that project 69 did not begin enrollment coverage until after the period 1996 and 1997 when all of its sample members had graduated high school.

***There is evidence that this reckless use of NSC data by Mathematica with such poor coverage for estimates of postsecondary enrollment and also for 2-year and below degrees earned has had strong influence on the conclusions in estimates when the large weighted project 69 is included (as is the case with all the estimates reported by Mathematica). Project 69's large weights result in estimates that are unstable and are especially subject to differential response and coverage patterns.*** For example, Mathematica's own estimate of attainment of "any postsecondary degree" based on responders to the Fifth-Follow-Up Survey adjusted for non-response show a positive significant 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 TOT estimate of 15 percentage point difference---*Fifth-Follow-up Report* appendix tables C-7 and C14)<sup>7</sup>. If Mathematica had followed the procedures it used in reporting for the Third and Fourth Follow-up reports of using just follow-up survey data adjusted for non-response these positive results would have been the impacts they reported. However, in their widely quoted conclusions to the final *Fifth Follow-up* they report that UB had no impact on award of any credential but "certificates." Against PPSS's recommendation and that of the IES external reviewers to be "conservative in use of NSC" Mathematica chose to report in the text tables and conclusions only those estimates that use 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 a degree or certificate." ***The significant and large positive results note above, tabulated by Mathematica itself are included in their appendix C but never mentioned in the report of study conclusions which reports that there were "no detectable effects on the award of any degree or certificate" (see Appendix A).***

It's clear that using the NSC as the only source for the survey non-responders when uneven weights and differential coverage bias is probable due to clustering of UB grantees and the most applicable period was one in which NSC was not even collecting 2-year degree and less than 2-year certificate information is a violation of NCES and WWC coverage and attrition standards and seriously underestimates the extent of degree or certificate attainment below the bachelor's degree. **The**

---

<sup>7</sup> In this case the survey only based estimates with project 69 included also may overestimate the positive UB impact size due to large weighted certificate receipt responders to the survey. The survey only based estimates without project 69 also show significant positive impact but with not as large an effect size. As noted above it is probable that the Fifth Follow-up survey only based outcome estimates somewhat over-estimate the percent of the sample (treatment and control) who have obtained degrees even when adjusted for non-response. However, response rates for the treatment and control group were reportedly about the same (74 percent) so differential attrition may be less of an issue than in earlier rounds of the study.

choosing to highlight and include in the report text tables and Executive Summary conclusions estimates that clearly violate NCES standards for coverage, and may be biased while at the same time ignoring the Fifth Follow-up survey adjusted for non-response large positive results reported in Mathematica Fifth Follow-up appendix Tables C-7 and C-14 for award of “any degree” appears to be non-objective biased reporting that violates the ED Information Quality Guidelines and the basic proprietary standards for unbiased reporting of evaluation research of the Joint Program Evaluation Standards.

#### **4-D-2 Correction Required to Meet Standards and Information Quality Guidelines in Area**

As originally recommended by PPSS Technical Monitors, Mathematica should not use NSC data from this early period for enrollment estimates as NSC does not have enough coverage (25 percent) and there is evidence of bias due to clustering of UB participants in the grantee institutions. Therefore the best estimates for enrollment are ones that use all of the follow-up surveys (Third through Fifth) to maximized response and that use the 10 years of Federal Student Aid (SFA) files for evidence. Mathematica should also not use NSC data for 2-year degrees and below 2-year credentials. For award of degrees or certificates below the BA degrees which cannot be obtained from the federal aid files the best estimates are the combined surveys ---Third through Fifth--adjusted for non-response with cautions that they may overestimate the degree/certificate attainment. For BA receipt the best estimates are the three combined follow-up surveys (third through fifth) plus the NSC data as 4-year degree coverage had increased. Estimates with and without NSC can be used for BA degree and compared (see Exhibit 14).

#### **4-E. 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.**

One of the most difficult challenges of random assignment studies, especially of voluntary support service federal programs, concerns establishing and maintaining clearly distinguished treatment and control groups. *What Works Clearinghouse Standards* require that the intervention whose effects are being measured ***can be clearly attributed to the intervention*** and that the only difference between the treatment and control group is the intervention.

#### **4-E-1 Failure to Adequately Acknowledge Issues with Control Group Service Substitution and Treatment Group Dropping out**

This issue has been repeatedly raised by stakeholders concerning the Mathematica Upward Bound evaluation. It also formed the basis of the arguments made in Congress against the new UB evaluation study begun in late 2006, and cancelled by ED in early 2008 following Congressional cutting off of funding. It was argued that it would be unethical to purposively increase recruitment, and then to deny services to half of those recruited. If services were not denied and alternative

services were provided then the results might be confounded by control group substitution and treatment group dropping out.

The Mathematica Upward Bound study baseline and follow-up surveys contained questions (sometimes quite detailed) about other pre-college support or supplemental service participation, although these questions were somewhat different in each of the applicable survey rounds and suffer from the fact that the students were also in different grades at the time they completed the various survey rounds. They also suffer from the fact that the control group was not asked directly about any regular UB participation.<sup>8</sup> However, sufficient information was collected to classify whether the student reported any other pre-college support or supplemental services with an academic component and whether the study participant participated in UBMS.<sup>9</sup> This information summarized below can be used to gain some understanding of how much of an issue equivalent and/or similar service receipt was for this study.

Exhibit 10. 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 to 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%)	2,702 (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, UBMS, or other) pre-college support or supplemental service program	149 (10%)	1,690 (8%)	522 (40%)	8,651 (40%)	671 (24%)	10,342 (24%)
Reported participated in any type (UB, UBMS, or other) of pre-college support or supplemental service program	1,375 (90%)	20,176 (92%)	798 (61%)	13,215 (60%)	2,173 (76%)	33,390 (76%)
<b>NOTE:</b> Percents given in parentheses. UB = Upward Bound; UBMS = Upward Bound Math/Science. Weighted data use poststratified weights for longitudinal file. <b>SOURCE:</b> Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.						

<sup>8</sup> Control group members were given a list of specific and general programs that did not contain the name "regular UB program" and could indicate in an "other specify" space that they were in UB. They were asked about UBMS participation.

<sup>9</sup> Information was collected on the surveys about length of participation and type of programs on the various surveys that could be analyzed in more detail.

**Participation in the UB Program by the Treatment and Control Group.** About 26 percent of the Treatment Group maintained in the ITT analysis was coded as “waiting list” dropouts and about 20 percent of the treatment sample reported on the First Follow-up that they did not enter Upward Bound. A portion of this 20 percent could not remember being given the opportunity when asked about it a year later. Conversely about 12 to 14 percent of the control group reported they entered into Upward Bound Math Science or Upward Bound. *The Mathematica Fifth-Follow-up Report*, while emphasizing Intent to Treat (ITT), includes some Treatment on the Treated (TOT) analysis taking into account the 12-14 percent UBMS crossovers and the treatment non-UB participants (unlike the Third Follow-up report which did not recognize UBMS participation by the control group as a crossover). *However as one of the IES external reviewers noted given that one fourth to one-fifth of the treatment group did not enter UB and 12-14 percent of the control group was in UBMS, the TOT estimates may be more meaningful statistic for this study.* In the study year recruitment procedures were altered to ensure there would be double the number of applicants as openings. Those who completed the baseline surveys were considered on a “waiting list” for participation and in the study years no one could get on the UB “waiting list” without completing the baseline survey. All of the students were minors and over half were in middle school when completing the baseline survey; hence their actual entry into the UB program that next summer which was typically a residential program was related to parental permissions and family mobility. Low income families have high levels of mobility.

**Failure to Address Alternative Service Receipt Contamination Issues.** *Examination of study survey data revealed that 60 percent of the control group reported participating in some form of supplemental pre-college programs including other TRIO programs such as Talent Search before or after randomization (Exhibit 10).* Presumably most of these programs were less intensive than Upward Bound. Cahalan reports that PPSS requested that Mathematica use the information from the baseline and follow-up surveys on alternative service receipt to address issues of service substitution; however, Mathematica declined to conduct these analyses. Instead the *Mathematica Fifth Follow-up Report* continued to take the viewpoint also expressed in the *Third Follow-up Report* that the evaluation study was examining the impacts of Upward Bound over and above what supplemental services the students would be getting if the Upward Bound were not there. *The Upward Bound grantee practitioners repeatedly argued that this was a faulty logic model because the students getting other similar services were often placed in or sought the other time sensitive alternative services precisely because they did not get assigned to Upward Bound.* Moreover the students often applied to Upward Bound because of middle school services provided by TRIO programs such as Talent Search. This is the same issue addressed by noble laureate James Heckman, and

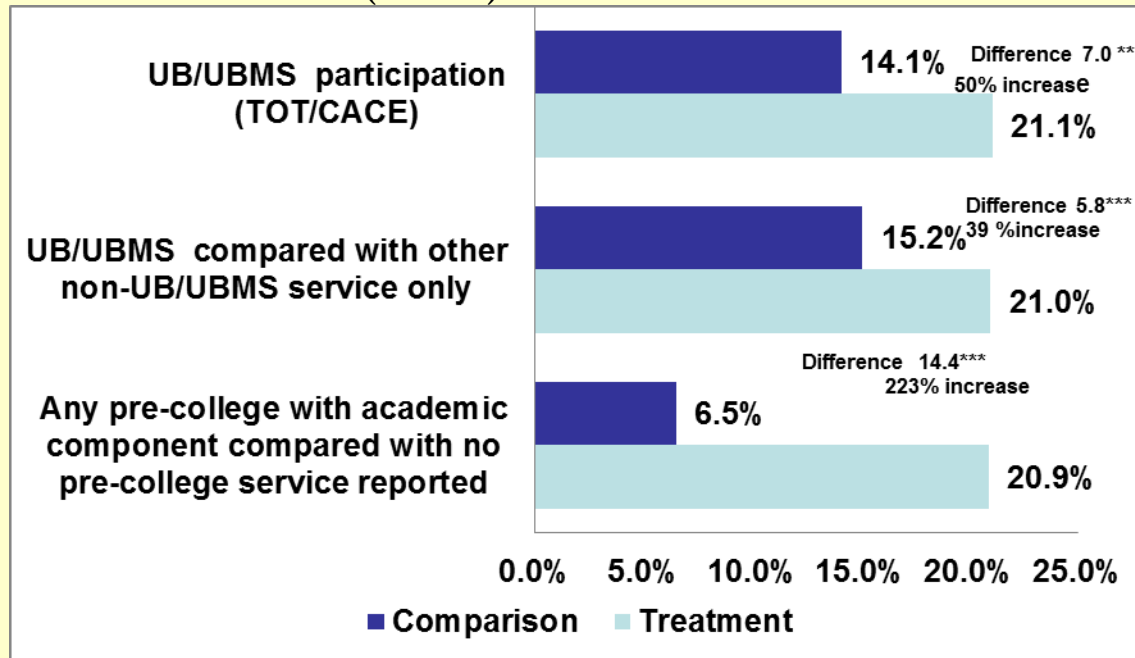


co-authors (Heckman Hohman, Smith, and Khoo 2000) re-analysis of the Job Training Partnership Act (JTPA) evaluation in which they considered the interpretation of evidence from social experiments when persons randomized out of a program being evaluated have good substitutes for it, and when persons randomized into a program do not enter the program or drop out. Using data from an experimental evaluation of JTPA classroom-training programs, they documented the empirical importance of control group substitution and treatment group dropping 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).

None of the Mathematica reports include comparisons of UB participants with those getting only other services or a serious consideration of the contamination issues related to the receipt of other similar but less intensive non-UB/non-UBMS services that were received by 60 percent of the control group. *Cabalan reports that concern with the serious contamination issues, led one internal PPSS reviewer of the Mathematica Fifth Follow-up Report, Dr. Jay Noell, the PPSS UB Evaluation Technical Monitor/COR prior to Dr. Cabalan, to recommend that Mathematica acknowledge that the random assignment study had failed because it had too high a level of contamination to be valid. He recommended that the data be analyzed as a quasi-experimental design. Dr. Noell recommended that the ITT analyses be placed in an appendix, but not be used in assessing program effectiveness.*

Cahalan, in her re-analysis, choose to present as much observational information about alternative supplemental services receipt for both the treatment and the control group as she could and to recognize that the ITT and TOT impacts she and Mathematica were estimating must be considered to be conservative estimates of UB program impact. The *Cabalan Re-Analysis Report* presents the ITT and TOT analyses using models and methods similar to Mathematica but also includes some additional observational quasi-experimental design analysis using instrumental variables regression. These analyses compare outcomes for those who were in UB/UBMS with those who reported they participated in some other non-UB/UBMS pre-college supplemental service (see *Cabalan Re-analysis Report* chapter 4). These analyses found positive impacts for UB/UBMS when compared to those participating only in some other form of non-UB/non UBMS pre-college services (Exhibit 11). Cahalan also compared those receiving any service (UB or other service) with those reporting no-service and found large effects for those reporting some form of pre-college supplemental service controlling for baseline differences--- although Cahalan points out the “any service vs. no service” comparisons, even using instrumental variable regression are more likely to suffer from unmeasured and uncontrolled selection bias issues (Exhibit 11).

**Exhibit 11. Instrumental variable Regression Results from the National Evaluation of Upward Bound for BA attainment in +6 years after Expected High School Graduation Year (EHSYG)**



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

NOTE: TOT = Treatment on the Treated (TOT); UB = Upward Bound; UBMS = Upward Bound Math Science. All estimates significant at the .01 level or higher. 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 and representational issues. We use a 2-stage instrumental variables regression procedure to control for selection effects. SOURCE: Data tabulated January 2010 using: National Evaluation of Upward Bound data files, study 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 1992-9 to -2003-04.

#### 4-E-2 Correction to Meet Standards and Information Quality Guidelines in Area

To adequately address the evaluation issues originally put forth for this study, a corrected report published by ED would need to include recognition of the serious contamination issues noted above that do not allow the study to meet WWC standards in this area and should at a minimum include quasi-experimental design analysis of the relative outcomes of those receiving various services and amounts of services such as those presented using instrumental variables regression in Exhibit 11.

**4-G The study design, data collection, analyses and reporting have integrity and the study is transparent, replicable and provides full documentation of procedures and pertinent limitations and sources of error**

**4-G-1 Reasons Report Does Not Meet Guidelines and Standards in the Area**

As Dr. Goodwin has noted in the email in *Exhibit E* summarizing his reasons for concern with the Mathematica report--- “*Virtually all the issues summarized above are not evident from reading the Mathematica report.*” The Mathematica report gives the appearance of being very procedurally correct and complete which may account for the fact that IES internal staff and external reviewers reportedly did not object to publishing the report. *This failure to fully acknowledge and indeed to mask some of the issues with the study in an effort to defend past design, analyses and reporting decisions is among the most serious of the problems with the Mathematica report.*

Specifically, the Mathematica Fifth Follow-up Report does not meet the integrity and transparency information quality guideline noted above in three major aspects:

- a) ***The Mathematica Fifth Follow-up Report contains no acknowledgment of credible alternative strong positive results and conclusions for the Upward Bound program*** that are found when study error issues are addressed in a straightforward and transparent manner with full documentation (see appendix P)
- b) ***The Mathematica Fifth Follow-up Report does not provide adequate documentation of statistical procedures*** used in arriving at the impact estimates that would allow the readers to assess their validity or correctness or to replicate the findings. In addition the report does not provide information on cell sizes for sub-group impact analyses and several of these comparisons do not have adequate cell sizes to support impact assessment
- c) **Failure to Adequately Acknowledge and Discuss Limitations and Provision of Misleading Information About Limitations.** ***The Mathematica Fifth Follow-up Report contains misleading statements*** concerning the limitations of the study. Specifically these concern the representativeness of project 69, the extent of treatment-control group non-equivalency and bias in favor of the control group, and the size of the impacts found when project 69 is removed and estimates are made only for the 74 percent of the population of interest not misrepresented by project 69 (66 of the 67 projects in the sample).

- d) ***Lack of ED Review Process Integrity.*** The Mathematica *Fifth Follow Up Report* was published in a manner that did not follow PPSS standard review procedures and it also may be that there was improper political influence on IES and PPSS in the review process

***4-G-1-a No Acknowledgment of Credible Alternative Positive Results and Conclusions.***

The report contains no acknowledgment of what external reviewers and Dr. Goodwin have called the “credible alternative analyses” conducted by the PPSS staff responsible for technical monitoring of the study. These analyses identified, mitigated, and corrected for recognized study limitations and found quite different conclusions concerning the UB program effectiveness.

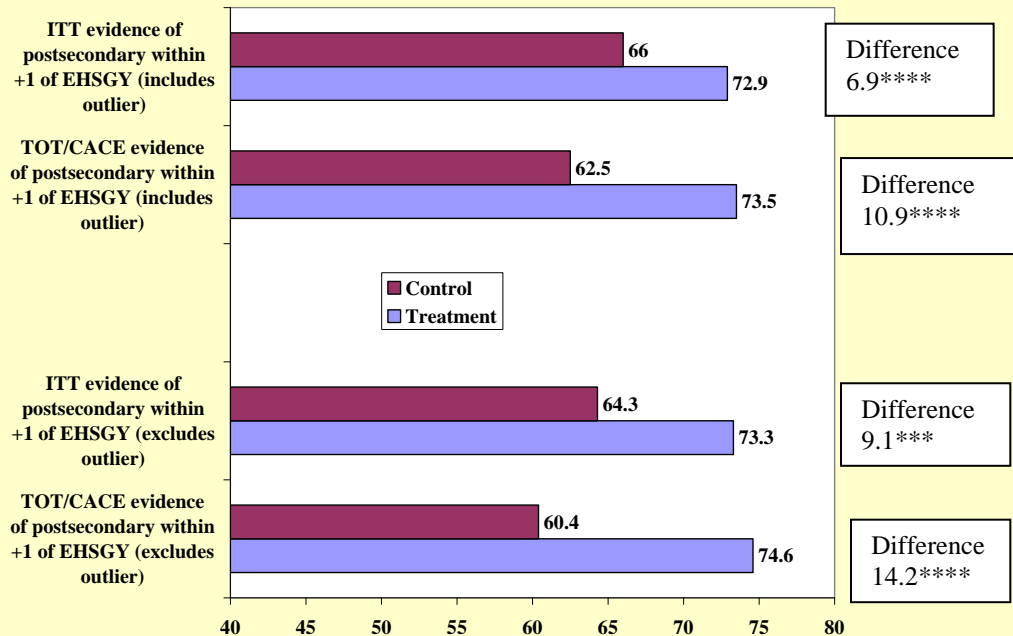
***Failure to Acknowledge Strong Results on Enrollment and Federal Aid Application and Receipt With and Without Project 69.***

*It is the position of this request for correction, in agreement with the IES reviewer and PPSS Technical Monitoring recommendations, that the estimates with project 69 are flawed and should not be used as a measure of program impact by ED.* However, the fact also needs to be noted that when Cahalan standardized outcome measures and used all three survey follow-up rounds to maximize response and used the student federal aid (SFA) files and avoided using the low coverage and bias introducing NSC data that she found positive results on key outcomes both with and without project 69. While impacts are consistently larger for the 74 percent of UB not represented by project 69, as can be seen in Exhibits 12 and 13 there is also clear evidence that even with the bias introducing project there are significant positive impacts for UB when outcome measures are standardized to expected high school graduation year and when the NSC data with low and biased coverage is not included. Documentation of the variables included in the models and sample output from the statistical procedure is provided in Appendix P. Cahalan’s re-analyses also showed positive impacts on application and award of federal student aid and for the attainment of any postsecondary degree or credential (see *Cahalan Re-Analysis Report* Executive Summary results excerpted in Appendix C).

Exhibit 13 taken from the *Cahalan Re-Analysis Report* shows the large impact on postsecondary entrance for those classified as “academically at risk” and more modest but also significant positive impacts for those in the top 80 percent on academic indicators. Previous drafts of the *Mathematica Fifth Follow-up Report* also contained appendix tables that showed positive impacts when the top 80 percent and the bottom 20 percent sub-groups are considered separately, but not for the overall impact. In this case, the result is related to the fact that project 69 is contributing a disproportionate proportion in the bottom 20 percent on academic indicators from its treatment group. The two academic risk divided sub-groups thus each have a more balanced treatment-control group equivalency than does the overall sample with project 69

included. This is because when the sub-groups are formed the heavily weighted project 69 treatment group is more likely than the project 69 control group to be in the bottom 20 percent while the higher performing control group from project 69 cases are more often in the top 80 percent—so paradoxically some of the imbalance between the treatment and control group on academic risk bias in the overall sample is removed by this sub-grouping.

**Exhibit 12. Estimated rates of postsecondary entrance within +1 (about 18 months) of expected high school graduation year (EHSGY) for Upward Bound Opportunity (ITT) and UB/UBMS participation (TOT), with and without outlier: study conducted 1992-93 to 2003-04**



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

**NOTE:** UB = regular Upward Bound; UBMS = Upward Bound Math/Science; ITT = intent to treat; TOT = treatment on treated; CACE = complier average causal effect. Model based estimates based on STATA logistic and instrumental variables regression taking into account the complex sample design. Weighted estimates use poststratified weights. See table 5 in body of the report for detailed note. **SOURCE:** Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Program Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), US Department of Education: study conducted 1992-93 to 2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04. (Excerpted from the *Cabalan Re-Analysis Report*, Figure IV)

Exhibit 13.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 <i>Horizons</i> 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 <sup>th</sup> 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 <sup>th</sup> 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.</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 (; 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: Students with higher risk were in the bottom 20 percent of academic achievement in 9<sup>th</sup> grade; Students with lower risk were in the top 80 percent of academic achievement in 9<sup>th</sup> grade based on student transcript information. See table 5 for detailed general notes. Appendix tables give examples of actual model results.</p> <p><b>SOURCE:</b> Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; federal Student Financial Aid (SFA) files 1994-95 to 2003-04; and National Student Clearinghouse Data 1995-2004. (Excerpted from the <i>Cabalun Re-Analysis Report</i>, Table 12)</p>				

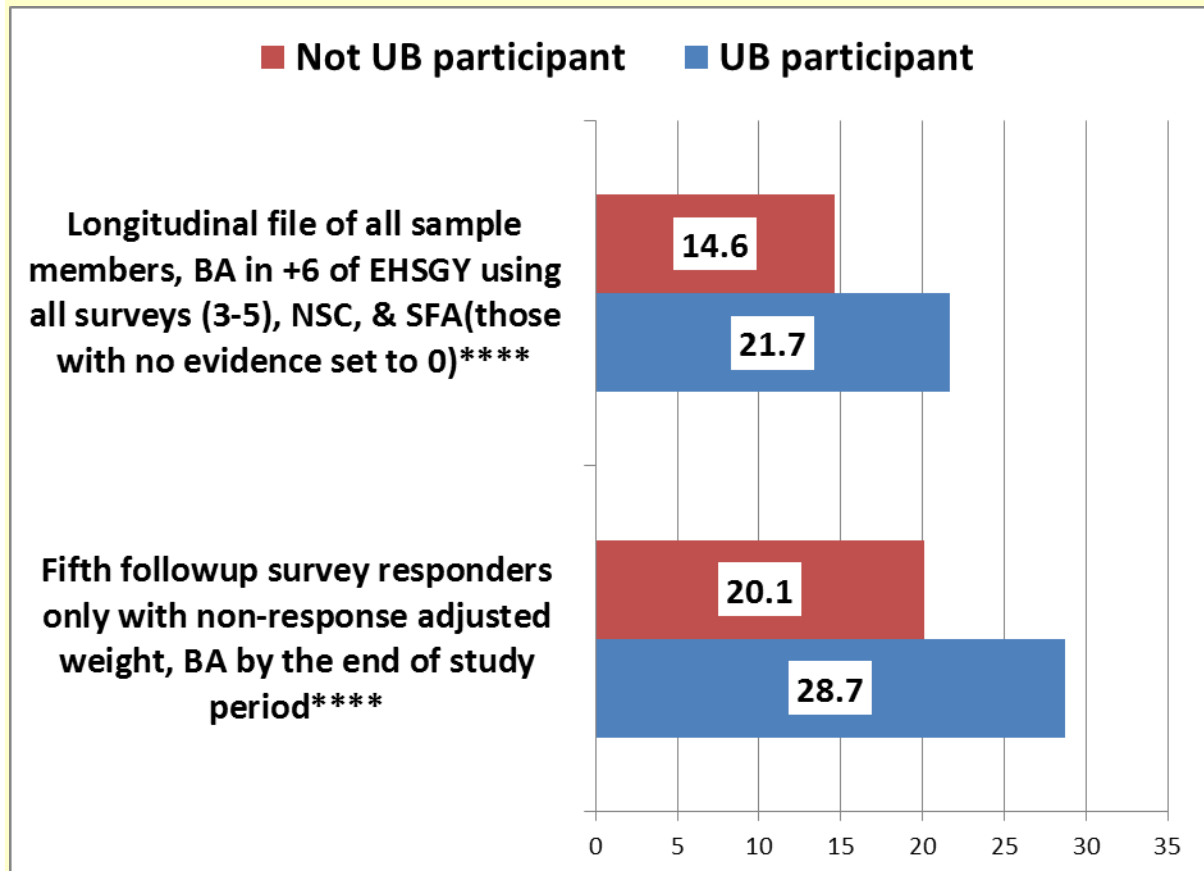
**Failure to Acknowledge Strong Positive Results on BA Attainment When Bias Introducing Project 69 is removed and estimates are presented only for the strata not represented by Project 69 (for 66 of the 67 projects in the study).** The *Mathematica Fifth Follow up Report* misleadingly indicates that the results without project 69 are not strong enough to change the major conclusions. Using very similar BA attainment results as reported by Cahalan in Table 10 of her *Re-Analysis Report*, they also maintain that standardization does not make a difference. *These misleading statements in the Mathematica report led at least one of the external IES reviewers to fail to recognize the serious significance of standardization issues or the problems with project 69—they repeated the Mathematica false assertion that the sample design project 69 issues and standardization of outcomes were very legitimate concerns but that they did not make a substantial difference in the conclusions.*

As noted above and can be seen from Exhibit 14, unlike the results for postsecondary entrance and financial aid awards, Cahalan also did not find positive results for BA receipt when she standardized outcomes to EHSYG in the BA estimates that included project 69. The academic risk (and other academic indicators) factors in favor of the control group introduced by project 69 and the 4-year-2-year representational issues for the heavily weighted project 69 are too strong a source of baseline non-equivalency bias in favor of the control group. ***However, what Mathematica did not report and masks in the Fifth Follow-up Report are the very large significant impacts on BA receipt found for the 66 projects that taken together meet the What Works Clearinghouse standard that the treatment and control group be equivalent on factors related to outcomes.*** Sampled cases from these 66 projects also did not suffer from the representation issues outlined above related to the grantee being a 2-year and certificate granting institution that was supposed to be the sole representative of a diverse 4-year stratum with extreme and unequal weights. **As can be seen in Exhibits 14 and 15 below there are large UB impacts on BA receipt for 66 of the 67 projects in the sample.**

Exhibit 14. Evidence of attainment of any postsecondary degree or certificate and attainment of a BA within +6 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 study sampling strata		One outlier project removed (remainder represents 74 percent of <i>Horizons</i> waiting list)	
	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)	Given Opportunity (ITT)	Participated in UB/UBMS (TOT/CACE)
<b>Evidence of attainment of any postsecondary degree or credential by end of study period—fourth follow-up survey data only with non-response adjustment</b>				
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**)
<b>Evidence of attainment of any postsecondary degree or credential by end of study period-- Fifth follow-up survey data only with non-response adjustment</b>				
Fifth follow-up survey responders only	pr-T = 51.9 pr-C = 41.4. Difference = 10.6**  ( pr-T = 49.0 pr-C = . 44.6 Difference = 4.4**)	xb T = 54.4 xb C = 39.3 Difference = 15.2**  (xb T = 50.4 xb C = 43.5 Difference = 6.9**)	pr T = 47.5 pr C = 42.6. Difference = 4.9**  (pr T = 48.6 pr C = 44.6 Difference = 3.9**)	xb T = 49.3 xb C = 41.5 Difference = 7.8**  (xb T = 50.0 xb C = 43.6 Difference = 6.4*)
<b>Evidence of attainment of any postsecondary degree or credential by end of study period-- Fifth follow-up survey and NSC data</b>				
Fifth follow-up survey and NSC data used	pr-T = 35.0 pr-C = 30.8 Difference = 4.6***  (pr T = 36.3 pr C = 33.4 Difference = 2.9**)	xb T = 37.4 xb C = 30.7 Difference = 6.7****  (xb T = 38.3 xb C = 33.2 Difference = 5.1**)	pr T = 34.4 pr C = 30.7 Difference = 3.7***  (pr T = 36.2 pr C = 33.3 Difference = 3.0**)	xb T = 37.0 xb C = 30.9 Difference = 6.1***  (xb T = 38.3 xb C = 33.1 Difference = 5.2**)
<b>Evidence of BA in +6 of EHSGY--- All applicable follow-up surveys, Pell Award Files, NSC—responders and non-responders included – longitudinal file</b>				
Uses all applicable follow-up surveys, NSC, and Pell graduation variable; standardized to EHSGY; 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***)
<b>Evidence of BA in +8 of EHSGY--- All applicable follow-up surveys, Pell Award Files, NSC—responders and non-responders included – longitudinal file</b>				
Uses all applicable follow-up surveys, NSC, and Pell graduation variable; standardized to EHSGY using first followup variable; longitudinal file poststratified weight.	pr-T = 16.6 pr-C = 16.3 Difference = .3 NS  (pr T 18. 9 pr C = 16.6 Difference = 2.3**)	xb T = 19.1. xb C = 18.0 Difference = 1.1 NS  (xb T = 22.0 xb C = 18.2 Difference = 3.8**)	pr T = 17.5 pr C = 13.7 Difference = 3.8****  (pr T = 18.9 pr C = 16.1 Difference = 2.8***)	Xb T = 21.7 xb C = 14.6 Difference = 7.1****  (xb T = 22.3 xb C = 16.6 Difference = 5.7***)
<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; 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><b>NOTE:</b> 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><b>SOURCE:</b> Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to -2003-04; federal Student Financial Aid (SFA) files 1994-95 to 2003-04; and National Student Clearinghouse Data 1995-2004. (Excerpted from the <i>Cabalan Re-Analysis Report</i>, Table 10)</p>				



**Exhibit 15. Impact of Upward Bound (UB) on Bachelor's (BA) degree attainment: Instrumental Variables Regression models for Treatment on the Treated (TOT) 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/.01/0.00 level; NS = not significant at the .10 level or below. NOTE: UB = regular Upward Bound; UBMS = Upward Bound Math/Science; 'TOT' = treatment on treated; CACE = complier average causal effect. . EHSYG = Expected High School Graduation Year; NSC = National Student Clearinghouse; SFA = Student Financial Aid All estimates significant at the .01 level or higher. 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 and representational issues. Model based estimates based on STATA logistic and instrumental variables regression taking into account the complex sample design. Weighted estimates use poststratified longitudinal weights or non-response adjusted weights as noted. We use a 2-stage instrumental variables regression procedure to control for selection effects for the Treatment on the Treated (TOT) impact estimates. SOURCE: Calculated January 2010 using: National Evaluation of Upward Bound data files, study 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 1992-9 to -2003-04 (Parts Excerpted from the *Cabalan Statement See Appendix D*)

These very strong results for UB for BA receipt that Cahalan found for 66 of the 67 projects taken together are among those that are most troubling from their absence in the Mathematica reports—especially given the historical and continuing emphasis on 4-year college completion of the Upward Bound program . For example, with the complete longitudinal file using three rounds of survey data, financial aid records data, and NSC data, the Treatment on the Treated (TOT) instrumental variables regression estimates of impact show a large increase in BA receipt by +6 years after expected high school graduation--- *going from estimated 14.6 percent for the control group to 21.7 percent for the treatment group—an increase of 49 percent (Exhibits 14 and 15)*. Similar impacts were found using unstandardized variables derived by Mathematica for data by the end of the study period from the *Fifth Follow up Survey* responders adjusted for non-response (Exhibit 15-- second comparison in chart).

Statistically significant and substantial impacts were also found for the Intent to Treat (ITT) estimates with an increase in BA attainment in +6 years of about 28 percent. The combination of having a 2-year and certificate awarding grantee to represent the largest 4-year stratum, with a treatment group on average more appropriate for the grantee institution and a control group from this project that on average resembled UBMS participants appears to have led the Mathematica analysts into a Type I Error of overestimating the impact of UB on certificate attainment and a Type II Error of failure to detect the substantial impacts of UB on BA receipt which is observed among the 66 other projects which when taken together have a balance between the treatment and control group on baseline attributes related to outcomes.

#### **4-G-1-b Lack of Adequate Documentation and Inability to Check or Replicate Results**

While the Mathematic Fifth Follow-up report contains nine technical appendices (A to I) and includes literally over 1000 comparisons and over 20 ways of computing each impact estimate along with standard error information, in the final assessment the report does not provide enough information to check or replicate the major conclusions or the numerous comparisons made in the tables. *All of the impact tables include the bias introducing project 69, with the exception of some experimentation with differential weighting in Appendix G*. Most of the 27 estimates in the tables and importantly those selected to be discussed in the body of the report and executive summary conclusions include National Student Clearinghouse (NSC) data when coverage was found by PPSS staff to be too low with evidence of bias due to clustering to be included. The Mathematica impacts are stated to include an interaction term for project 69; however, the estimates reported differ from those that PPSS staff found when including an interaction terms for project 69 (see Appendix E-in the *Cahalan Re-Analysis Report*). The Mathematica report does not provide adequate and transparent discussion of the technical issues of major concern or enough information to replicate results.

It's also clear that some of the overall and sub-group results are suffer from the unequal weighting issues that make the impact estimates sensitive to the fact that different sample members responded to the different rounds of the surveys with gradual decline in response over time. In some cases the sub-group divisions change the treatment-control non-equivalency introduced by project 69 into the overall sample. For example, the Mathematica race/ethnicity sub-group analysis found positive impacts for whites but not for blacks in the sample. As none of the project 69 sample members (treatment or control) are white, (being 60 percent Black and 40 percent Hispanic) when we look only at treatment control differences in the overall white sample we have, in effect, removed some of the bias introduced by the fact that project 69's control group has on average higher education expectations and higher academics than project 69's treatment group. Conversely when we only look at the Black sample sub-group, project 69 is contributing an even higher proportion of the weight and (treatment-control group non-equivalency) than it does to the overall sample. As noted a similar sub-group issue happens with regard to academic risk categorization sub-groups (top 80 percent and bottom 20 percent). There are also clear cell size issues with some of the sub-group analyses that are not pointed out.

**4-G-1- C. Failure to Adequately Acknowledge and Discuss Limitations and Provision of Misleading Information About Limitations.** Cahalan reports that because of the objections and comments PPSS staff gave to Mathematica that the final version of the *Fifth Follow-up report* does acknowledge some of the issues with the study sample design that were previously unreported in the 1999 and 2004 reports. For example, the *Mathematica Fifth Follow-up Report* published in 2009 ---10 years after the first report on high school impacts (1999) was published, is the first of the published reports to include information on the fact that one project was carrying 26 percent of the weight. *The report, however, does not acknowledge and indeed seemingly deliberately masks other key issues such as the representational issues with project 69, the extent of treatment-control group baseline non-equivalency with an uncontrollable bias in favor of the control group, the control group alternative service contamination issues, and the extent of lack of coverage of the NSC data and potential for biased coverage.*

**Evaluation Ethical Issues Related to Non-Disclosure.** As is indicated from the tables sent to PPSS concerning project 69 in 2005 by Mathematica (see appendix H), at least since 2005, Mathematica has been aware that the results presented in their reports were subject to the inclusion or exclusion of project 69. Cahalan reports that since 2006, Mathematica has also been made aware that when Federal Aid Files data was used that the unstandardized estimates for the Fourth Follow-up for the entire sample with and without project 69 were significant. Since 2008 when Cahalan standardized results to EHSGY, Mathematica has also been made aware that when results were standardized by expected high school graduation year that there were statistically significant results for postsecondary entrance with and without project 69 for estimates of enrollment in +1 and +4 years and for application for federal aid in +1 and +4 years using all of the applicable rounds of follow-up surveys and the federal aid files.

**As noted PPSS staff report they were not made aware of the seriousness of the unequal weighting issues until 2005.** As noted previously at a conference call meeting held in 2005 the former Mathematica staff person responsible for analysis and drafting the *Fourth Follow-up Report* drew this fact to ED's attention. Dr. Stuart presented estimates that showed that her results were sensitive to one project that had very negative individual results. At this time PPSS staff did not have copies of the UB data files and it was not until much later that **PPSS staff came to understand that project 69's negative results were not because of the poor- performance of the project but due to the extreme difference between the treatment and control group in favor of the control group on factors related to outcomes in that project (see Exhibits 4 to 6).** Unfortunately as noted these differences introduced uncontrolled for bias into all of the estimates reported in the Mathematica reports but are not acknowledged in the *Mathematica Fifth Follow-up Report*.

Although reportedly asked to do so by PPSS staff, it's not clear whether Mathematica staff ever did any qualitative or quantitative analysis in 2005 to find out why project 69 efforts had such (seemingly to Mathematica) negative results. The site was reportedly visited by Mathematica staff in the early days of the study and it was informally reported that project 69 was involved in training construction workers as a follow-up to the CTE target school it served. It's not known if the site visitor staff involved remembered or ever knew that the stratum project 69 was supposedly representing was a 4-year one and not a two year/certificate granting institution. *The evidence from the strong impacts on academically at risk students, for which project 69 contributed 30 percent of the treatment weights suggests that the project was actually doing a good job of serving its target population—it just was not an appropriate representative of its 4-year stratum and for whatever reason it did not have an equivalent treatment and control group and therefore could not be used to assess the effectiveness of the grantee UB project.*

*It is a clear violation of Program Evaluation Propriety Standards not to fully reveal to stakeholders these issues, especially when the stakeholders are being given “ineffectual ratings” on the basis of the faulty results and have had zero funding recommendations justified because of the results. These results continue to be used in Congress in 2011 to justify funding cuts to the program.* Even if project 69 were not so problematic and if it really was a site that had very poor results (which was clearly not demonstrated); it would still violate ethical standards not to reveal that for 66 of the 67 sites taken together that demonstrated the equivalency between the treatment and control group (required by the What Works Clearinghouse) there are strong positive impacts.

**Clear inability of sample design to make inferences to population.** As IES Statistical Reviewer C pointed out, it was not ever logically possible for only one project to represent this diverse 4-year stratum which included projects with majority white and majority black students and included the major research universities that had UB projects as well as historically black 4-year liberal arts colleges. Research Universities have little in common with project 69 which resembles a commuter branch community college campus with a focus on work force training. As noted 4-year stratum was defined as non-Hispanic and includes UB projects that were

majority black and those that were majority white. However, none of the project 69 participants (treatment or control group) were white—60 percent were black and 40 percent Hispanic.

***4-G-1-d. Lack of ED Review Process Integrity.—Political Interference and IES Compliance***

Examination of the facts (see appendix G) concerning the review process for the report raises serious questions about the integrity and procedural capacity of a review processes that led to publication of a report that clearly did not meet ED’s own information quality guidelines. In this case a contractor’s conclusions were accepted and published despite serious concerns and clear evidence of “credible analyses” to the contrary prepared by the person assigned by ED to technically monitor the study. Mathematica repeatedly refused to make the changes requested by Dr. Cahalan and by the end of the report review period also by her supervisor Dr. Goodwin which (in their best technical judgment) would have made the report meet ethical and technical standards for evaluation research (see appendices G to O).

As the PPSS Team Leader and COR assigned to monitor the Technical Quality of the report Dr. Cahalan had been seriously questioning Mathematica procedures with regard to the study since 2006. She consulted with external statistical experts who held the PPSS Statistical Technical Assistance contract at RTI and reviewed the data files from the study herself, and briefed staff in IES, Budget Service and OPE staff on what she had found. The review by RTI statistical staff under Dr. Chromy in 2006-07 included replication of Cahalan’s positive impact findings using data from the 4<sup>th</sup> follow-up. However, Cahalan was overridden repeatedly by political appointees in performing her job as Technical Monitor, seemingly with the compliance of IES. She objected in writing to the decision to put the report into the final Ex Sec review process in November of 2008 and again to the decision to publish the report in January of 2009 (see appendices L to N). Dr. Cahalan’s supervisor, Dr. Goodwin—head of the unit—Policy Analysis Services—PAS within PPSS had strongly defended Mathematica in internal ED publication debates of earlier results in 2004 for this controversial study. However, by 2008 he also expressed serious doubt in writing about the conclusions both to Mathematica and to his superiors (see Appendix K.). He publically repeatedly stated that he believed Cahalan’s estimates were “more credible” or “at a minimum equally credible.” The one IES external reviewer knowledgeable about statistical sampling had also stated that the inferential results for the entire sample including project 69 used by Mathematica to base conclusions were not “robust.” The Office of Postsecondary Education (OPE) out of whose allocations the funding for the study was derived also “formally disapproved” of publishing the report in the Ex. Sec. Review of December of 2008 and submitted written comments none of which were ever addressed in any manner. Yet ED, reportedly with IES concurrence, published the report in the last week of the Bush Administration in 2009 and has not seen fit to remove it.

Cahalan reports that in late December 2008 after the report was officially classified in the Ex Sec review process as “returned to PPSS for rewrite” due to the OPE disapproval, she was forbidden by the OPEPD front office political appointee staff to engage in the usual procedure of communicating with Mathematica concerning the comments from the Ex Sec review. In December of 2008 she sent a draft of a memo (that had already been reviewed by her supervisor

Dr. Goodwin) to the front office asking for permission to send the memo. The draft memo was addressed to the new Mathematica President Decker and detailed the changes needed to the report. Dr. Cahalan was denied permission to send the change memo to Mathematica and told not to communicate with Mathematica except through the front office (see appendix M). Two weeks later the report was published without Cahalan's or OPE's issues being addressed. The OPEPD front office political appointee staff was reportedly engaged in a private negotiation with Mathematica (and reportedly IES) about the report without PPSS staff's involvement.

In early January 2009, Cahalan and Goodwin were surprised when Goodwin was ordered by Ze've Wurman (special assistant to Assistant Secretary Evers and the OPEPD front office liaison with PPSS for report review) to publish the report and post the report to the PPSS website before the end of the Bush Administration—by January 15. PPSS was told that the front office had negotiated with Mathematica and that they had made changes that satisfied IES. PPSS was not given any information concerning these negotiations or what these changes were or why they were considered sufficient. Upon review of the revised draft, Cahalan and Goodwin found that their basic issues with the report had not been addressed and moreover that several misleading paragraphs had been inserted into the report.

With regard to IES the question can then legitimately be raised as to how a report that clearly violated so many of the IES/NCES and IES/What Works Clearinghouse standards and evaluation research ethical standards concerning stakeholder rights received IES approval to publish? The answer to this question is unknown to COE. Below is what is known and some speculation provided by Dr. Cahalan and others involved.

The IES career staff person who was assigned to review the Mathematica report when it was submitted (over Cahalan's objection) into Ex Sec final review in December of 2008, has stated to Dr. Goodwin that soon after she began to prepare her written review of the *Mathematica Fifth Follow-up Report*, she was directed by IES front office staff to stop preparing comments and not to prepare a written review. She was told that a decision had been made by IES leadership that IES internal staff would not submit comments in the Ex Sec review reportedly because there had already been the external IES review conducted earlier. Cahalan reports that this behavior of not having internal staff from the IES/National Center for Education Evaluation (NCEE) provide extensive comments with regard to PPSS reports was reportedly very atypical for IES. There is a history of IES giving extensive and critical comments on PPSS reports such that often reports are held up for considerable time and several rounds of comments and revision happen. For example some reports such as the sister study Student Support Services (SSS) final evaluation report (prepared by Westat) was held up for more than two years and had at least three rounds of Ex. Sec. review process submissions and revisions. Similarly the Middle School GEAR UP evaluation report also prepared by Westat was held up for over a 2 years by the need to address IES extensive comments. Both the SSS and Middle School GEAR UP reports had to have extensive re-write's before IES would sign off on the publishing of the reports.

**Why was the *Mathematica Upward Bound Fifth Follow-Up* final report treated differently by ED and IES than other Reports?** Cahalan reports that in her view part of the answer is that the study employed what was considered to be the strongest type of research

design and the favored method of IES/NCEE --random assignment-- and the draft Mathematica report gave the appearance of being procedurally correct and very thorough. Also as noted the report masks some of the most serious issues and IES external reviewers were not given enough information to know whether the report was correct or not. Cahalan reports that another factor was that Mathematica leadership openly accused Cahalan of being an advocate for the program and overstepping in interfering with Mathematica's "objectivity." The Department of Education did not want to appear as not being objective in evaluating one of its own programs. In addition, the Department of Education Budget Service and IES staff evaluators had recently lost battles in Congress over a new IES Upward Bound Random Assignment Evaluation begun in 2006 and for which Congress cut off funding in 2007 (ED formally cancelled this study in 2008), so the climate at ED with regard to TRIO in IES and Budget Service was not favorable. Cahalan also reports that ED Budget service staff commented to her in one of her briefings that publishing her findings would make it even harder to get funding for future evaluations from Congress.

Another factor is the fact that Mathematica is the What Works Clearinghouse (WWC) contractor and also had been awarded a major portion of IES/NCEE contract funding over the past ten years. There was also the presence of high level ex-Mathematica staff at IES. Mathematica has a reputation for highest quality work and Cahalan reports that she perceived there was a reluctance to even entertain that Mathematica could have made mistakes, especially as they held steadfastly to their positions with memo's defending their analyses. Thus there was a convergence of: 1) political expediency for the Bush Administration political appointees and Budget service who had justified their zero funding budget requests and TRIO policy reform efforts on the basis of Mathematica's earlier UB evaluation reports--- and 2) IES's strong respect for and desire to defend Mathematica's work and also to defend the IES favored random assignment method that the study represented. This convergence apparently led IES leadership and staff to not take very seriously the technical concerns or evidence presented by PPSS Monitoring Staff and reportedly resulted in the IES concurrence leading to publishing the report.

Another factor noted by Cahalan is that IES/NCEE staff and external reviewers had expertise and faith in the random assignment method to deal with the study error issues raised by Cahalan and, with one exception, not as much expertise or understanding of survey sampling and non-sampling error issues that were impacting the study results obtained by Mathematica. There is also not as much appreciation for the need for stakeholder involvement in evaluations or concern with the stakeholder rights such as those addressed in the *Joint Committee Education Evaluation Standards* for evaluations, especially with regard to high stakes evaluations that effect a programs existence and funding. Stakeholders and anyone who expressed concern for their rights were regarded as interfering with the "objectivity" represented by the "expert independent outside contractors."

#### **4-G-2 Correction Required to Meet Standards and Information Quality Guidelines in Area**

**Correction Needed With Regard to the Report.** It is incumbent on ED as the sponsor of the study to fully reveal the information standards violations noted in this request for correction

to the public and UB stakeholders. Integrity is listed as one of the data quality factors in the Department of Education Quality of Information Guidelines. Note is made that certain information used for policy decision making should be subject to special concerns as to robustness. One of the factors stressed in Obama administration OMB documents concerns transparency within the government and fostering data driven decision making. As noted --- The Joint Committee Program Evaluation Standards (see Exhibit B for a summary) also address Integrity under the heading of Propriety. Standard P6 noted below discusses the full disclosure of findings.

***P6 Disclosure of Findings*** *The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results(Standard P6 in Joint Committee on Education Evaluation Standards).*

**Corrections Needed With Regard to the ED Publication Review Process and Evaluation Ethics With Regard to High Stakes Evaluations.** Given the evidence presented throughout this request for correction it is imperative that ED act responsibly and remove the Mathematica report from its website until such time as it is adequately corrected to meet Information Quality Standards. A public errata statement concerning the Mathematica UB reports should be issued and an apology given to the stakeholders who have been misled by this evaluation for more than 10 years.

Given the irregularities of the review process noted above there is a need for reform of the ED publication processes to avoid political interference. Given the apparent failure of IES (leadership, and internal and external reviewer staff) to understand the seriousness of the research ethics and technical issues involved with the high stakes UB evaluation there is also a need for reform at IES. There is a strong need for IES/NCEE staff and external reviewers to be given training in evaluation ethical standards and stakeholder rights. There is also a need to have more staff training in NCEE in study sampling and non-sampling error issues.



## Appendix A

### ***Conclusions from the Fifth Follow-up Report of the National Evaluation of Upward Bound prepared by Mathematica Policy Research Published in January of 2009 by US Department of Education Study results***

By comparing the study's treatment group to its control group, this evaluation estimates the value-added effect of the opportunity to participate in Upward Bound—an unusually intensive precollege program—for the students who seek that opportunity and are eligible to participate in the program. The main findings are:

- ***Upward Bound had no detectable effect on the rate of overall postsecondary enrollment or the type or selectivity of postsecondary institution attended for the average eligible applicant.*** About four-fifths of both treatment group members and control group members attended some type of postsecondary institution, including four-year institutions, two-year colleges, and vocational schools, and the estimated impact is an increase of less than 2 percentage points in the rate of enrollment (effect size = 4 percent). For enrollment at four-year colleges and universities, the estimated impact is 1 percentage point (effect size = 3 percent). These effects are not statistically significant.
- ***Upward Bound had no detectable effect on the likelihood of applying for financial aid, or, the likelihood of receiving a Pell Grant.*** The 1 and 2 percentage point increases in the rates of financial aid application and Pell Grant receipt (effect sizes = 3 and 5 percent) are not statistically significant.
- ***Upward Bound increased the likelihood of earning a postsecondary certificate or license from a vocational school. It had no detectable effect on the likelihood of earning a bachelor's degree or the likelihood of earning an associate degree.*** While about 4 percent of control group members received a vocational certificate or license, nearly 9 percent of treatment group members did, implying an impact of 5 percentage points (effect size = 23 percent). The impacts on receiving any postsecondary credential and receiving a bachelor's degree are 2 and 0 percentage points (effect size = 5 and 0 percent), respectively, and are not statistically significant.
- ***Upward Bound increased postsecondary enrollment or completion rates for some subgroups of students.*** For the subgroup of students with lower educational expectations at baseline—that is, the students who did not expect to complete a bachelor's degree—Upward Bound increased the rate of postsecondary enrollment and the likelihood of receiving a degree, license, or certificate by 6 and 12 percentage points, respectively, raising the overall postsecondary completion rate to about the level observed for students with higher expectations. Because targeting on the basis of lower educational expectations might be challenging if it creates an incentive for applicants to understate their expectations, further analyses were conducted to examine the effects of Upward Bound on subgroups that might be more readily targeted. According to these exploratory analyses, Upward Bound increased postsecondary enrollment rates for students who were in tenth grade or above at the time of application, students who took a mathematics course below algebra in ninth grade, and students with a ninth grade GPA above 2.5. The estimated impacts were 3, 7, and 3 percentage points, respectively. Additional analyses suggest that Upward Bound also had positive impacts on postsecondary outcomes for some other subgroups defined by student- and project-level characteristics.

- ***Longer participation in Upward Bound was associated with higher rates of postsecondary enrollment and completion.*** *An additional year of Upward Bound participation was associated with a 9 percentage point increase in the rate of enrollment at four-year institutions and a 5 percentage point increase in the likelihood of receiving a bachelor's degree. Completing the Upward Bound program was associated with increases of 27 and 21 percentage points, respectively. These findings are based on nonexperimental methods, and the validity of causal inferences based on these estimates depends on the validity of strong assumptions.*

## Appendix B

### Summary of the *Joint Committee on Standards for Educational Evaluation (JCSEE)*

<b><u>Utility Standards</u></b>
The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.
<b>U1 Stakeholder Identification</b> Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.
<b>U2 Evaluator Credibility</b> The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.
<b>U3 Information Scope and Selection</b> Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders
<b>U4 Values Identification</b> The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.
<b>U5 Report Clarity</b> Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.
<b>U6 Report Timeliness and Dissemination</b> Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.
<b>U7 Evaluation Impact</b> Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.
<b><u>Feasibility Standards</u></b>
The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.
<b>F1 Practical Procedures</b> The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.
<b>F2 Political Viability</b> The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.
<b>F3 Cost Effectiveness</b> The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified
<b><u>Propriety Standards</u></b>
The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.
<b>P1 Service Orientation</b> Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.
<b>P2 Formal Agreements</b> Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.
<b>P3 Rights of Human Subjects</b> Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.
<b>P4 Human Interactions</b> Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.
<b>P5 Complete and Fair Assessment</b> The evaluation should be complete and fair in its examination and

recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.
<b>P6 Disclosure of Findings</b> The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results.
<b>P7 Conflict of Interest</b> Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.
<b>P8 Fiscal Responsibility</b> The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate
<b><u>Accuracy Standards</u></b>
The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.
<b>A1 Program Documentation</b> The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.
<b>A2 Context Analysis</b> The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.
<b>A3 Described Purposes and Procedures</b> The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.
<b>A4 Defensible Information Sources</b> The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.
<b>A5 Valid Information</b> The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.
<b>A6 Reliable Information</b> The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.
<b>A7 Systematic Information</b> The information collected, processed, and reported in an evaluation should be systematically reviewed, and any errors found should be corrected.
<b>A8 Analysis of Quantitative Information</b> Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.
<b>A9 Analysis of Qualitative Information</b> Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.
<b>A10 Justified Conclusions</b> The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.
<b>A11 Impartial Reporting</b> Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.
<b>A12 Metaevaluation</b> The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

## Appendix C

**Re-Analysis Findings reported by Cahalan, the PPSS Technical Monitor Executive Summary of *Addressing Study Error in the Random Assignment National Evaluation of Upward Bound: Do the Conclusions Change?* [http://www.coenet.us/files/files-do\\_the\\_Conclusions\\_Change\\_2009.pdf](http://www.coenet.us/files/files-do_the_Conclusions_Change_2009.pdf). (referred to here as the *Re- Analysis Report*).**

Major findings from analyses that attempt to correct or mitigate the identified study errors were as follows:

- Contrary to previously published results, when study error issues are addressed by using federal student financial aid (SFA) administrative records to supplement data for survey non-responders and adjusting outcome measures for students' expected high school graduation year (EHSGY), we found significant positive impacts of Upward Bound on postsecondary entrance and for applying for financial aid within +1 and +4 years of EHSGY. For example, we found impacts of 6.9 percentage points for "UB opportunity" or Intent to Treat (ITT) estimate, and 10.9 percentage points for the Treatment on the Treated (TOT) estimate for postsecondary entrance in +1 year. As these results include the bias introducing project 69 they probably underestimate the true effect of Upward Bound (figure IV).
- More robust results, estimating effects for the 74 percent of the sample not represented by project 69 show impacts of 9.1 percentage points for the ITT result and 14.2 for the TOT result for postsecondary entrance evidence in +1 year of EHSGY. Similar results were obtained using only the Student Financial Aid files to observe rates of applying for financial aid (tables 5-8 in report body).
- In observational two-stage instrumental variables regression taking into account but not eliminating selection effects, Upward Bound/Upward Bound Math/Science (UBMS) participation was also found to be significantly associated with positive outcomes relative to those who participated only in some other type of (presumably less intensive) "non-UB/non-UBMS pre-college support or supplemental" service (tables 7 and 9 in report body).
- Consistent with previously-published findings, large statistically significant positive effects were found on postsecondary entrance for the sub-group deemed to be of higher academic risk (bottom 20 percent on 9<sup>th</sup> grade academic indicators). Statistically significant positive findings, however, were also found for those in the top 80 percent on the same indicators (table 8 in report body).
- Overall, both the re-analysis and the Mathematica analyses found positive significant results for ITT and TOT estimates for UB for the attainment of any type of postsecondary degree or credential by the end of the study period in 2003-04 that was 7 to 9 years after expected high school graduation (table 10 and appendix table B-6). Weighted results based on fifth follow-up responders adjusted for non-response found ITT impacts of 10.6 percentage points (51.9 for the treatment compared to 41.4 for the control group) and TOT differences of 15.2 percentage points (54.4 compared to 39.3). These large differences were driven by positive impacts on the award of postsecondary certificates, a programmatic emphasis of project 69 which was supposedly representing a 4-year program stratum. Smaller but also significant impacts were found when project 69 is removed.
- As with postsecondary entrance, results for attainment of any degree or credential were also seemingly very large for those with lower expectations and also for those in the bottom 20 percent

on academic indicators (deemed more at risk) at baseline (table 10 and appendix table B-6). However, unequal weighting and the outlier project 69 characteristics emphasizing programs below the bachelor's degree may be affecting these results.

- Estimates for the attainment of the BA degree in +6 years that included the bias introducing project 69 were not significant. In estimates considered more robust, among the 74 percent of UB of the sample not represented by project 69 (based on the other 66 projects in the sample), there is a 28 percent increase in the probability of attaining a bachelor's degree in +6 years (17 percent for the treatment group and 13.3 for the control group) for the Intent To Treat (ITT) estimate and ***very importantly a 50 percent increase for the Treatment on the Treated (TOT) estimate (21.1 percent for the treatment group and 14.1 for the control group) (table 10 in report body).***
- In contrast to the results for any postsecondary degree or credential, considering BA receipt only, among the bottom 20 percent on 9<sup>th</sup> grade academic indicators, only three percent (25 unweighted cases) had evidence of attaining a bachelor's degree within +6 years of EHSGY. This sample number is too few for treatment-control group comparisons.
- Among the top 80 percent, on academic indicators, about 24 percent had evidence of a BA in +6 years and positive significant and substantial effects were found for the UB program for estimates with and without project 69 (table 14 in report body).

**Appendix D: Final Performance Report for the Final of Three Contracts for the National Evaluation of Upward Bound**

**U.S. Department of Education**

CONTRACTOR PERFORMANCE INFORMATION		
Contractor Name and Address (Identify Division) Mathematica Policy Research 600 Alexander Park Princeton NJ 085406346  (Please correct the above as needed.)		1. Contract Number: ED-04-CO-0152
		2. Type of Contract: Cost plus fixed fee
		3. Contract Value (Current plus any unexercised options): \$1,120,947.00
		4. Period of Performance (including any option periods): October 2004-November 2007
5. Description of Requirement: This contract was the final of three contracts for a longitudinal random assignment study of the Upward Bound (UB) and Upward Bound Math Science Program (UBMS). The requirements of this contract included processing of a final batch of postsecondary transcripts, analyses of fifth follow up data, and preparation of final reports for UB and UBMS. This contract did not cover the fifth follow up survey collection which was under a previous contract.		
6. Ratings: Summarize contractor performance and circle or type in the number below that corresponds to the performance rating for each category. Please see the attachment, which explains the rating scale.		
Quality: Quality of Product or Service -	0 1 2 3 4	<p>Comments: This rating is for the final contract, the last of three for the Upward Bound (UB) study. The extensive data collections for this difficult study were conducted carefully and were well done; and potentially provide much useful information. The previous contracts for this study have been given high ratings. The major tasks for this final contract were data analyses and report writing for the fifth follow up for the regular UB study and for the UBMS study. Other work for this contract consisted of updating the postsecondary transcripts. Approximately \$800,000 of the budget was spent on the analyses and reports and \$300,000 on the transcript updating.</p> <p>For this contract, my overall rating is (2) FAIR for the Quality of the Work. I first discuss the UBMS report and then spend the bulk of this review on the regular fifth follow up UB report which was a more major focus of the work under this contract. The UBMS report was adequate and would have been given a rating of 3. Much of the UBMS content was unchanged from a report from the third and fourth follow up rated previously. The UBMS report would have benefitted from some update of the descriptive information much of which was repeated word for word from the 2007 report. It would have also benefited from a more careful edit for consistency of format and for missing references.</p> <p>Unfortunately, my rating for the fifth follow up regular UB report is POOR. The basis for this rating is problems with the accuracy of the report conclusions. As noted previous ratings for the study have been positive. However, since those ratings were completed new information previously unknown to the Department of Education has come to light. This information was discovered in an in-depth QA review of the actual data files and the sample design. New information was also revealed when the sampling frame was finally delivered to ED at the end of the contract and the identity of the sampled institutions was revealed to ED allowing an assessment of the representativeness of sampling and weighting employed.</p> <p>As noted in PPSS internal reviews and in the IES external reviews, the analyses presented in the fifth follow up report used standard statistical modeling procedures, produces numerous impact estimates, and gives the appearance of careful research. However, as is also noted in the internal and external reviews the conclusions based on the impact estimates presented <i>cannot be considered robust</i>. PPSS QA re-examination found study impact estimates suffer from: 1) inadequately controlled for bias in favor of control group; 2) unequal weighting with one project with 26 percent of weight; and 3) serious</p>

	<p>representational issues with regard to 4-year grantees. When these issues are addressed using standard statistical methods—the study shows strong positive impacts for the UB program that are not acknowledged in the Mathematica report.</p> <p>PPSS found in QA analyses of the data that the Mathematica reports published in 2004 and sadly again in 2009 were flawed with an inadequately controlled for bias in favor of the control group. When these issues are addressed by standardizing the outcome measures by years since high school graduation there are statistically significant and substantial positive impacts clearly demonstrated on the major outcome measures for: postsecondary entrance, application for financial aid, award of the PELL grant, and attainment of postsecondary credentials for the entire sample. The Mathematica report concludes that UB had no detectable impact on “postsecondary entrance” or “financial aid,” when there is clear evidence to the contrary.</p> <p>In addition, the impacts reported by Mathematica do not reveal the fact that there were substantial and significant positive average effects for the program for 66 of the 67 projects in the sample taken together in virtually all of the outcome analyses using the same procedures and controls as those performed by Mathematica itself in the 2004 or the 2009 reports. One project in clear violation of statistical standards was selected as the sole representative of the largest 4-year stratum, --and carried 26 percent of the weight. Significantly, as is never revealed in the Mathematica published reports, it is a former community college with largely less than two year programs. The UB project partners with a job training program and cannot provide adequate representation for the diverse 4-year stratum it for which it is supposedly serving as the sole representative.</p> <p>Importantly, also not transparent in the published reports, this project (known as project 69) also introduces inadequately controlled for bias into the overall impact estimates in favor of the control group. Within this project for example, 80 percent of the high academic risk sample members were in the treatment group and 20 percent in the control group; 79 percent of the students expecting an advanced degree at baseline were in the control group and 21 percent in the treatment group; and 77 percent of the younger students were in the treatment group and 23 percent in the control group. This explains the seemingly negative findings detected for this one project considered by itself. This large bias combined with the large weights resulted in an overall control group that on average was older (56 percent higher grade students were in the control group and 44 percent in the treatment group), had higher academic indicators (58 percent of the high risk students were in the treatment group and 42 percent in the control group), and had higher educational expectations at baseline, (58 percent of the high expectation, defined as expecting an advanced degree at baseline were in the control and 42 percent in the treatment group). Controls used in the regression models by Mathematica did not adequately control for these differences in favor of the control group. (ie. there were no controls used for academic risk variables and the control used for grade at baseline was not keyed to a fixed academic year).</p> <p>As one would expect in a random assignment study, the 66 other projects taken together had about a 50/50 split between treatment and control group on these attributes found to be highly related to outcomes. Unfortunately, given its extreme weight project 69 confounded the overall results reported by Mathematica in 2004 and 2009; and masked the substantial and significant average positive findings for the 66 other projects taken together. For example, the models did not find significant effects on BA attainment in either the PPSS analyses or the Mathematica analyses when the bias introducing project 69 is included. However, both the Intent to Treat (ITT) and Treatment on the Treated (TOT) estimates for the average of 66 of the 67 projects show very large significant impacts for BA attainment--including a 50 percent increase in BA attainment in +6 and +8 years after high school for</p>
--	--



		<p>the TOT estimate (14 percent for control to 21 percent for treatment in the +6 years after high school estimate). The large significant findings for BA receipt (the goal of the UB program) are totally missed in the Mathematica reports that state there was “no effect detected on BA attainment.” An IES external reviewer stated that the estimates with project 69 could not be considered robust and indicated that those without project 69 could be considered to be more robust...but nowhere are these findings acknowledged to the reader of the Mathematica report.</p> <p>While formulaically implementing inverse probability of selection weighting and estimation procedures, the report demonstrates a lack of technical awareness and sensitivity to the complexity of the weighting and bias issues present in this study that have confounded the impact estimates. In addition, despite, the inclusion of 27 different estimates for each outcome measure, the outcome measures were unstandardized by expected high school graduation date—despite clear indication that the treatment and control group were not as balanced as one would have expected in a random assignment study on the distribution of older and younger students in a fixed academic year. The report is also not fully transparent in the appendix in discussion of the representational issues with project 69—never including information that it was a former community college with largely less than 4-year programs partnering with a job training program that was the sole representative of the largest 4-year public stratum....explaining the large apparent positive UB impact on obtaining certificates.</p> <p>Unfortunately, each of these issues was raised with Mathematica by PPSS prior to publishing of the final report and they chose not to revise their report to adequately address these issues. Alternative analyses that attempted to address these issues and that resulted in different conclusions were repeatedly made known to the contractor by ED and they refused to even acknowledge the existence of these analyses that reached different conclusions in their published report. The contractor also did not provide requested documentation for derived variables used in the analyses or exact descriptions of the modeling procedures used in estimation of impacts so that QA replication could be attempted. The description of modeling procedures provided in the report appears to be somewhat different than the final estimation procedures used.</p>
<b>Problem Resolution:</b>	0 1 2 3 4	<p>Comments: Rating of FAIR---The contractor concentrated on defending past work and decisions rather than being open to admitting the serious problems and addressing them. Critiques of the study were dismissed as advocates without a constructive attempt to understand the study issues. The study has had serious consequences for the UB program resulting in “ineffective” PART ratings, zero funding in Administration budgets, reform efforts based on study results, following objections from the TRIO community in Congress cancellation of funding for a more recent IES UB study, and finally legislative prohibitions in Congress for studies using the same RA procedures as this study. Much of this sad history could have been avoided if problems with the study had been transparently revealed and addressed at an earlier phase when they were first recognized.</p> <p>The following is a list of recommendations that would have increased the quality of the product that Mathematica produced on this contract and some suggestions for problem solving procedures that may improve performance in the future.</p> <ol style="list-style-type: none"> <li>1. Sample Design Flaws—Mathematica has indicated that the problems with the sample design stem from ED’s request that they ensure they could produce</li> </ol>

		<p>estimates for various sub-groups of interest to the program. However, an examination of the design indicates that there was not a chance that they could produce adequate estimates for these strata several of which only have one project included. It is a violation of commonly accepted standards for sampling when inverse probability of selection weights are to be applied, to only select one representative from any stratum and especially one that contains the largest number of cases, so that it will inevitably have the lowest probability of selection and hence a very large weight (in this case 26 percent of the study weight in one project of the 67 sampled). The design ended up with 339 strata at the student final stage with an average of only 8 cases per stratum. This was not an acceptable design. Mathematica, hired for expertise in this area, could have advised ED that this was not an acceptable design option and that serious issues would develop with this design.</p> <ol style="list-style-type: none"> <li>2. One of the major sources of problems with this study was the attempt to combine producing precise national estimates of all UB projects to give the design external validity with a random assignment experimental design measuring impacts of the intervention. It is very difficult to satisfy these dual goals (detection of impacts for an experimental design intervention and producing national estimates of the average effects). A recent systemic review of all random assignment studies sponsored by IES from 2003-2009 revealed that not one of them attempted to apply national inverse probability weights to the impact estimates. Communication to ED that this was not a reasonable approach would have been a reasonable problem solving solution to the OMB and ED request to have nationally representative estimates.</li> <li>3. One might have been able to excuse the initial sample design issues as not being fully apparent when the first stage sample was drawn. The unequal weighting issue was apparently made worse by the fact that project 69 had an unusually large number of applicants relative to the openings and hence a much larger control group than treatment group, as well as a larger proportion of the total sample weight by the end of the student stage. However, it is difficult to understand continuing to put forth estimates known to be so problematic in the Fifth Follow up report. To mitigate the sample design flaws and bias issues Mathematica was repeatedly asked by ED to report the data unweighted and weighted; and with and without project 69. This suggestion was repeatedly refused even when expert opinion was given that the national estimates were not robust in the external IES review. These suggestions first made by the PPSS COR were then repeated by the PPSS Division Director following the IES review and were again not implemented.</li> <li>4. It's clear that the sample suffered from what is known as a bad draw—a case randomly selected that is atypical of the projects in the stratum it is representing. Mathematica should have checked the sample by stratum when it was first selected to ensure that sample members were representative of their stratum. The common procedure followed in most sample studies with weights is to always check a sample before using it to ensure that it can lead to good estimates of “known totals”. Checks should have been done using IPEDS to ensure that the types of degrees awarded were similar to those of the stratum members and that the race/ethnicity was representative of the average distribution of the stratum. In an informal discussion, ED has been told that a member of the Mathematica staff actually visited project 69 and observed that they partnered with a job training project and gave out construction certificates. This should have been a red-flag that this one project was not an adequate representative of a 4-year stratum that contained the major research universities that had UB projects at the time. When ED at the end of the contract was finally</li> </ol>
--	--	--

	<p>provided with a listing of the sample frame (after repeatedly requesting it), we found that project 69, was a former 2-year college with largely two year and less than two year degrees awarded, and that it is very different from the other projects in its stratum which included the major research universities who had UB projects, and historically black 4-year colleges, as well as 4-year colleges in which the majority of the participants were white. There were also no white participants in the project 69 sample (60 percent black and 40 percent Hispanic). The discussion of project 69 in the appendix G of the report, no where reveals the issue with project 69 having primarily less than 4-year programs; and indeed maintains that project 69 is an adequate representative of its stratum. When ED asked that this information be added to the report Mathematica refused. Mathematica has admitted in conversations that project 69 was a “bad draw”, as if the UB program was handed a bad poker hand and there was nothing they could do about it. But statistical sampling and random assignment with an adequate number of sampled cases is done precisely to avoid the imbalances of a poker game with a limited number of cards leading to one group having a hand that is much more favorable than another. The normal procedure in sampling, when one knows one has a “bad draw” as manifest by poor capacity of the sample to represent its stratum on known totals or large imbalances between the treatment and control group is to re-draw the sample for the stratum. This should have been done before the study ever started, but once it was realized—presumably some time later, it seems inexcusable that Mathematica would not have corrected this error in some manner and would continue even now to refuse to be transparent in admitting this publically to the stakeholders in this study.</p> <p>5. PPSS first became aware of the project 69 issue in 2005, when a Mathematica staff person (now no longer employed there) sent us tabulations that showed the results for the fourth follow up were sensitive to only one project with a large number of cases (85) and seemingly very negative outcomes and also for the first time informed ED that the project carried 26 percent of the weight (this fact was not in previously published reports in 2004 and was only put into the fourth and fifth follow up reports at ED’s insistence). At this time PPSS asked that Mathematica do some research to try to explain why the UB services by this project supposedly encouraging college going and giving academic preparation would actually result in negative impacts on college going for those selected for treatment. We never received any answer from Mathematica, but when we began to look at the data files a year later, the reason became clear. As noted above there were very strong bias in favor of the control group on grade at a fixed time point; academic risk factors (not revealed in Mathematica’s discussion of project 69) and in educational expectations. These differences explain the seemingly negative impacts of this project. The other 66 projects taken together are well balanced in the treatment/control group attributes and taken together consistently demonstrate positive impacts. Mathematica should have investigated these negative findings when asked to by ED—if only to find out what the project was doing wrong—so that something could be learned from the \$14 million dollar study. In fact there is evidence that this project performed very well in getting its at-risk treatment group participants into college and obtaining certificates. <b>Unfortunately these uncontrolled for treatment/control group biases, extremely large weights, and representational issues resulted in a distorted overall view of UB –making it seem like it had no effect on entering college and no effect on BA receipt, but did have a large effect on obtaining less than two year certificates!</b></p> <p>6. The stakeholders in this study (UB projects, Department of ED program office (OPE) and the PPSS COR, and finally by the end of the contract the Division Director who had been the original COR for the study when it was begun in</p>
--	--



		<p>1992 have all expressed serious concerns about the estimates from the study. PPSS provided detailed reviews of the fifth follow up report, and gave concrete suggestions for how this report could have addressed the problems noted above using standard statistical techniques. PPSS consulted with outside statistical experts who gave recommendations that were given to Mathematica—present the data weighted and unweighted and with and without project 69; standardize outcome measures to expected high school graduation to deal with the control group being older than the treatment group; and do not include National Student Clearinghouse data in the early days of the clearing house (1996 &amp; 1997) when coverage was 25 percent and it was known that project 69 with such a large number of cases was not a member when its students were graduating high school. None of these suggestions that would have mitigated the problems with the sample design were implemented by Mathematica in their reports upon which they based conclusions. Problem solving would have been helped if Mathematica had demonstrated a bit more sensitivity to the programmatic knowledge and legitimate rights and concerns of the stakeholders to this evaluation. Mathematica never adequately addressed the technical concerns expresses about their reports by PPSS and dismissed legitimate concerns as those of “advocates” for the program.</p>
<b>Cost Control:</b>	0 1 2 3 4	<p>Comments: (3) Good—study requirements were completed within the budget; However, this was a contract in which over \$700,000 was spent on analyses and the two reports for the study. Several analyses requested and files documentation of derived variables and statistical output was not provided with the excuse that there were not adequate resources in the contract.</p>
<b>Timeliness:</b>	0 1 2 3 4	<p>Comments: (3) Good— delays were for the most part related to additional data procurement</p>
<b>Business Relations:</b>	0 1 2 3 4	<p>Comments: (2) FAIR Mathematic's corporate leadership was not responsive to ED's concerns about the study and did not try to facilitate problem resolution.</p>
<b>Customer Service:</b>	0 1 2 3 4	<p>Comments: POOR---The customers are UB eligible population, US Congress, ED, Upward Bound Projects, and public tax payers. The study potentially provides much useful information. However, given that the reports contain unwarranted conclusions concerning the UB program, it's difficult to assert that the customers were well served overall. The UB grantees being evaluated had a right to an accurate and fair evaluation and in my view this right was violated. For more detailed information on the technical issues and alternative analyses implementing the PPSS recommendations discussed above please see-- - Cahalan, Margaret, <i>Addressing Study Error in the Random Assignment National Evaluation of Upward Bound: Do the Conclusions Change?</i>, Council for Opportunity in Education, 2009 <a href="http://www.coenet.us/files/files-Do_the_Conclusions_Change_2009.pdf">http://www.coenet.us/files/files-Do_the_Conclusions_Change_2009.pdf</a></p>
<b>7. Total score: 13</b>		

**Evaluated by:**

Agency/Organization Office of Planning, Evaluation and Policy Development (OPEPD), Policy and Planning Studies Services (PPSS) \_\_\_\_\_

Date 11-15-2009 \_\_\_\_\_

(In accordance with the Federal Acquisition Streamlining Act, the following information will not be released to the contractor.)

Name and Title \_Margaret Cahalan \_Team Leader Secondary Postsecondary Cross Cutting Team and COR for contract

Telephone number 202-401-1679

Signature \_\_\_\_\_*Margaret Cahalan*\_\_\_\_\_ E-mail address \_margaret.cahalan@ed.gov\_\_\_\_\_

## Appendix E

### Re-analysis finds Strong Positive Results from the Random Assignment National Evaluation of Upward Bound (UB)

**By Margaret Cahalan** *All views and results reported are the sole responsibility of the author and do not reflect any review or authorization by the US Department of Education. Dr. Cahalan serves as the Secondary-Postsecondary Cross-Cutting Team Leader within the Policy and Program Studies Services (PPSS) within the US Department of Education, Office of Planning, Evaluation and Policy Development (OPEPD).*

This fall the American Youth Policy Forum (AYPF) published an informative report entitled *Success at Every Step*, reviewing research results concerning 23 pre-college programs (<http://www.aypf.org/publications/SuccessAtEveryStep.htm>). Upward Bound (UB) is among the programs reviewed and *Success at Every Step* summarizes the lack of overall impact on postsecondary outcomes results reported in the 2009 final report from the National Evaluation of Upward Bound authored by Mathematica Policy Research available at the following Department of ED web address (<http://www2.ed.gov/about/offices/list/oepd/ppss/reports.html#higher>).

In the light of the continued widespread quoting of the Mathematica results, I'm writing this article to explain why I believe that these conclusions are unwarranted concerning the UB program. Put simply, as the Department of Education (ED) Technical Monitor for the final period of the study, in a QA examination of the data, confirmed by consultation with independent statistical experts, I found that the postsecondary results in the Mathematica report *suffer from inadequately controlled for bias in favor of the control group, and a sample with serious representational flaws and unequal weighting issues*. In the interest of full government transparency this article is also written as an effort to comply with professional evaluation standards that state:

*The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results, (Standard P6 in Joint Committee on Education Evaluation Standards).*

I also wish to share some alternative results that I believe, give a more credible representation of the findings from this evaluation. These re-analyses show that when identified study flaws are addressed using standard statistical procedures there are positive impacts for the UB program on the goals of the program—postsecondary entrance, award of financial aid, and degree or credential attainment. My results were published by the Council on Opportunity in Education (COE) last fall in a report entitled *Addressing Study Error in the Random Assignment National Evaluation of Upward Bound: Do the Conclusions Change?* ([http://www.coenet.us/files/files-do\\_the\\_Conclusions\\_Change\\_2009.pdf](http://www.coenet.us/files/files-do_the_Conclusions_Change_2009.pdf)) (referred to here as the “re-analysis report”).

This article is not intended to be a critique of the random assignment method. While I strongly disagree with the research transparency and analysis choices made by Mathematica, I also believe that the National Evaluation of Upward Bound is among the most carefully conducted and useful studies we have in the area of pre-college research. The study collected detailed information on services and outcomes from a baseline and five follow up surveys conducted over 10 year period with high response rates, the lowest of which was over 70 percent.

**Concern with the Published Report.** Dr. David Goodwin, the original UB study monitor and former Director of the ED division in which the study was conducted, now of the Gates Foundation, in the context of requesting an IES external review of my re-analysis report wrote: “*I believe the Mathematica report currently published by ED is seriously flawed.*” (July 2009 email memo quoted with permission of the author). Dr. Goodwin has summarized the issues with the report as follows:

1. The design involves a nationally representative sample of 67 Upward Bound projects. One of these projects (called project 69) carries 26 percent of the total weight. A sample design placing one-fourth of the weight on one project would not be regarded by sampling statisticians as sound practice (see figure I).

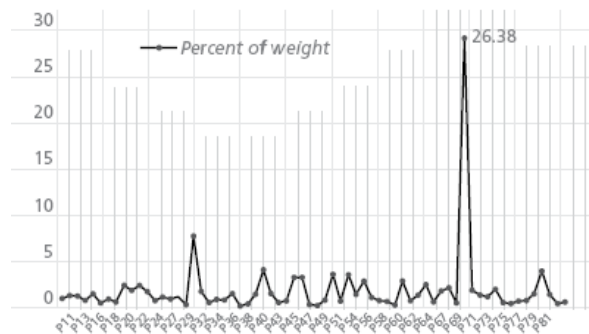


Figure I shows that one project known as Project 69 accounts for fully 26.4 percent of the total weight for the sample.

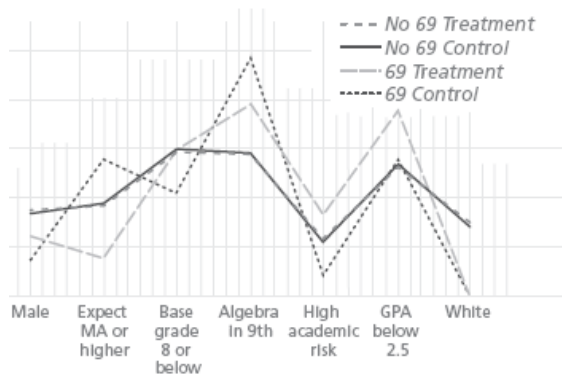
**FIGURE I. Percentage distribution of sum of the weights by project for the 67 projects making up the study sample**  
National Evaluation of Upward Bound,  
study conducted 1992-93 to 2003-04

**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 Program Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), U.S. Department of Education; study conducted 1992-9 to -2003-04.

2. The Mathematica study results, which do not standardize outcomes by expected high school graduation year, depend on the presence or absence of project 69. Mathematica’s own internal analysis (using unstandardized outcome measures) indicates that without project 69 there are significant positive results.
3. A closer look at project 69 indicates that there are large differences between treatment and controls on characteristics likely to affect postsecondary outcomes, several of which cannot be adequately controlled in the statistical analysis. These differences explain the seemingly negative impacts in this site. Within project 69, controls were less likely to be considered “academically at risk,” had higher educational aspirations, and entered UB at a higher high school grade. (For example 80 percent of the “academically at risk students” from project 69 were in the treatment group and 20 percent in the control group). Considering the rest of the study sample taken together, the treatment and control groups were well matched. However, combining the project 69’s weight with these treatment/control differences and the inability to adjust for academic risk, raises serious questions about biases in the impact estimates (see figure II and figure III).

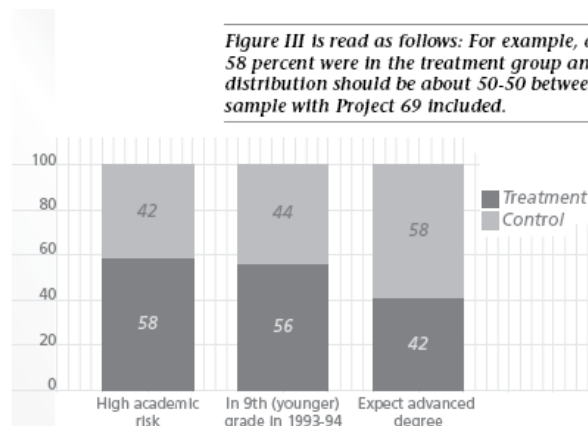
Figure II shows that the UB treatment and control group are well matched without Project 69 on the variables in the chart; however, in Project 69 the treatment and control group manifest substantial differences. For example, 56 percent of the control group in Project 69 expected an MA or higher at baseline compared with 15 percent of the treatment group. In contrast, among the other 66 projects in the sample, 38 percent of the control group and 37 percent of the treatment group expected an MA or higher.



**FIGURE II. 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 to 2003-04

**NOTE:** Project 69 tabulation based on the 85 sample cases from Project 69 (52 controls and 33 treatment cases—poststratified weighted to 11,536 cases—5,768 treatment and 5,768 controls). The category “No 69 treatment” and “No 69 control” 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 Program 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.



**Figure III is read as follows:** For example, among those who were classified as higher academic risk, 58 percent were in the treatment group and 42 percent in the control group. In a random assignment study distribution should be about 50-50 between treatment and control group; figure shows imbalance in overall sample with Project 69 included.

**FIGURE III. Percentage distributions for all 67 sampled projects (including Project 69) between treatment and control groups among those sample members who were a higher academic risk, in the 9th (younger) grade in 1993-94, and who expected an advanced degree at baseline**  
National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04

**NOTE:** High academic risk includes those sample members in the bottom 20 percent of the sample on 9th grade GPA and other academic indicators.

**SOURCE:** Data tabulated April 2009 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04.

- Project 69 is not representative of the strata it was chosen to represent—not-majority Hispanic, not-rural, 4 year college. Until a few years prior to the study, project 69 had been a community college. It was subsequently taken over by a nearby 4-year college and reclassified; however, in offerings and other characteristics it remained very much like a community college. The project 69 UB program partners with a job training program and was not an adequate sole representative of this 4-year stratum that included the major research universities that had UB grants at the time. It also lacked a summer residential component, a key feature of almost all Upward Bound projects at 4 year colleges, and relatively rare at community colleges.

The Mathematica reports are especially troubling in that they do not provide information that would allow the reader to understand the full extent of study sampling and non-sampling error issues. Dr. Goodwin goes on to note: that “*Virtually all the issues summarized above are not evident from reading the Mathematica report. Although Mathematica may have followed procedurally acceptable methods, it seems clear that their main findings are not robust to alternative analyses that, at a minimum, are equally valid.*” Internal and external ED reviewers concluded that the national impact estimates could not be considered robust and there is a high probability that the combination of the issues noted above resulted in a Type II error of lack of detection of UB impacts on postsecondary entrance and on BA receipt, and a Type I error

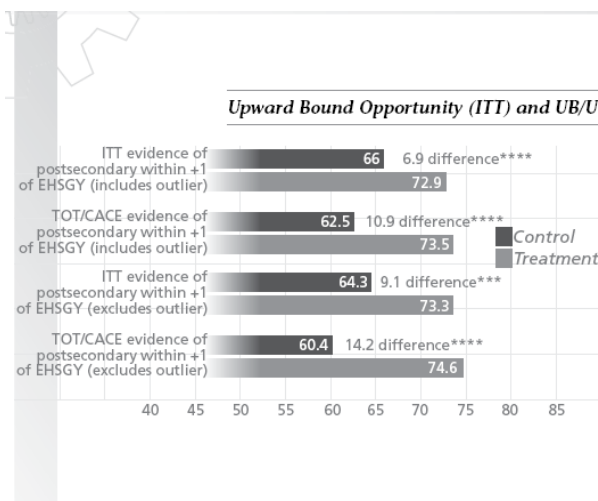


of over estimation of the size of the UB impact on the award of certificates, a programmatic emphasis of the project 69.

**The Re-Analysis.** The re-analysis report uses similar logistical and instrumental variables regression procedures and models to estimate impacts as the Mathematica analyses. It also presents Intent to Treat (ITT) and Treatment on the Treated (TOT) impact estimates using the same treatment and control groupings. However, consistent with the suggestions of ED internal and external reviewers, the re-analysis report differs from the Mathematica reports in three major ways.

1. Impact estimates in the re-analysis report are presented both weighted and unweighted and with and without the outlier and confounding project 69.
2. Most outcome measures are standardized to reflect common periods after high school (for example—+18 months, +6 years). Expected high school graduation year (EHSGY) in the sample spanned from 1995 to 1998. Mathematica reports combine outcomes as of date of survey rounds and the QA analysis found there was an inadequately controlled for imbalance with the control group being more frequently in a later grade at a fixed time point. Using unstandardized outcome measures also confounded the relationships of the control variables to the outcome measures in the models.
3. The re-analyses takes a more conservative approach than Mathematica to using the National Student Clearinghouse (NSC) data for postsecondary entrance. NSC reports on their website that they only reached 25 percent of enrollment coverage by 1996 and notably project 69 was not participating in the most applicable period. Coverage was much higher for 4-year institutions for the time frame when bachelor's degrees might have been obtained and the re-analysis uses NSC data for looking a BA receipt.

**Re-Analysis Findings.** In contrast to the Mathematica reported results, the re-analysis found that if outcome measures were made more precise by standardizing by expected high school graduation year (EHSGY) using data from each of the three applicable follow up surveys (third through fifth) and ten years of federal aid records, that there were substantial and significant positive effects for evidence of postsecondary entrance and financial aid application and award for the ITT and TOT estimates weighted and unweighted and with and without the bias introducing project 69. Effect sizes are consistently larger without project 69.



**FIGURE IV. Estimated rates of evidence of postsecondary entrance within +1 (about 18 months) of expected high school graduation year (EHSGY) for Upward Bound Opportunity (ITT) and UB/UBMS participation (TOT)**  
National Evaluation of Upward Bound, Study conducted 1992-93 to 2003-04

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

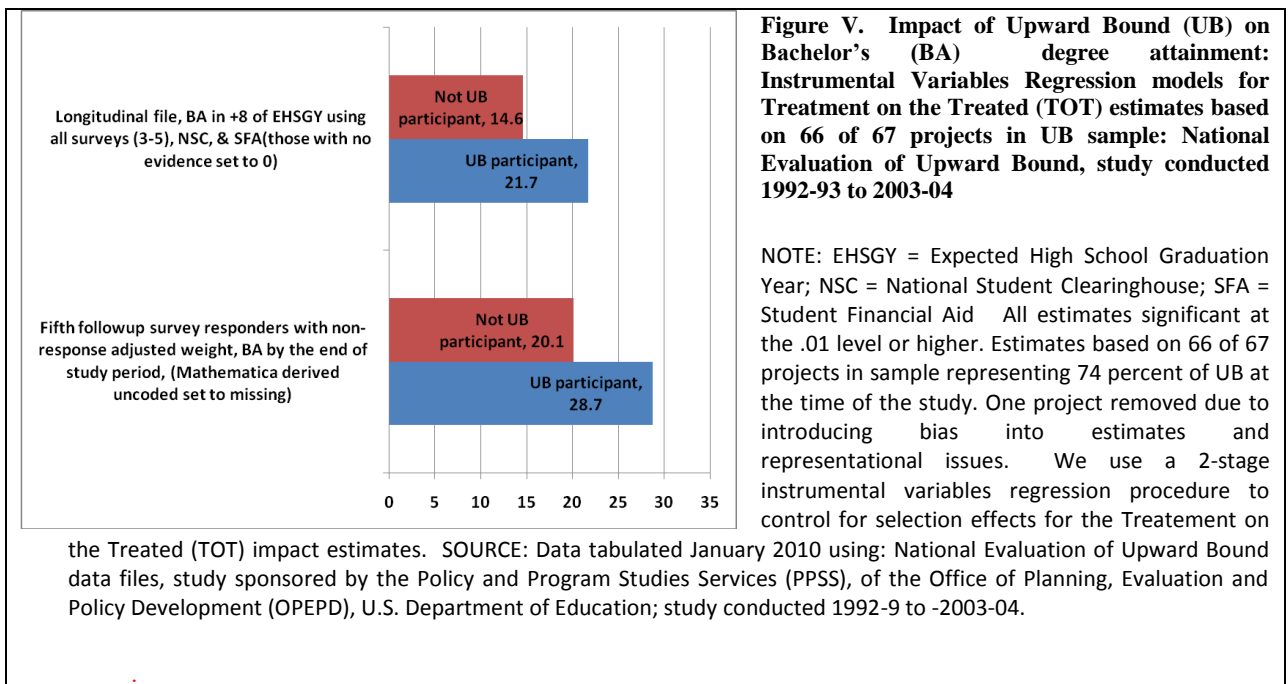
**NOTE:** UB = regular Upward Bound; UBMS = Upward Bound Math-Science; ITT = intent to treat; TOT = treatment on treated; CACE = complier average causal effect. Model based estimates based on STATA logistic and instrumental variables regression taking into account the complex sample design. Weighted estimates use poststratified weights. See table 5 in body of the report for detailed note.

**SOURCE:** Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.

The re-analyses found significant and substantial impacts for the award of any postsecondary degree or credential by the end of the study period; however, with the bias introducing project 69 included neither the Mathematica analyses or the Re-analyses found results for BA attainment. With project 69 included the finding of positive

impacts for attainment of any degree or credential were driven by large impacts for “certificate” attainment reflecting the type of postsecondary credentials awarded by the highly weighted project 69 that was supposedly representing the largest 4-year public grantee stratum.

Among the results that are most troubling in their absence in the Mathematica reports are the substantial impacts found on BA attainment for 66 of the 67 projects taken together when the confounding project 69 is removed (figure V). For example, with the complete longitudinal file using survey data, financial aid records data, and NSC data, the Treatment on the Treated (TOT) instrumental variables regression estimates of impact show a large increase in BA receipt by +8 years after expected high school graduation--- *going from estimated 14.6 percent for the control group to 21.7 percent for the treatment group—an increase of 49 percent (figure V)*. Similar impacts were found using variables derived by Mathematica from the fifth followup survey responders only by the end of the study period (figure V second comparison in chart). Statistically significant impacts with lower effect sizes were also found for the Intent to Treat (ITT) estimates.



## Appendix F

### ***Email from Dr. David Goodwin, former director of PPSS Unit (PAS) responsible for the study requesting review of Cahalan paper and indicating why he believed that the Mathematica report was “seriously flawed” (July 2009)***

**From:** Goodwin, David  
**Sent:** Wednesday, July 22, 2009 12:02 PM  
**To:** Ricciuti, Anne  
**Cc:** Ginsburg, Alan  
**Subject:** Upward Bound external review

I'm following up our conversation of about a week ago to find out if IES is willing to arrange an external peer review for Margaret Cahalan's reanalysis of data from the Mathematica Upward Bound evaluation. Since you indicated some skepticism based upon an IES staff review of Cahalan's paper, *I'd like to briefly summarize why I believe the Mathematica report currently published by ED is seriously flawed:*

The design involves a nationally representative sample of 67 Upward Bound projects. One of these projects (called project 69) carries 26 percent of the total weight. Each of the remaining 66 projects typically carry about 3-5 percent weight. A sample design placing one-fourth of the weight on one project would not be regarded by sampling statisticians as sound practice.

The study results depend totally on the presence or absence of project 69. Mathematica's own internal analysis indicates that without project 69 there are significant positive results, but with 69 included there are no impacts on postsecondary attendance.

A closer look at project 69 indicates that there are significant differences between treatment and controls on characteristics likely to affect postsecondary outcomes, several of which cannot be adequately controlled in the statistical analysis. Within project 69, controls were less likely to be considered “academically at risk,” had higher educational aspirations, and entered UB at a higher high school grade. Considering the rest of the study sample, treatments and controls were quite similar. However, combining the project 69's weight with these treatment/control differences and the inability to adjust for academic risk, raises serious questions about biases in the impact estimates.

Project 69 may not be very representative of the strata it was chosen to represent—non-white, urban 4 year college. Until a few years prior to the study, project 69 had been a community college. It was subsequently taken over by a nearby 4 year college and reclassified; however, in offerings and other characteristics it remained very much like a community college. It lacked a summer residential component, a key feature of almost all Upward Bound projects at 4 year colleges, and relatively rare at community colleges.

Virtually all the issues I've summarized above are not evident from reading the Mathematica report. Although Mathematica may have followed procedurally acceptable methods, it seems clear that their main findings are not robust to alternative analyses that, at a minimum, are equally valid. This is why I think it is important that the Department publish the Cahalan paper. Given the IES staff review, without an independent external review of Cahalan's research, it's not going to happen.

David Goodwin  
Policy and Program Studies Service  
U.S. Department of Education  
400 Maryland Ave., SW  
Washington, D.C. 20202  
202-401-0263 (tele)  
202-401-5943 (fax)

**Appendix G**  
***Chronology of Events Related to the Publication of the Mathematica Fifth Follow-up Report: Spring 2005 to January 2009 (Prepared Margaret Cahalan)***

**Spring 2005**—Mathematica lead analyst for the Fourth Follow-up prepared tables that were sent to PPSS showing that one project that carried a large weight (26 percent) had strong negative impacts and that inclusion or exclusion of this project changed the conclusions of the study (see Appendix H). PPSS asked Mathematica to investigate the project further.

**Summer 2005**—Mathematica delivered draft of *Fourth Follow-up Report*. PPSS comments express concern about a number of issues including: sample design issues, lack of standardization of outcomes, lack of use of administrative records, lack of recognition of UBMS participation as a crossover issue, and obvious effects of unequal weighting on the stability of the estimates.

**Spring of 2006**—Mathematica submits revised draft *Fourth Follow-up Report*. Major technical issues raised in review of the report are not addressed to PPSS satisfaction

**Summer of 2006**—PPSS arranges for obtaining the Federal Student Aid (SFA) files for entire UB sample for 10 years of data to be used in addition to the survey data in revising the *Fourth Follow-up Report*

**Summer of 2006**—PPSS requests and receives copies of the Mathematica data files through the Fourth Follow-up and PPSS merges data with the Federal Student Aid files. Cahalan conducts re-analysis of Fourth Follow-up data and obtains different results from those in the Mathematica draft report.

**October 2006**—PPSS briefs OPE, Budget and IES concerning the different findings when the data are merged with the Federal Student Aid (FSA) files. For the Fourth Follow-up data significant positive impacts were found on key postsecondary outcomes for the entire sample

**November 2006**—PPSS briefed Mathematica on the findings on the *Fourth Follow up* data and requests changes to the draft *Fourth Follow-up Report* based on the issues raised and the alternative findings when the Federal Aid Files are included.

**January-February 2007**—Mathematica responds that there are not enough budget resources to both revise the *Fourth Follow-up* report and also completed the *Fifth Follow-up Report*. A joint PPSS-Mathematica decision was made to concentrate on the *Fifth Follow-up* report for which analysis was well underway. Mathematica subsequently requested and was granted \$25,000 additional from OPE to complete the Fifth Follow-up report funds (bringing the total amount for the Fifth Follow-up Analysis contract to \$1,120,947. The total for the three contracts for the Evaluation was about \$14 million)

**January 2007**—PPSS begins consultation with RTI (who held the PPSS Statistical Analysis and Technical Support Contract) to obtain an outside expert opinion concerning the sampling design issues. PPSS sends the Cahalan paper to RTI for statistical review. PPSS also sends UB data files to RTI for review of the sample. PPSS also asks RTI to try to replicate Cahalan's findings.

**March 2007** -- RTI's Dr. James Chromy, a Fellow of the American Statistical Society and recognized expert in both sampling statistics and random assignment for clinical trials, as well as NAEP and NPSAS sampling issues reviewed the UB sample design.. Dr. Chromy met with Dr. Cahalan and Dr. Goodwin in DC and reported that the issues Cahalan had raised concerning the sample design and implementation were valid and serious concerns. Dr. Chromy also reported that his staff had replicated Cahalan's STATA results using their SUDANN program. It was agreed that PPSS would request a copy of the sampling frame used to draw the sample from Mathematica so that RTI could do further checks on the sampling design and bias issues of concern. Subsequent to this request from RTI for the sampling frame PPSS requested that Mathematica send a copy of the sampling frame to RTI. Mathematica responded that they were unable to locate a copy of the sampling universe frame used in 1992 but would continue efforts to locate the frame. (Note ten months later after the contract ended in December of 2007, Mathematica delivered a PDF scanned copy of a printout of the frame to PPSS).

**March-April 2007**—Based on consultation from Dr. Chromy, PPSS recommends that in the draft report for the *Fifth Follow-up* data be reported with and without project 69 and weighted and unweighted. PPSS recommended that Mathematica acknowledge that there were serious sample design issues and that they could not make robust estimates for the entire population but only for the 74 percent of the population not represented by project 69. PPSS also recommended that the National Student Clearinghouse data be used very conservatively due to lack of coverage concerns. These recommendations were rejected by Mathematica.

**August 2007**—Mathematica submitted the first draft of the *Fifth-Follow-up Report* to PPSS. Margaret Cahalan and Jay Noell prepared detailed comments on the report requesting substantial changes to the report similar to what they had recommended prior to the report drafting.

**November 2007**—As all funds had been spent for the project, Mathematica did not request a time extension. Even though a report acceptable to PPSS had not been completed, ED could not legally request that Mathematica continue to work unless more funds were provided. No additional funds were available and the Time and Materials contract officially ended on November 30 2007.

**December 2007**—Mathematica sent to PPSS a PDF of the sampling frame for the study which enabled PPSS for the first time to learn the identity of the project 69 grantee. Prior to this time Dr. Cahalan and Dr. Goodwin and Dr. Chromy were unaware of the serious representational issues with project 69 and indeed the Mathematica reports indicated and still do indicate that it was "typical" of the other projects in its stratum. Upon learning the identity of the project and

the characteristics of the grantee institution and the project, and observing the characteristics of the other projects in its stratum, Cahalan and Goodwin became more concerned about project 69's capacity to adequately represent its stratum which was carrying 26 percent of the total weight for the study.

**December 2007**—A meeting was held between PPSS staff (Goodwin, Cahalan, and Noell) and Mathematica staff (retiring President, Chuck Metcaff, Mary Moore, Alan Schirm and Neil Sefter). At this time PPSS went over again its recommended changes. Mathematica disagreed with the recommendations and refused to make the major changes but agreed to some minor changes. Dr. Metcaff suggested that Cahalan prepare a “minority” report on the topic.

**January 2008-April 2008**—Cahalan first began working on the detailed task of standardizing the outcomes by Expected High School Graduation Year (EHSGY). In addition to ascertaining the expected high school graduation year, this effort also required determining the years of entrance into postsecondary and also the years of any degree awards. This meant that all applicable surveys (baseline and five follow-ups) and 10 years of Federal aid files and NSC data had to be examined to derive the variables needed. Cahalan also did additional work on project 69 and uncovered the extent of the bias coming from the non-equivalence of the treatment and control group from this project and also the representational issues—that it was very atypical of the stratum it was representing. Knowing the identity of the project helped explain the large UB impact on “certificate attainment.”

**April 2008**—After review of Cahalan's positive findings when results are standardized by Expected High School Graduation Year (EHSGY), Goodwin arranged for Cahalan to brief OPE, Budget and IES on her findings. Cahalan found positive impacts for postsecondary entrance in 1 and 4 years after EHSGY and similar findings for Federal Aid Application and Pell Receipt.

**April 2008**—Cahalan sent copy of her paper with her results when standardization to EHSGY to Mathematica and suggested they acknowledge these findings in their revisions to the Fifth Follow-Up Report upon which they were working.

**April 2008**—Mathematica sends a revised draft of the *Fifth Follow-up report* in response to PPSS comments made in September of 2007 to December of 2007. The report does not address key concerns raised by PPSS reviewers or acknowledge in any way the Cahalan paper results with positive findings contradicting the conclusions they reached..

**April 2008**—Goodwin recommended to OPEPD front office that ED publish both the Mathematica report and also the Cahalan report.

**May 2008** Mathematica responded to Cahalan's paper with memo's defending their technical decisions and questioning Cahalan's method of establishing EHSGY. In response Cahalan prepared a memo that included estimates using an alternative means to establish EHSGY (see Appendix P) and found the same results using two different ways to estimate EHSGY.

**May 2008**—Assistant Secretary Evers asks IES to conduct external review indicating that there cannot be two different conclusions from the same data. Goodwin and Cahalan assume the review will be of both reports (Mathematica and Cahalan's).

**May 2008**—Cahalan drafts memo to Assistant Secretary Evers asking permission to give the same briefing that she gave to ED staff to the COE concerning the issues with the study (see Appendix I). She is advised by Alan Ginsburg not to send the memo but to wait until after the IES external review before taking any action.

**Late May 2008**—PPSS is informed by the IES review office that they have been directed by OPEPD and IES leadership that only the Mathematica report is to be considered/reviewed for publication and that Cahalan's report will not be considered for publication. Dr. Goodwin wrote a memo to Ze've Wurman indicating his position that either both reports be published or none be published. However, he was overridden by the OPEPD front office. Cahalan also reports that PPSS was not allowed to submit the questions they had prepared for the blind IES reviewers and unlike the normal procedure in external review of reports, the IES reviewers were not given copies of PPSS staff (Cahalan and Noell's) extensive reviews of the Mathematica Fifth Follow-up report itself.

**June 2008** —After learning that her paper will not be considered for ED publication and that the IES reviewers were not even to be given her extensive comments on the Fifth Follow up report, Cahalan decides that she must no longer delay in informing the Stakeholders to the evaluation. Goodwin and Cahalan then consulted an ED lawyer in OGC. The lawyer indicated that a review of regulations seemed to indicate that acting on her own time and not as a representative of the government, Cahalan could prepare and deliver a critique of the already published *Third Follow-up* report (Myers et.al 2004). . Without asking for permission from her supervisor's, acting on her own annual leave time--- Cahalan then arranged to meet with Maureen Hoyler as a representative of the stakeholders. Cahalan shared her concerns and gave her a copy of a paper that critiqued only the Third Follow-up published results (see Appendix J, Cahalan's email to ED colleagues indicating what she had done and why she did it).

**July-August 2008**—The comments of the three IES external reviewers are forwarded to PPSS. As noted, the blind IES external review which was set up by Assistant Secretary Evers in July of 2008 was not done in a manner that PPSS expected or requested. Dr. Goodwin requested and expected that both Mathematica's and Cahalan's reports be reviewed for PPSS publication. However, IES and OPEPD leadership decided that only the Mathematica report would be reviewed and considered for publication. Unlike the standard procedure for such external reviews that involves having all comments from previous reviews be submitted along with the report, the blind reviewers were not given copies of the extensive internal reviews that PPSS staff members Dr. Jay Noell and Dr. Cahalan had already completed. Instead they were given a copy of Cahalan's paper on the Third Follow-up re-analysis which did not provide a direct comparison to the Fifth Follow-up Report or comments on the actual report they were reviewing. The IES reviewers noted that Mathematica had followed generally accepted procedures for random assignment studies. At the same time although specifically instructed not to comment on the Cahalan paper, they also upheld the technical correctness of issues

Cahalan raised and praised her paper as well researched. They all agreed with her that the sample design was seriously flawed, that standardization of outcomes was correct, and that there should be conservative use of the NSC data. But only one of the reviewers (the one reviewer who was a sampling statistician) gave clear suggestions as to what to do about the issues or clearly states that the Mathematica impact estimates for the entire population could not be considered robust. Moreover, not having access to the data, the IES reviewers relied on Mathematica's misleading statements in the report that project 69 did not make a significant difference in the conclusions. *While they commented that the Cahalan paper was well researched and technically sound, they stated they wished there had been a more direct comparison to the Fifth Follow-up.*

**August 2008**—A memo is drafted to Mathematica summarizing the changes required. The memo was first drafted by Jay Noell and then re-drafted by Dr. Goodwin and then edited to require fewer changes by Ze've Wurman on Bill Evers staff. The memo called for presenting the data weighted and unweighted and with and without project 69.

**September --November 2008**—Mathematica sends memo rejecting the major recommendations outlined by Dr. Goodwin. In early November, Mathematica sent a revised draft of the report (see response of Goodwin in Appendix K to Mathematica's outline of changes they planned to make)

**September 2008**—Cahalan again acting as a private citizen and not a representative of PPSS/ED on her own time, presented her critique of the 2004 published Third Follow-up report to the COE Board.

**November 2008**—Cahalan with permission from Dr. Goodwin presents her critique of the Third Follow up published report at the American Evaluation Association meetings

**November 2008**—In reaction to Cahalan's AEA presentation, Mathematica posts its *Fifth Follow-up Report* to the Mathematica web-site without obtaining ED approval

**November 2008**—PPSS is ordered to put the Mathematica *Fifth Follow-up Report* into the final clearance stage (Ex Sec Clearance) in an effort to get the report published before the end of the Administration. Cahalan sends formal memo to Dr. Goodwin—objecting that the report does not meet quality standards (Appendix L)

**November 2008**—In the Ex-Sec formal review process, OPE formally disapproves of publishing the report, IES does not comment and Budget service approves publishing the report

**December 2008**—Following the OPE formal disapproval the report is officially classified as "returned to PPSS for rewrite." Cahalan sends comments to Mathematica and has a long phone conversation with the President of Mathematica, Paul Decker. In this conversation Dr. Decker asked PPSS to write down the major changes and indicated he would consider them. Following this conversation, Cahalan drafted a memo to Dr. Decker summarizing the recommended changes (Appendix M-2) ---As noted below the memo was never sent as permission was denied by OPEPD leadership.



**December 19 2008** Cahalan asked for and was granted a meeting with Assistant Secretary Evers and his Deputy Diana Perez where she presented a power point presentation on her concerns with the report. At this overtly friendly meeting she was told that henceforth all communication with Mathematica had to go through the front office (Diana Perez) and she was asked not to communicate with Mathematica directly. The reason given for this to Cahalan at the time had to do with the fact that in November of 2008 Mathematica had gone ahead and posted their draft *Fifth Follow-up Report* on their own web-site without having obtained ED approval to do so. As noted this action was taken by Mathematica in response to Cahalan being given permission by Dr. Goodwin to present results of her critical re-analysis of the published Mathematica *Third Follow Up Report* data at the American Evaluation Association Meeting in early November. Cahalan was told that ED legal advisors were looking into whether Mathematica had violated the contract. Cahalan was actually in favor of Mathematica being allowed to post their results to own website, but objected to ED publishing.

**December 30 2008.** Cahalan sent a memo to Diana Perez asking for permission to send the memo to Mathematica President Paul Decker concerning the recommended changes to the report. This permission was not granted and the memo was not sent (see email in appendix M-1).

**January 2009** Dr. Goodwin was informed by Ze've Wurman (the PPSS liaison with Assistant Secretary Bill Evers office) that Mathematica had made changes to the report that were acceptable to IES and to the OPEPD Assistant Secretary Bill Evers. He then asked Dr. Goodwin to post the report to the PPSS website immediately with an expedited editorial clearance so that the report could get out by the end of the Administration. After Executive Secretarial Approval was granted along with several other reports in the final days of the Administration, the report was posted to the PPSS website in the last week of the Bush Administration. Dr. Goodwin and Dr. Cahalan subsequently found that no substantial changes had been made to the report to address their concerns and PPSS never received any documentation of these final negotiations between the OPEPD front office (Ze've Wurman), IES, and Mathematica. PPSS staff did not even know they were taking place until they were ordered to publish the report. PPSS staff and OPE's objections were not addressed. (Appendix N contains the last email that Dr. Cahalan sent to Assistant Secretary Evers on January 13 requesting that he re-consider his decision to publish the report)

**February 2009--** OPE career staff discovered that the report had been published. OPE was not informed of the irregular posting of the report over their formal written disapproval. Normally reports are not posted until all of the POC offices have given approval and have been informed of how their comments have been addressed or if not addressed a justification memo is provided for why they have not been addressed. See email from OPE staff person sent to Dr. Cahalan requesting information about who decided to post the report over OPE's disapproval (appendix O)

## Appendix H

### Email and Table from the Fourth Follow-up concerning project 69 that took place between PPSS staff and Mathematica in March of 2005

-----Original Message-----

From: Goodwin, David  
Sent: Thursday, March 03, 2005 12:10 PM  
To: Cahalan, Margaret  
Subject: FW: Tables for Friday's meeting

Do you want to participate in a conference call on Friday at 2 to discuss the attached. Apparently, there is one project in the UB study that carries a great deal of weight and if excluded from the study, would alter the results.

-----Original Message-----

From: Allen Schirm [mailto:ASchirm@Mathematica-Mpr.com]  
Sent: Thursday, March 03, 2005 10:57 AM  
To: Goodwin, David  
Cc: Elizabeth Stuart; Robert Olsen  
Subject: Tables for Friday's meeting

David,

Attached please find draft tables to be discussed at our 2:00 meeting on Friday. Although we'd love to see you, I don't think that we need to meet face to face. So, if you want to save the trip across the street or the security hassles, etc. of our coming to your place, we can just have a conference call. If that's your pleasure, shall we call you in your office?

Allen

<<Project69.xls>>

---

Allen L. Schirm  
Associate Director and Senior Fellow  
Mathematica Policy Research  
600 Maryland Avenue, SW, Suite 550  
Washington, DC 20024-2512  
202-484-4686 (voice)  
202-863-1763 (fax)  
aschirm@mathematica-mpr.com

**Appendix H-1 Table Referenced in Email above from 2005**

Table H.1							
Comparison of Students in and out of Project 69							
Includes All Self-Reported Enrollment and Completion							
	Treatment Group Means				Control Group Means		
	In 69	All Others	Difference		In 69	All Others	Difference
<b>Postsecondary Enrollment (%)</b>							
Any postsecondary institution	74	78	-4		88	75	13 ***
Highest level of schooling attended							
Four-year college or university	52	51	1		63	45	18 ***
Two-year college	15	22	-7 ***		23	25	-2
Vocational postsecondary institution	7	5	2 *		2	5	-3 ***
<b>College Selectivity (%)</b>							
Most selective four-year college or university							
More selective	6	10	-4 **		4	7	-3 ***
Less selective	26	33	-7 ***		51	30	21 ***
<b>Any Financial Aid Received (%)</b>	54	70	-17 ***		73	65	8 *
<b>Postsecondary Credits Earned (mean)</b>							
Two- and four-year colleges and universities	41	51	-11 **		61	34	26 ***
Four-year colleges and universities	35	39	-4		49	28	21 ***
Two-year colleges	6	12	-6 ***		12	10	2
<b>Postsecondary Completion (%)</b>							
Any degree, certificate, or license	37	33	4		36	28	8 ***
Highest degree, certificate, or license earned							
Bachelor's degree or higher	20	15	5 **		25	11	14 ***
Associate's degree	0	6	-6		5	8	-3 ***
Certificate or license	17	12	5 ***		6	9	-3 ***
Source: ub4analysis.log							
Note: Means calculated using weights that adjust for differential sampling probabilities and nonresponse to the fourth follow-up survey (see Appendix A for more details).							
*/**/** Impact estimate is statistically significant at the 0.10 / 0.05 / 0.01 level.							

## Appendix I-1

*In April of 2008, Dr. Goodwin arranged for Dr. Cabalan to brief ED staff from OPE and Budget Service on her findings with regard to the UB Evaluation. OPEPD political appointee staff was invited but did not attend. In May of 2008, Dr. Cabalan drafted an email to Assistant Secretary Evers (below), asking permission to give the same briefing to Congressional staff and stakeholders to the evaluation. At Dr. Goodwin's request Dr. Cabalan first sent the draft of the memo to Dr. Alan Ginsburg, Director of PPSS. After consultation with Dr. Ginsburg who advised her to wait until after the IES reviews were completed, Cabalan decided not to send the email.*

*In early June Dr. Goodwin and Dr. Cabalan learned of the decision by IES leadership and Assistant Secretary Evers that Cabalan's report would not be considered for publication and that only the Mathematica paper would be IES externally reviewed for publication. This meant that the alternative Cabalan paper would have no chance of being published by ED. Following this decision not to review her paper, Dr. Cabalan made the decision to move ahead without asking permission from her supervisor's or the Assistant Secretary to inform the COE stakeholders of what she found. In mid-June she met with COE and on June 18 after the fact she sent an email to ED colleagues informing them that she had briefed COE (see Appendix G-3).*

**From:** Cahalan, Margaret

**Sent:** Thursday, May 22, 2008 11:20 AM

**To:** Ginsburg, Alan

**Cc:** Goodwin, David

**Subject:** FW: Request for permission to brief stakeholders and congressional staff

Hello Alan,

The last document on this email, also printed below, is a request addressed to Bill Evers to be given permission to brief those who have a stake in this evaluation concerning the current status of the evaluation. I gave a similar briefing to ED staff in April. David has suggested that I consult with you before I send the request to Bill Evers.

Thank you,  
Maggie

## Appendix I-2

**MEMORANDUM (This memo was not sent to Assistant Secretary Evers. After Cahalan learned of the refusal of IES to review her paper to be considered for publication by ED as well as the Mathematica draft report, she went ahead and briefed COE on her own time ---- not acting as a representative of the Department of Education )**

To: Bill Evers

CC: Ze've Wurman  
David Goodwin  
Alan Ginsburg  
Sandra Furey

From: Margaret Cahalan

Subject: Request for permission to brief Congressional staff responsible for Re- Authorization of Higher Educational Act (HEA) and Upward Bound evaluation subjects of concerning status of the evaluation

Date: May 22 2008

As the Secondary Postsecondary Cross Cutting (SPCC) Team Leader within PPSS and as the last Contracting Officers Representative (COR) and technical monitor for the study, this memo is written to formally ask your permission, as the Assistant Secretary responsible for OPEPD, to brief the subjects of the National Evaluation of Upward Bound as represented by the Council of Opportunity in Education (COE) and Congressional staff responsible for re-authorization work on the Higher Education Act concerning the current status of the evaluation. This briefing would be similar to the update on the evaluation already provided to US Department of Education staff by PPSS in April of 2008.

As a professional working in the area of program evaluation, I believe I am bound to try to meet the ANSI approved Program Evaluation Standards. Standard P6 on disclosure of findings states;

*P6 The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and others with expressed legal rights to receive the results.*

I respect that you have acted responsibly in directing PPSS to have the Mathematica Fifth Follow up Report reviewed by an external reviewer and to also have my paper so reviewed as a precursor to any future ED publications, and I am not asking for permission to publish my paper. However, this is not a simple case of two independent academic analyses of the same data reaching different conclusions. My paper was written as a QA paper in reference to results already published by ED concerning this program in 2004; and to summarize concerns that I have repeatedly raised with the contractor, as the team leader and technical monitor for the study concerning the Fourth and Fifth follow up draft reports that I was asked to review. It updates a paper originally drafted in 2006, about which ED staff was also briefed.

As the attached documents indicate my concerns about the conclusions published concerning the program in the Third Follow Up report (2004) date from 2005 when the contractor for the study, first sent ED a tabulation that showed that the results for the unpublished Fourth follow up were sensitive to the inclusion of only one project, with 26 percent of the weight. Gradually over the period since 2006, I have obtained the data files from Mathematica and done extensive QA review of the data files. This recently obtained information included the original sample frame, which although requested earlier, was only delivered to ED in December of 2007 after the contract officially ended in November of 2007.

The Mathematica findings, presenting one perspective on the UB program, have been widely shared and were last published by ED in 2004. They have been used to declare the program “ineffective” in the PART rating system and have been used to justify the Administration recommending zero funding in FY05 and FY06. They have also been used to justify significant changes in the regulations governing the program. These actions by OMB and the Administration have been justified by the lack of effect results from this study, known to be sensitive to only one project at least since analyses done prior to the Fourth follow up report preparation in 2005.

As the SPCC team leader and COR for the study, I have repeatedly questioned my own ethical judgment in remaining publicly silent on these issues to stakeholders and to Congress for three years, during which time there were intense public debates on related issues in Congress. In my judgment, as soon as it was recognized that the results were dependent on one project, a statement should have been issued that the results were sensitive to only one project. In addition it has also been demonstrated in 2006 and again in 2008 by the internal papers I have prepared, that results are also sensitive to the use of administrative data to supplement survey data, model specification, poststratification of the weights, and standardization by expected high school graduation year of outcome measures (see attached internal papers from 2006 and 2008).

I do not believe that it is in the best interest of our common goal of working for the mission of the US Department of Education, to delay any further in making those responsible for the re-authorization of the Higher Education Act in our Congress and the subjects of the evaluation aware of these serious concerns, which have been raised over a period of more than two years, by the COR formally entrusted with the responsibility for technical monitoring of the study.

Respectfully,  
*Margaret Cabalan*

Margaret Cahalan, National Evaluation of Upward Bound COR.

## Appendix J

After meeting with Maureen Hoyer of COE in mid-June of 2008 to inform her, as the stakeholder representative of her QA findings with regard to the UB evaluation, Cahalan sent the email below. In the email she informs her colleagues and her supervisors of what she had done and her reasons for taking this step.

**From:** Cahalan, Margaret  
**Sent:** Thursday, June 26, 2008 11:59 AM  
**To:** Goodwin, David; Bergeron, Frances; Byrd-Johnson, Linda; Clement, John; Ginsburg, Alan; Bergeron, David; Anderson, Judith; Cichowski, Carol  
**Cc:** Wurman, Ze'ev  
**Subject:** FW: Status of Upward Bound Evaluation

Dear Colleagues,

I'm writing this email to bring you up to date on certain steps I have taken with regard to the UB evaluation. As you know, ED decided to move forward with an IES external review of the draft Mathematica Fifth Follow-up report and decided not to also move forward with a review of the paper I prepared. My paper will only be considered as comments on the Mathematica report in the review process. While I respect that the Department has the authority to determine what is considered for publication by ED, this decision presents a serious ethical dilemma for me. As you know my concerns about the published results from the study are long standing and stem from spring of 2005, and have become more serious as I have gained access to the data and completed QA analyses. As a government employee responsible for monitoring the contract and also as a professional working in the area of program evaluation, I believe there is a long overdue ---ethical obligation to inform those who have been affected by this evaluation of any concerns with published results in a timely manner. I believe I am bound to meet the ANSI approved *Joint Committee on Program Evaluation Standards* that in standard P6 state: "The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and others with expressed legal rights to receive the results."

For these reasons I have revised the paper I presented to you in April (and revised in May in response to Mathematica comments) to remove all use data/discussion of findings from the 5<sup>th</sup> follow-up survey which is now under IES review. This latest June version of the paper has the entire section on degree-attainment removed. The paper now focuses only on early enrollment and financial aid application in +1 and +4 years and does not consider the results of the 7 to 9 year period covered in the Fifth Follow-up. The paper is now only an alternative secondary analysis dealing with previously published conclusions put forth in Mathematica's 2004 Third Follow up report.

To fulfill my obligation to inform stakeholders of information known to me concerning the 2004 published results of this evaluation, I have met with a representative of a national organization representing the subjects of this evaluation and shared my revised paper. I have also informed her that the Fifth follow-up report is under review. A copy of this June version of my paper is attached for your information.

Respectfully,  
*Maggie Cahalan*

Maggie Cahalan, COR UB Evaluation

## Appendix K

This memo was sent by David Goodwin in September of 2008 to Allen Schirm the study Mathematica Project Director at the time and Mary Moore, Vice President and Director of the DC Office

September 25, 2008

To: Allen Schirm and Mary Moore  
Mathematica Policy Research

From: David Goodwin  
PPSS/OPEPD

Subject: Revisions to Upward Bound Evaluation Final Report

In response to your memo of 9/12/08, I am writing to recommend that you proceed to revise the report. Given that some of your responses are fairly general, I'll review the report when it is received, and at that point, decide whether the report should be submitted for POC review leading to publication by the Department of Education. I'm sure we all agree that it's long past time this study was concluded.

The changes outlined in your memo did not, for the most part, accept most of my recommendations or Maggie's. However, that need not prevent PPSS from submitting the revised report for POC review. Although the recommendation of the COR is for ED not to publish the report with the changes outlined, I may, reluctantly, choose to reject this recommendation upon examining the revised report.

In the remainder of this memo, I'd like to make some comments about how you responded to recommended changes, and my own personal view of the issues that have been raised regarding the draft final report. I make three basic points: 1) standardization for expected date of high school graduation by itself could change the findings; 2) the sample design was seriously flawed and something other than a tepid warning is warranted; and 3) Project 69 may or may not be representative of its stratum (I think it is not), but when joined with the weighting issue becomes more than just a "bad draw."

First, it should be clear that I have no agenda in the outcome of this debate between and Maggie Cahalan and Mathematica. As a former COR for this study, I approved and defended earlier reports concluding that the Upward Bound program had no overall impact on its participants. Absent an alternative analysis, I would have almost certainly supported Mathematica's final report. However, having seen Cahalan's reanalysis of study data, I became convinced that the Mathematica findings are considerably less than robust, at a minimum, and quite possibly incorrect.

**Standardization.** One issue concerns the standardization of time to compare students with similar dates of expected high school graduation. Only one outside reviewer (B) addressed this issue; his/her comments suggested that standardization was a valid approach but due to randomization, unlikely to make much difference in results. However, Cahalan's analysis suggests that it very likely does change results.



## Appendix K (September 2008 Memo from David Goodwin to Mathematica-continued)

Table 6 (attached) shows a variety of analyses.<sup>10</sup> The ITT analysis for the entire sample is the core analysis. Initially, Cahalan replicates (absent the specific code her estimates are fairly close to yours --1 percentage point difference versus 3 percentage points in your report) using only survey responses. She then introduces several changes, none of which seem unreasonable. When standardizing for expected high school graduation, using all survey as well as administrative records, the impact estimates on postsecondary enrollment are statistically significant. Without even addressing any of the other issues (sample weighting, project 69, etc.) these results by themselves should be enough to raise serious questions about the robustness of the Mathematica reported results.

I was disappointed in your memo's response on the issue of standardization. You indicated that such analysis "was never specified in prior statements of work..." Frankly, I never thought I'd hear that defense on a major study from a company with the reputation of Mathematica. You are better than that, and after \$14M in contracts for this study should not be hiding behind such a flimsy excuse.

**Sample Design.** A second issue concerns the sample design, and specifically the extreme weight accorded to one project. All reviewers concluded that the sample design was a major problem, but did not agree on what to do about it. Although post-hoc changes are certainly undesirable, statistical adjustments may be necessary as the only way to mitigate serious design flaws. I previously suggested that impact estimates be produced for both the entire sample and the 74 percent that could be generalized to the national population. You reject that recommendation arguing, in part, that there are other projects with "larger than average weights." Of course there are other "above average" weighted projects-- it's an average. However, the attached Figure 1 clearly demonstrates the extreme weighting accorded one particular project. This is not a matter of modest variation in project weights across the sample.

Another approach, strongly implied by the IES Action Editor, is to present unweighted estimates for the entire sample. She states, "I would also recommend that the report clarify upfront that national population estimates of program impacts for any of the other subgroups are not affected by the sampling limitations, and *all of the study findings are robust when interpreted in the context of the study sample, not as national population estimates* (my emphasis)." The phrase "in the context of the study sample" would seem to mean without national weights attached. I'd suggest that you consider presenting unweighted as well as weighted estimates for the entire sample. I'm aware of your concern with interpreting results because of the way that certain groups were overrepresented in the sample. I just don't see a perfect solution and would rather err on the side of providing readers with more information—not less.

**Project 69 Outlier.** Finally, there is the question about whether Project 69 is truly an outlier that biases the results. The reviewers did not think it clearly did, and generally did not think this was a major problem because of either random assignment or the regression adjustments used in analysis; in any case, this issue was less important to them than the sample weights. I disagree. There are two issues of concern—whether Project 69 adequately represents its stratum, and the other is whether treatment-control group differences are a source of bias. I think there is reason to believe that Project 69 is not representative of a 4-year college stratum. It had previously been a community college and had been reclassified as a four-year institution when taken over by another college. It's Upward Bound project operated more like one at a community college in that it did not offer a summer residential experience. Whether or not this project was just a "bad draw" would not be an issue if it did not also carry such extreme weight.

---

<sup>10</sup> Table 6 applies to data collected in the third follow-up. It's conceivable that such analysis applied to data in the 4<sup>th</sup> and 5<sup>th</sup> follow-up surveys could result in different findings.

## **Appendix K (September 2008 Memo from David Goodwin to Mathematica-continued)**

On the issue of bias, it's clear that the treatment and control groups are well matched on 9<sup>th</sup> grade academic risk indicators, educational expectations and grade distribution without project 69, but much less so when this project is included. (Figure II). Although your analysis includes controls for some of the differences (but appropriately, not the academic risk indicators), I believe that those controls, coupled with the weights, are inadequate. Treatment-control differences in particular projects can occur despite random assignment and are accepted when weights are relatively evenly distributed, but that is not the case here.

Nothing I've said is new or likely to resolve any current disagreement. The points simply represent my personal views. I am NOT saying that the Mathematica report is unworthy or fatally flawed. With the exception of the sample design, the external reviewers mostly seemed to believe that the study procedures were consistent with what is generally understood to be acceptable evaluation practice. That's not to say, however, that there's only one valid approach, and that another equally plausible analysis could not reach different conclusions. I believe that is the situation here, and therefore, urge Mathematica to be far more circumspect about the strength of evidence supporting its findings than has been the case so far.

One other issue is how the report characterized the selection of the highly stratified sample design with which we ended up. The report states "After considering alternative designs and weighing their relative strengths the Department of Education chose a design in which selection probabilities varied substantially across strata... (P.20)" That statement is self-serving on the part of Mathematica, for it ignores the fact that Mathematica never told ED that such a sample design was problematic, never raised the topic at a TWG meeting, and never published the sample weights in its reports. Mistakes were made in the sample design from which we've all learned, but the choice of a sample design was not as one-sided as the report indicates.

Finally, despite all the disagreements associated with this report, I remain highly respectful of the professional skill and integrity of the Mathematica organization, and in particular, the individuals who have worked on this project. I would look forward to working with Mathematica on future projects.

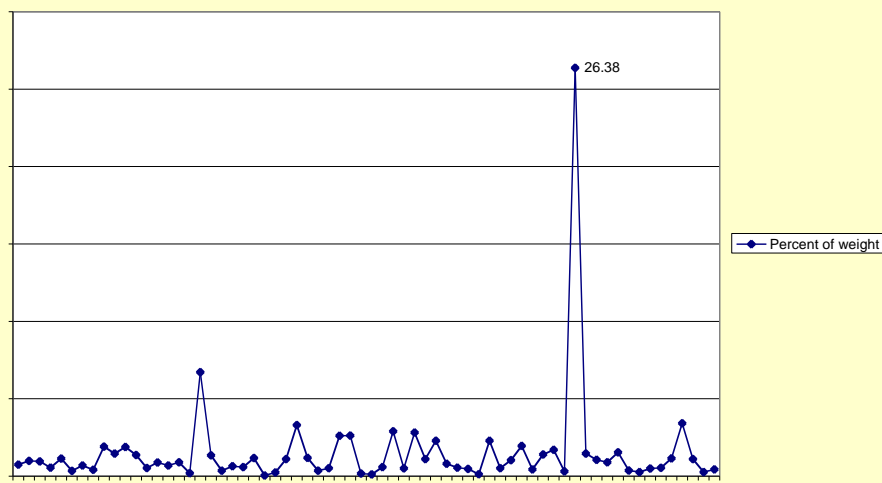
Cc: Maggie Cahalan

**Appendix K (September 2008 Memo from David Goodwin to Mathematica--continued)**

<b>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</b>				
	<b>All sampling strata</b>		<b>One project with bias removed (the remainder represents 74 percent of Horizons waiting list)</b>	
	<b>Given Opportunity (ITT)</b>	<b>Participated in UB/UBMS (TOT/CACE)</b>	<b>Given Opportunity (ITT)</b>	<b>Participated in UB/UBMS (TOT/CACE)</b>
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 f3wgt su (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 f3wgt su 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 f3wgt su	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><b>NOTE:</b> Please see table 5 for detailed notes</p> <p><b>SOURCE:</b> 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>				

Appendix K (September 2008 Memo from David Goodwin to Mathematica-continued)

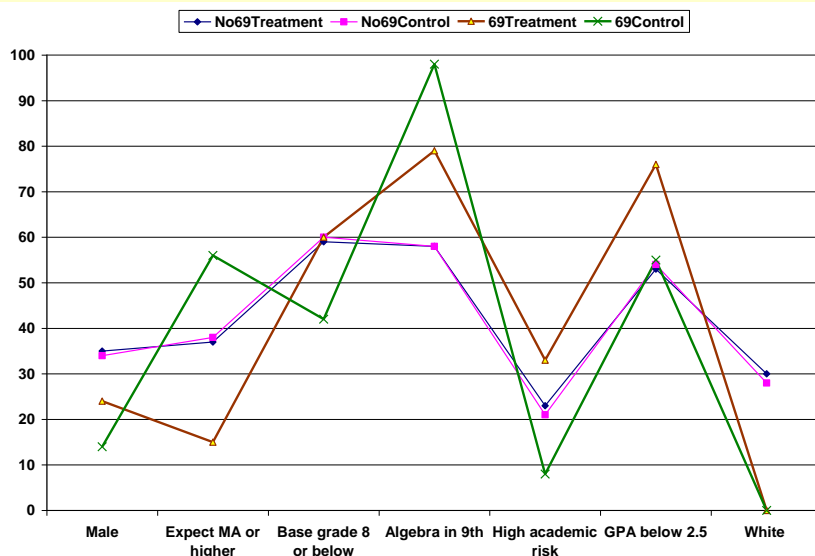
Figure I. Percentage of sum of the weights by project of the 67 projects making up the study sample: National Evaluation of Upward Bound, study conducted 1992-93-2003-04



**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.

Figure II. 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



## Appendix L

*In November of 2008, Mathematica submitted a revised draft of the Fifth Follow-up report. This revised draft did not address the major issues with the report that PPSS had raised consistently and that were re-iterated in memo's to Mathematica from Dr. Goodwin and Dr. Cahalan. In this November 2008 memo after reviewing the revised draft, Cahalan recommends that ED no longer consider the report for publication, but allow Mathematica to publish their report on their own website. Cahalan also objects to putting the report into the Ex. Sec final review process. Cahalan was overridden in her objections by the Assistant Secretary's representative and the report was placed into Ex. Sec review in mid-November*

Memo—November 10, 2008

**To: David Goodwin, Director PAS**

**From: Maggie Cahalan, National Evaluation of Upward Bound Technical Monitor - COR**

**Subject: Fifth Follow-Up Report**

I have reviewed the November 2008 revised Fifth Follow-up Report which is the final report for the study; and the Mathematica memo summarizing the changes they made in this revision. As you know these revisions were made in response to your memo of August 2008 following the IES external review and my memo also of August 2008 summarizing remaining technical issues. These August 2008 PPSS requests for revisions were fully consistent with changes that had also been recommended earlier on the basis of PPSS review in September of 2007, and April of 2008. While, Mathematica has included some additional qualification statements and it's clear the report represents a substantial effort; the changes made are not responsive to the specific changes consistently recommended by PPSS over a period of more than 2-years. More importantly **the report, as revised, does not meet known standards for statistical or evaluation research; and very importantly contains incorrect and misleading conclusions concerning the Upward Bound program.**

Therefore sadly, at this time, as the Team Leader for the Secondary Postsecondary Cross-Cutting Team (SPCC) and the COR-Technical Monitor for the National Evaluation of Upward Bound **I am writing to formally recommend that the Fifth Follow up report as first prepared by Mathematica in August of 2007 and revised most recently in November of 2008, not be published by ED and not be sent out of PPSS for Ex Sec review. For ED to publish this report, knowing that the conclusions are not justified would be highly irresponsible.**

I recommend that Mathematica be given permission to self-publish their report, if they so choose, after the data files have been released by ED for restricted use---but that ED not pursue further attempts to make this report an official ED publication. Over the course of more than 2-years, Mathematica has been given consistent recommendations for needed changes and most recently these recommendations were made by you, yourself in your memo following the external review. These were in the form of "requirements" for change from PPSS. These recommendations and requirements are consistent with the comments of the IES external blind reviewers; and as importantly they are fully consistent with commonly accepted standards for evaluation, and for statistical and scientific research.

Mathematica has been given numerous opportunities to revise their report and they have repeatedly chosen to disregard these recommendations; and indeed have prepared a report that seems primarily designed to defend past mistakes. They are well aware that the consistency of their conclusions are dependent on maintaining some of the flawed procedures that they have followed in the past with regard to this influential and controversial study; and this may explain their frankly shocking refusal to implement PPSS recommendations. I do not know their motivation. I do know that over the course of three years, I have consulted with numerous statistical experts, and extensively reviewed the data from this study. These consultations and analyses have made me aware that the Mathematica conclusions made on the basis of this study, are seriously flawed. Therefore I cannot recommend publication of the report and I wish to formally stand in the way as much as possible.

As you know, these same misleading conclusions have been published in the Third Follow Up report and were in the unpublished Fourth Follow Up Report. They have had very serious consequences for the program resulting in the PART “ineffective” rating—something received by only 3 percent of the over 1000 programs that have been PARTED by the federal government. In this case the PART rating resulted in zero funding recommendations, and an intense series of debates in Congress concerning the program and evaluation methods. These debates have resulted in the cancellation of a new study and an ED Absolute Priority; and a prohibition against the random assignment recruitment methods used in the evaluation.

To continue to publish reports that are misleading concerning a program, we are evaluating is inexcusable on PPSS’s part and is a serious violation of the AERA standards that the conclusions be warranted and transparent. It also violates the Propriety standards of Joint Program Evaluation Standards. Ethically, I cannot be a party to this being done for a study for which I was given the responsibility to be the Technical Monitor.

The contractor, Mathematica Policy Research, has had the technical issues pointed out to them since 2005 when the Fourth Follow up report was reviewed and throughout the period of both preparation and review of the Fifth Follow up report. The problems were also detailed in written comments by Jay Noell and myself to the first draft of the Fifth Follow-up report in September of 2007 and orally in December of 2007.

The contractor was not transparent in failure to reveal the weighting design issues in the Third follow up published report. Mathematica also failed to ever inform ED of the representational issues with project 69, and delayed, despite repeated requests in sending the sample frame until after the contract had ended in December of 2007. As you know, these issues were discovered by PPSS itself ---after the contract ended.

I have detailed the problems with the Mathematica report numerous times, but do so again below with a specific focus on research standards violations.

- **Everyone who has examined the sample design has noted it has serious flaws. It does not meet the fundamental design principal when probability of selection weights are to be applied that there must be more than one member of each strata; and seriously has an unacceptable end stage weighting with one project accounting for 26 percent of the weight.** External reviewers have noted the national estimates are not robust under these circumstances. For more than 2 years, PPSS has been recommending that the results be reported weighted and unweighted and with and without the outlier 26 percent weight project. Mathematica has repeatedly refused to implement this requirement.

- **The analyses presented is not transparent in presenting information concerning the serious representational issues with project 69;** and indeed appendix G is outright misleading in its discussion of project 69.
- **The analyses inappropriately use National Student Clearinghouse data for their main postsecondary entrance estimates when it is known to have only 25 percent coverage and known to have lack of coverage for project 69 in the applicable time frame introducing bias in the estimates reported—a clear violation of NCES standards on coverage and bias**
- **The analyses use imprecise outcome measures that are un-standardized by expected high school graduation year despite evidence of treatment control group non-equivalency on this distribution.** This is also a violation of commonly accepted standards for research. It is similar to comparing student scores in the first grade with those in the fourth grade—something that just should not be done.
- **QA analyses have repeatedly shown that results are sensitive to the inclusion or not of only one program out of the 67 sampled and that this project has very serious bias in favor of the control group.** It is a clear violation of scientific standards not to make this fact very clear in reporting findings. The results reported by Mathematica are not transparent and to make them transparent it must be stated that the conclusions are dependent only on this one outlier weight project of questionable representation, and with evidence of serious bias.
- **The report contains no actual results from the models run and does not describe adequately the procedures followed in arriving at the estimates.** I have attempted to replicate some of the subgroup findings using the methods as described in the report and have been unable to do so.
- **Its clear that some of the numerous comparisons with sub-groups are based on insufficient cell sizes—this is never acknowledged**

There are other issues with this report that have been commented on in earlier comments and that remain unaddressed.

Thank you,  
Maggie Cahalan

## Appendix M-1

**December 30 2008 Email Exchange between Dr. Cahalan and Deputy Assistant Secretary Diana Perez in which Cahalan is refused permission to send the a memo to Paul Decker, President of Mathematica on Recommended Changes (see Appendix M-2). This draft memo was prepared by Cahalan following the December OPE Disapproval of the Fifth Follow-up Report in the POC Review Process**

**From:** Cahalan, Margaret  
**Sent:** Tuesday, December 30, 2008 11:08 AM  
**To:** Perez, Diana  
**Cc:** Wurman, Ze'ev; Evers, Bill; Goodwin, David; Ginsburg, Alan  
**Subject:** Communication with Mathematica about UB report

Hi Diana,

As you know, the report *The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation (Referred to as the Fifth Follow Up Report)*, prepared by Mathematica Policy Research was reassigned in the OS Exec Sec tracking system on December 15<sup>th</sup> to "PPSS for rewrite" following the OPE disapproval in the POC review. Prior to the meeting I had with you and Bill, as is standard procedure, I had sent the OPE comments to Mathematica and subsequently I had a conversation with Paul Decker the President of Mathematica about the report. At the end of the conversation, he indicated that ED should write down the specific recommendations for revision and Mathematica would review them. I have now done this, and would like to send these recommendations to Mathematica as the COR for the contract.

You indicated at our meeting that we should coordinate all communication with Mathematica through you, so I am sending this memo to you before sending it out. Please let me know if it is ok to send this to Mathematica.

Thanks, Hope your holidays are going well,  
Maggie

**From:** Perez, Diana  
**Sent:** Tuesday, December 30, 2008 1:54 PM  
**To:** Cahalan, Margaret  
**Cc:** Wurman, Ze'ev; Evers, Bill; Goodwin, David; Ginsburg, Alan  
**Subject:** RE: Communication with Mathematica about UB report

Good Afternoon Margaret,

Per our discussion, all communications with Mathematica should be done through the OPEPD front office. Thank you for forwarding the memo. If appropriate, the OPEPD front office will send all comments and changes to Mathematica once we have finished our review of the report and all comments.

Hope your holidays are going well,  
Diana



## Appendix M-2

**December 30-2008 Draft Memo to Paul Decker President of Mathematica. Cahalan was denied permission by the OPEPD front office to send this memo in the email above (appendix M)**

This memo outlines PPSS and OPE recommended changes following OPE disapproval of publishing the report in the ED Ex Sec Review process (the memo was never sent). The Fifth Follow –up report was published two-weeks later disregarding the PPSS monitoring staff objections and formal OPE disapproval. It was reported to PPSS staff (by Ze’ve Wurman, consultant to Bill Evers and the PPSS liaison with OPEPD front office) that IES had signed off on the report although any documentation of this “sign-off” was never provided to PPSS staff.

### **Memorandum (draft never sent because permission denied by OPEPD Assistant Secretary’s political appointee staff in late December 2008)**

**To:** Paul Decker, President Mathematica Policy Research

**CC:** Allen Schim, Don Lara, David Goodwin, Alan Ginsburg, Diana Perez, Bill Evers, Ze’ev Wurman, Cynthia Duncan, Carol Matamora

**From:** Margaret Cahalan, Technical Monitor for the National Evaluation of Upward Bound

**Date:** 12/30/2008

**Re:** *Report: The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation (Referred to as the Fifth Follow Up Report)*, prepared by Mathematica Policy Research

---

Thank you for taking time from your busy schedule to discuss the report with me on the phone last week following the “disapproval to publish” recommendation by the Office of Postsecondary Education (OPE) in the POC review. The official status of the document is now “returned to PPSS for rewrite.” At that time you indicated that if PPSS had specific recommendations for changes at this time, we should write them down and Mathematica would consider them. This memo is in response to that invitation.

I respect, as you have noted in our call, that as a company you are highly committed to “objective” evaluation research and are a recognized leader in the area of random assignment studies in education. I also recognize that this study was very difficult to conduct and that the UB sample design was trying to be responsive to various often contradictory study goals such as the need for national estimates and the desire to represent relatively uncommon types of projects. To that end we offer the following recommendations that we hope will satisfy both ED’s need for an accurate and fair evaluation of the Upward Bound program, and Mathematica’s need for objective research.

These recommendations are based on PPSS QA review of the data files and report, the IES external reviewers comments, and the recent POC review comments of OPE. As was agreed upon in the last contract modification for the study in 2007, and given that the contract has been over for more than a year and there are no more funds available for Mathematica to “rewrite,” PPSS is willing to make the recommended changes to the 5<sup>th</sup> follow up report, and submit them for Mathematica oversight review, before re-submitting the report into ED POC review.

1. *Given the sample design flaws recognized by PPSS, Mathematica, and the IES external reviewers and OPE reviewers, PPSS recommends that the results be presented and the report discuss conclusions on the basis of models run with and without project 69 and weighted and unweighted.* The IES reviewers have noted that the national weighted estimates with project 69 cannot be considered to be robust, but that the estimates without project 69 representing 74 percent of the sample can be considered more robust. They have also noted that the estimates for the sample considered to represent the 67 projects in the sample only (implied unweighted) could also be considered more robust. *Therefore PPSS has concluded that the best way to proceed is to include the results with and without project 69 and weighted and unweighted with clear labels. Discussion of conclusions should also be updated to include the results for the 74 percent of the sample not represented by 69 and also the unweighted estimates.* As was recommended by one of the IES external reviewers, we also recommend that TOT estimates be included in the same tables as the ITT estimates. As the OPE reviewer also noted, given the control group UBMS crossover and treatment group dropout issues faced by this study, it is important that the TOT estimates be discussed.
2. *PPSS recommends that, given the 4-5 year spread in high school graduation year among the sample, that outcome measures be carefully standardized to expected high school graduation year (EHSGY) and that all applicable follow up surveys be examined to obtain the exact dates of any postsecondary entrance and of any degree or credential awards.* The potential validity of the standardization of outcomes as a concern in any study was recognized by two of the IES external reviewers, although one noted it should not be a concern in a random assignment study. However, given that this was not a simple random sample but one with 339 strata not based on EHSGY, and that we have evidence that the control group was indeed older than the treatment group on average, there is a need to standardize the outcome measures if this bias is to be controlled. Given that this is the final report from the study, and that this standardization was not done in the past published reports, we also recommend that some results for postsecondary entry be included for earlier points in time after high school that are more

relevant than the 7 to 9 year period for examination of the impacts of a high school program. This might be for example +1 +4 +6.

One concern with regard to the 5<sup>th</sup> follow up time frame (7 to 9 years out), especially for BA degree completion is that not all of the sample members had reached 9 years or even 7 years by the time of the 5<sup>th</sup> follow-up. Based on survey information on EHSGY--- about 10 percent had only reached 6 years; 30 percent 7 years; 34 percent 8 years and 19 percent 9 years and 5 percent had reached 10 years by 2003-04 when the 5<sup>th</sup> follow up was conducted. This means that unlike the standardizations covering a date that everyone had reached ---6 years, it may be necessary to do the analyses only on those who had reached the same number of years out. As noted this would seem to be especially important for BA receipt, as we know from treatment control group distribution comparisons, the control group had a higher proportion of older students and non-standardized estimates do not take this difference into account.

3. *Consistent with NCES coverage standards ([http://nces.ed.gov/statprog/2002/std3\\_1.asp](http://nces.ed.gov/statprog/2002/std3_1.asp)), given that the National Student Clearinghouse (NSC) only began operating in 1994 and reports only 25 percent coverage in the period the UB sample was graduating high school and most likely to enter postsecondary, PPSS recommends that NSC data not be used for postsecondary entrance comparisons of treatment and control group. This recommendation is based on the 25 percent coverage in the period combined with the known potential bias issues due to UB clustering in grantee institutions and known lack of participation of project 69 with 26 percent of weight. PPSS thinks that these coverage and bias issues outweigh the argument Mathematica has made that some postsecondary entrance of survey non-responders who did not apply for or get financial aid might be missed. PPSS does recommend that the National Student Clearinghouse data be used for BA award, but we do not recommend it for less than BA degrees due to known lack of participation in less than 4 year schools and bias issues. While the NSC data is becoming increasingly helpful in tracking students it continues to suffer for sporadic and incomplete coverage that can seriously mislead conclusions (Cunnington 2004; Roderick M, Nagaoka J, Allensworth 2006-- *From High School to the Future: A first look at Chicago Public School graduates' college enrollment, college preparation and graduation from fourth-year colleges*). PPSS recommends that the "applied for aid" data in federal aid files be also used to give evidence of postsecondary as it covers more students than the Pell award data and is not sensitive to differences in income levels.*
4. PPSS recommends that the appendix G discussion of project 69 be updated to acknowledge academic risk differences between treatment and control group and also differences in EHSGY and also to deal with the identified project 69 representation issues. In general a more transparent and informed treatment of 69 issues is needed.
5. The sub-group analyses discussion needs to reflect an awareness of some of the project 69 high weight and peculiar treatment/control group issues. For example, the Mathematica race/ethnicity sub-group analysis found positive impacts for whites but not for blacks in the sample. As none of the project 69 sample members are white; when

we look only at treatment control differences in the white sample we have, in effect, removed some of the bias introduced by the fact that project 69's control group has on average higher education expectations and higher academics than project 69's treatment group. Conversely when we only look at the Black sample sub-group, project 69 is contributing an even higher proportion of the weight and (treatment-control group non-equivalency) than it does to the overall sample.

A similar sub-group issue happens with regard to academic risk categorization sub-groups (top 80 percent and bottom 20 percent). The Mathematica analyses finds positive impacts for both the top 80 percent and the bottom 20 percent in postsecondary entrance; and, in a seeming paradox, does not detect impacts for the overall sample comprised of the top 80 and the bottom 20 percent in postsecondary entrance. In this case, the result is related to the fact that project 69 is contributing a disproportionate proportion in the bottom 20 percent on academic indicators from its treatment group. The two academic risk divided sub-groups thus each have a more balanced treatment-control group equivalency than does the overall sample with project 69 included.

7. It's clear that some of the numerous comparisons with sub-groups for degree attainment especially are based on insufficient cell sizes—this needs to be reviewed and those estimates with insufficient cell sizes removed from tables.
8. As OPE recommends, there needs to be more recognition of valid alternative analyses of the data.

Hopefully, these changes will meet both Mathematica and ED's shared goal of producing the most accurate and objective final report from this important study. As indicated in an earlier communication, ED is available to meet with Mathematica to discuss this plan to respond to the concerns that have been raised by the report reviewers from PPSS, IES external review and from OPE.

## Appendix N

### **Email of Dr. Cahalan of January 13 2009 the last week of the Bush Administration to Assistant Secretary Evers Requesting Re-Consideration of his Decision to Publish the Fifth Follow up Report. (The Report was published on January 15, 2009)**

**From:** Cahalan, Margaret  
**Sent:** Tuesday, January 13, 2009 10:45 AM  
**To:** Wurman, Ze'ev; Evers, Bill; Perez, Diana  
**Subject:** Request that you re-consider Upward Bound Report Publication Efforts

Dear Assistant Secretary Evers and staff,

As you know, I have had concerns about the published and unpublished Mathematica results since 2005, when the lead analyst for the study sent ED tables showing that the “no effect” results were sensitive to only one project and that this project carried 26 percent of the weight. Subsequent QA analyses I have done over the last three years has revealed to me that this project has serious treatment control group imbalances in expected high school graduation dates, in educational expectations and in academic risk factors in favor of the controls and is introducing bias into the conclusions reported by Mathematica. It also has serious representational issues. Over the course of more than a year PPSS has repeatedly recommended to Mathematica that the data be reported with and without the outlier project and weighted and unweighted; and that the outcome measures be standardized to date of high school graduation, and that the National Student Clearinghouse data not be used for postsecondary enrollment estimation due to having only 25 percent coverage in the applicable period and clustering of UB participants enrollment in grantee institutions. None of these recommendations, the validity of which were confirmed by the external review, have been implemented by Mathematica in the revised report submitted in November. For these reasons, I must object once again to your pushing the report through in your last days in the Department. The attached memo provides more detail as to my unmet concerns with the draft Mathematica report.

I believe that ED and Mathematica have an unfulfilled obligation to the subjects of the evaluation, who have had serious negative consequences due to this evaluation, to make them aware of the study issues and alternative results. To move forward and once again publish results that are unwarranted and biased, and to suppress results to the contrary about a program one has been given the public trust to evaluate seems to me to violate the basic standards of evaluation research and good government. For these reasons I believe it would be a serious mistake to publish the Mathematica report as it is currently written.

I recognize that you are going to justify your actions by saying you have obtained the IES reviewers sign off on the report. I will not comment on the appropriateness of your forbidding my communication with Mathematica and then going around PPSS to IES and secretly negotiating with Mathematica, except to say that in my 25 years in government work in which I have had many reports go through review processes, I have never seen one in which the concerns of the technical monitor and the office sponsoring the study, are not required to be addressed before publication. Typically reports are not even sent to external review until they have addressed the concerns of the technical reviewers within the sponsoring agencies office.

At this point I am very weary of this matter and would very much like to put it behind me. In all honesty it has been the most painful of my 25 year career as a researcher, but I do not think I can in good conscience just let it drop. I have once again attached a summary of my concerns about the report. I have also attached the memo I drafted to Mathematica following the POC review and OPE disapproval with a summary of recommended changes to address remaining concerns. As you may recall on December 30 you denied me permission to send this memo to Mathematica.

Good luck to you in your future endeavors,

Maggie Cahalan

**Appendix O**

**February 24 2009 --Email from OPE Staff Member Requesting Information Concerning  
the Posting of the Report Over OPE's written objections**

**From:** Bergeron, Frances  
**Sent:** Tuesday, February 24, 2009 10:40 AM  
**To:** Cahalan, Margaret  
**Subject:** Final Report of UB Evaluation Released

Maggie:

I just found out today that the final report on the UB evaluation was released in January and has been posted to PPSS's web site. Do you know who made the decision to post the report over OPE's written objections.

Frances

## Appendix P

### Documentation of Modeling Results from the *Cahalan Re-Analysis Report*

*(Note this is Appendix B in the report from which it is excerpted)*

### **Examples of Detailed Model Results for Tables in the Body of Report and Comparison of Results when an Alternative Variable is Used for Standardization by EHSGY**

Appendix tables B1-B6 provide examples of the complete model results for the statistics reported in the body of the report. We also include some model results (labeled B1a to B4a) for an alternative variable used for establishing a “grade-year” reference for standardization by expected high school graduation year (EHSGY).

As observed in table 2 of the body of the report, treatment and control group non-equivalencies in grades reported on the study surveys indicated that there was a need to standardize outcomes relative to fixed time frames. Model results reported in the body of this paper are based on a standardization of Expected High School Graduation Year (EHSGY) based on a baseline survey variable (B1) present for 99 percent of the sample that read:

“What grade were you in during the LAST SCHOOL YEAR (1992-93 school year)?”

Because some students reportedly answered the question with reference to the 1991-92 school year instead of 1992-93, we included a correction in the tabulation that provided for a range going from -1 to +1 (or +4) of the year that was established on the basis of the grade reported on the baseline data file which ranged from grade 7 to a few in 11. As discussed in the body of the report and indicated in tables 2 and 3, estimates of EHSGY using different variables on the data files are not entirely consistent with each other when the baseline survey, first follow-up survey and the third follow-up survey data files are compared. For this reason, to check the models reported in the body of the text based on the baseline variable (B1) with the correction for the 1991-92 responders, we also calculated an alternate EHSGY based on the results to the first follow-up survey in which sample members were asked the following question (QA1):

“What grade (are you in/were you in during the 1993-94 school year) or (are/were) you not attending junior high or high school (now/then)?”

The models presented in appendix tables B1 to B4a show comparative results using the two alternative variables for tabulation of EHSGY. As can be observed in the tables, the two alternative bases for

standardization yield much the same impact estimates and significance test results. For example, table B1 and B1a present the same model with the two alternative variables used to standardize for the outcome of postsecondary entrance evidence by +1 of EHSGY. The impact estimate reported in table B-1 shows a 6.9 effect significant at the .004 level. These estimates are used in the body of the report. The alternative variable presented in table B1a shows a 6.7 effect significant at the .000 level for the same model. Results in tables B2 and B2a are for the same models as in B1 and B1a but exclude project 69. We see that the estimate of effect in B2 (and reported in table 6 in the body of the report) is 9.1 significant at the .000 level and the estimate in B2a using the alternative first follow-up variable for standardization shows an effect size of 12.7 and significance of .001. Tables B3 and B3a show instrumental variables two stage regression results for modeling TOT with the dependent variable of appearance on the aid file and show effect size of 9.3 significant at the .002 level for the baseline variable standardization; and show an effect size of 10.4 significant at the .001 level for the first follow- up variable used for standardization. Results in B4 and B4a for bachelor's degree receipt without project 69 show similar effects and significance levels when +7 instead of +6 is used with the baseline.



Table B-1. Intent to Treat (ITT) logistic regression results 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 (estimate reported in table 5 and 6)							
pr-T = 72.9 pr-C = 66.0 Difference = 6.9**** kenye2	Variable name	Coef.	Linearized Std. Err.	t	P> t	95% Confidence Interval	
FFUTC (random assigned to treatment)	Ffutc	0.395308	0.130178	3.04	0.004	0.131997	0.658619
Gr79293 (Grade 7 on baseline ref grade 9)	gr79293	0.165761	0.668542	0.25	0.805	-1.18649	1.518015
Gr89293 (Grade 8 on baseline ref grade 9)	gr89293	-0.80596	0.446533	-1.8	0.079	-1.70916	0.097236
Gr109293 (Grade 10 on baseline ref grade 9)	gr109293	0.223571	0.848204	0.26	0.793	-1.49208	1.939225
Gr119293 (Grade 11 on baseline ref grade 9)	gr119293	-1.2639	1.263728	-1	0.323	-3.82003	1.292227
Clowoy (Low income only)	Clowoy	0.189757	0.257292	0.74	0.465	-0.33066	0.710179
Cfgenoy (First generation only)	Cfgenoy	0.346913	0.212268	1.63	0.11	-0.08244	0.776266
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	-0.96561	1.159979	-0.83	0.41	-3.31189	1.380672
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	-0.36939	0.391458	-0.94	0.351	-1.16119	0.422409
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-1.117	0.638178	-1.75	0.088	-2.40784	0.173837
Cexdk (Baseline educational expectation was "don't know" ---ref BA)	Cexdk	-0.7174	0.136398	-5.26	0	-0.99329	-0.44151
Cexhs (Baseline educational expectation was high school only---ref BA)	Cexhs	-1.15535	0.263932	-4.38	0	-1.68921	-0.6215
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-1.08164	0.159795	-6.77	0	-1.40485	-0.75842
cex14aa (Baseline educational expectation was two-year---ref BA)	cex14aa	-0.62101	0.096702	-6.42	0	-0.81661	-0.42541
Cexma (Baseline educational expectation was Masters Degree---refer BA)	Cexma	0.130731	0.127043	1.03	0.31	-0.12624	0.387699
Cexphd (Baseline educational expectation was PhD---ref BA)	Cexphd	0.260035	0.125456	2.07	0.045	0.006277	0.513794
Cothrac (Race was not Hispanic, Black, or White—ref Black)	Cothrac	-0.11733	0.298544	-0.39	0.696	-0.72119	0.486535
Chisp (Hispanic—ref Black)	Chisp	-0.3342	0.21233	-1.57	0.124	-0.76368	0.095275
Cwhite (Race was White, not Hispanic—ref Black)	Cwhite	-0.50434	0.164489	-3.07	0.004	-0.83705	-0.17163
Cfemale (Female)	Cfemale	0.655618	0.074893	8.75	0	0.504132	0.807103
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	Parbefor	0.404186	0.15019	2.69	0.01	0.100399	0.707974
_cons	_cons	0.983643	0.520775	1.89	0.066	-0.06972	2.03701
NOTE: Results of this table appear in figure I and in table 5 and table 6. 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. See also table B-1a for results using an alternative variable for EHSGY estimation. Number of strata (wprstco)= 28; Number of PSU ( wprojid) = 67; uses postratified longitudinal baseline weight (v5bwgtp1).							
SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.							

Table B-1a. Intent to Treat (ITT) logistic regression results 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% Confidence 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 two-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, 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 Program 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; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.							

Table B-2. Intent to Treat (ITT), <i>excludes project 69</i> , 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 (estimate reported in table 5 and table 6)							
pr-T = 73.4 pr-C = 64.3 Difference = 9.1**** kenye2	Variable name	Coef.	Linearized Std. Err.	t	P> t	95% Confidence Interval	
FFUTC (random assigned to treatment)	ffutc	0.489536	0.128311	3.82	0	0.230002	0.74907
Gr79293 (Grade 7 on baseline ref grade 9)	gr79293	0.521967	0.625101	0.84	0.409	-0.74242	1.786352
Gr89293 (Grade 8 on baseline ref grade 9)	gr89293	-0.39121	0.286428	-1.37	0.18	-0.97057	0.188144
Gr109293 (Grade 10 on baseline ref grade 9)	gr109293	-0.45553	0.563581	-0.81	0.424	-1.59548	0.68442
Gr119293 (Grade 11 on baseline ref grade 9)	gr119293	-2.12228	0.983931	-2.16	0.037	-4.11247	-0.1321
Clowoy (Low income only)	clowoy	0.346651	0.249567	1.39	0.173	-0.15815	0.851449
Cfgenoy (First generation only)	cfgenoy	0.456902	0.247804	1.84	0.073	-0.04433	0.958132
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	0.152517	0.657632	0.23	0.818	-1.17767	1.482704
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	-0.10336	0.320505	-0.32	0.749	-0.75164	0.544922
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-1.08313	0.627791	-1.73	0.092	-2.35296	0.186695
Cexdk (Baseline educational expectation was “don’t know”---ref BA)	cexdk	-0.79163	0.149155	-5.31	0	-1.09333	-0.48994
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	-1.24988	0.364853	-3.43	0.001	-1.98787	-0.5119
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-0.96619	0.175638	-5.5	0	-1.32145	-0.61092
cex14aa (Baseline educational expectation was two-year---ref BA)	cex14aa	-0.65279	0.123614	-5.28	0	-0.90282	-0.40276
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	0.162114	0.149791	1.08	0.286	-0.14087	0.465095
Cexphd (Baseline educational expectation was PhD---ref BA)	cexphd	0.346946	0.121938	2.85	0.007	0.100304	0.593588
Cothrac (Race was not Hispanic, Black, or White—ref Black)	cothrac	0.062038	0.279237	0.22	0.825	-0.50277	0.626848
Chisp (Hispanic—ref Black)	chisp	-0.28652	0.3381	-0.85	0.402	-0.97039	0.397353
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.45507	0.167869	-2.71	0.01	-0.79462	-0.11552
Cfemale (Female)	cfemale	0.651833	0.088637	7.35	0	0.472549	0.831118
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.288657	0.175721	1.64	0.108	-0.06677	0.644085
_cons	_cons	0.519335	0.33867	1.53	0.133	-0.16569	1.20436
NOTE: Results of this table appear in figure I and in table 5 and table 6. 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. See also table B-2a for results using an alternative variable for EHSGY estimation. Number of strata (v5no69st) = 27; Number of PSU (wprojid) = 66; postratified longitudinal baseline weight (v5bwgtp1).							
<b>SOURCE:</b> Data tabulated June 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.							

Table B-2a. Intent to Treat (ITT), <i>excludes project 69</i> , logistic regression results 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 (EHSYG): National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04 (uses alternative grade variable for EHSYG standardization)							
pr-T = 73.3 pr-C = 60.6 Difference = 12.7**** npse18	Variable name	Coef.	Linearized Std. Err.	t	P> t	95% Confidence Interval	
FFUTC (random assigned to treatment)	ffutc	0.404011	0.113939	3.55	0.001	0.173547	0.634475
Ffgr9 (grade 10 ref)	ffgr9	0.37182	0.280346	1.33	0.192	-0.19523	0.938873
Ffgr11 (grade 10 ref)	ffgr11	-0.02695	0.159631	-0.17	0.867	-0.34983	0.295939
Ffgr12 (grade 10 ref)	ffgr12	-1.32323	0.547936	-2.41	0.021	-2.43154	-0.21493
Clowoy (Low income only)	clowoy	0.465974	0.202402	2.3	0.027	0.056577	0.87537
Cfgenoy (First generation only)	cfgenoy	0.46034	0.253056	1.82	0.077	-0.05151	0.972195
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	0.663936	0.314424	2.11	0.041	0.027953	1.299919
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	0.358413	0.261897	1.37	0.179	-0.17132	0.888151
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-0.4937	0.208163	-2.37	0.023	-0.91475	-0.07265
Cexdk (Baseline educational expectation was don't know---ref BA)	cexdk	-0.853	0.158962	-5.37	0	-1.17454	-0.53147
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	-1.38008	0.38735	-3.56	0.001	-2.16357	-0.59659
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-0.89794	0.182897	-4.91	0	-1.26788	-0.528
cex14aa (Baseline educational expectation was two-year---ref BA)	cex14aa	-0.67054	0.141661	-4.73	0	-0.95708	-0.38401
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	0.028171	0.146501	0.19	0.849	-0.26816	0.324498
Cexphd (Baseline educational expectation was PhD---ref BA)	cexphd	0.28854	0.102476	2.82	0.008	0.081263	0.495816
Cothrac (Race was not Hispanic, Black, or White—ref Black)	cothrac	0.189738	0.214378	0.89	0.382	-0.24388	0.623358
Chisp (Hispanic—ref Black)	chisp	-0.20354	0.293326	-0.69	0.492	-0.79685	0.389766
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.44839	0.182601	-2.46	0.019	-0.81774	-0.07905
Cfemale (Female)	cfemale	0.583567	0.095933	6.08	0	0.389524	0.77761
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.273002	0.128501	2.12	0.04	0.013084	0.532919
_cons	_cons	0.099075	0.269472	0.37	0.715	-0.44598	0.644134
NOTE: Model uses an alternative variable from the First Follow-up (A3) instead of variable B1 from the Baseline Survey on which to standardize EHSYG (See table B-2). SFA = Student Financial Aid files; Ref = left out reference in dummy variable sequence. See table 5 in text for additional note information; Number of strata (v5no69st) = 27; Number of PSU (wprojid) = 66; postratified longitudinal baseline weight (v5bwgtp1). <b>SOURCE:</b> Data tabulated June 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.							

<b>Table B-3. 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)</b>							
xb T = 62.9 xb C = 53.6 Difference = 9.3**** (Kaidhs)	Variable name	Coef.	Linearized Std. Err.	T	P> t	95% Confidence Interval	
xnewgp (Participated in UB/UBMS)	xnewgp	0.123149	0.037028	3.33	0.002	0.048253	0.198045
Gr79293 (Grade 7 in 1992-93 ref grade 9)	gr79293	-0.01293	0.132541	-0.1	0.923	-0.28102	0.255162
Gr89293 (Grade 8 in 1992-93 ref grade 9)	gr89293	-0.12538	0.048361	-2.59	0.013	-0.2232	-0.02756
Gr109293 (Grade 10 in 1992-93 ref grade 9)	gr109293	0.008279	0.125262	0.07	0.948	-0.24509	0.261646
Gr119293 (Grade 11 in 1992-93 ref grade 9)	gr119293	-0.24429	0.168436	-1.45	0.155	-0.58498	0.096407
Clowoy (Low income only)	clowoy	0.017819	0.049742	0.36	0.722	-0.08279	0.118432
Cfgenoy (First generation only)	cfgenoy	0.051589	0.051158	1.01	0.319	-0.05189	0.155066
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	-0.02485	0.141647	-0.18	0.862	-0.31135	0.261663
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	-0.04074	0.055938	-0.73	0.471	-0.15389	0.072402
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-0.14285	0.118341	-1.21	0.235	-0.38222	0.096515
Cexdk (Baseline educational expectation was “don’t know”---ref BA)	cexdk	-0.18389	0.040305	-4.56	0	-0.26541	-0.10236
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	-0.26295	0.0628	-4.19	0	-0.38997	-0.13592
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-0.18834	0.035934	-5.24	0	-0.26103	-0.11566
cex14aa (Baseline educational expectation was 2-year---ref BA)	cex14aa	-0.17929	0.024134	-7.43	0	-0.22811	-0.13047
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	0.022336	0.025866	0.86	0.393	-0.02998	0.074655
Cexphd (Baseline educational expectation was PhD---ref BA)	cexphd	0.042772	0.018322	2.33	0.025	0.005713	0.079831
Cothrac (Race was not Hispanic, Black, or White—ref Black)	cothrac	0.018877	0.046879	0.4	0.689	-0.07595	0.113699
Chisp (Hispanic—ref Black)	chisp	-0.06829	0.06105	-1.12	0.27	-0.19177	0.055199
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.10059	0.038615	-2.6	0.013	-0.17869	-0.02248
Cfemale (Female)	cfemale	0.139323	0.02847	4.89	0	0.081736	0.196909
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.032356	0.0231	1.4	0.169	-0.01437	0.07908
_cons	_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. See also table B-3a for results using an alternative variable for EHSGY estimation. 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 ffutc.							
<b>SOURCE:</b> Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.							

Table B-3a. 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 (uses alternative grade variable for standardization)							
pr-T = 65.4 pr-C = 55.0 Difference = 10.4**** knaidh1	Variable name	Coef.	Linearized Std. Err.	t	P> t	95% Confidence Interval	
xnewgp (Participated in UB/UBMS)	xnewgp	0.098531	0.02724	3.62	0.001	0.043433	0.153629
Ffgr9 (grade 10 ref)	ffgr9	-0.00902	0.052581	-0.17	0.865	-0.11538	0.097332
Ffgr11 (grade 10 ref)	ffgr11	0.014367	0.027164	0.53	0.6	-0.04058	0.069311
Ffgr12 (grade 10 ref)	ffgr12	-0.24887	0.064124	-3.88	0	-0.37858	-0.11917
Clowoy (Low income only)	clowoy	0.057143	0.048196	1.19	0.243	-0.04034	0.154628
Cfgenoy (First generation only)	cfgenoy	0.029202	0.039698	0.74	0.466	-0.0511	0.109499
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	0.178713	0.056336	3.17	0.003	0.064763	0.292663
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	0.073839	0.037345	1.98	0.055	-0.0017	0.149376
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-0.01392	0.062567	-0.22	0.825	-0.14047	0.112637
Cexdk (Baseline educational expectation was “don’t know”---ref BA)	cexdk	-0.17051	0.031567	-5.4	0	-0.23436	-0.10666
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	-0.24205	0.060803	-3.98	0	-0.36504	-0.11907
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-0.1457	0.02996	-4.86	0	-0.2063	-0.0851
cex14aa (Baseline educational expectation was two-year---ref BA)	cex14aa	-0.17534	0.029116	-6.02	0	-0.23423	-0.11645
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	0.012577	0.023624	0.53	0.597	-0.03521	0.06036
Cexphd (Baseline educational expectation was PhD---ref BA)	cexphd	0.014907	0.026489	0.56	0.577	-0.03867	0.068485
Cothrac (Race was not Hispanic, Black, or White—ref Black)	cothrac	0.038572	0.0459	0.84	0.406	-0.05427	0.131412
Chisp (Hispanic—ref Black)	chisp	-0.02733	0.036807	-0.74	0.462	-0.10178	0.047119
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.08829	0.03744	-2.36	0.023	-0.16402	-0.01256
Cfemale (Female)	cfemale	0.146096	0.026613	5.49	0	0.092267	0.199925
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.064862	0.021253	3.05	0.004	0.021875	0.10785
_cons	_cons	0.457948	0.040832	11.22	0	0.375358	0.540537
NOTE: Model uses an alternative variable from the first follow-up survey (A3) instead of variable B1 from the Baseline Survey on which to standardize EHSGY (See table B-3). 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 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 ffuttc. SOURCE: Data tabulated June 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; and federal Student Financial Aid (SFA) files 1994-95 to 2003-04.							

Table B-4 Intent to Treat (ITT), <i>excludes project 69</i> , logistic regression results for dependent variable of having evidence of attaining a BA degree in +6 of expected high school graduation year (EHSGY) from any applicable survey, SFA Files, or National Student Clearinghouse (NSC): National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04 (estimate in table 10)							
pr-T = 17.0 pr-C = 13.3 Difference = 3.7**** kbahs6	Variable name	Coef.	Linearized Std. Err.	t	P> t	95% Confidence Interval	
FFUTC (random assigned to treatment)	ffutc	0.362466	0.084166	4.31	0	0.192224	0.532708
Gr79293 (Grade 7 on baseline ref grade 9)	gr79293	0.092829	0.334572	0.28	0.783	-0.58391	0.769564
Gr89293 (Grade 8 on baseline ref grade 9)	gr99293	0.446231	0.304951	1.46	0.151	-0.17059	1.063053
Gr109293 (Grade 10 on baseline ref grade 9)	gr109293	0.474865	0.600971	0.79	0.434	-0.74071	1.690443
Gr119293 (Grade 11 on baseline ref grade 9)	gr119293	-1.46837	1.670039	-0.88	0.385	-4.84634	1.909606
Clowoy (Low income only)	clowoy	0.706581	0.265714	2.66	0.011	0.169123	1.244038
Cfgenoy (First generation only)	cfgenoy	0.534226	0.191623	2.79	0.008	0.146633	0.92182
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	-0.40178	0.652044	-0.62	0.541	-1.72066	0.917103
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	-0.38824	0.265334	-1.46	0.151	-0.92493	0.148449
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-0.53442	0.4097	-1.3	0.2	-1.36312	0.294275
Cexdk (Baseline educational expectation was “don’t know”---ref BA)	cexdk	-0.67588	0.215566	-3.14	0.003	-1.1119	-0.23985
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	-2.17255	0.908448	-2.39	0.022	-4.01006	-0.33504
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-0.6227	0.277369	-2.25	0.031	-1.18374	-0.06167
cex14aa (Baseline educational expectation was two-year---ref BA)	cex14aa	-1.28374	0.274614	-4.67	0	-1.8392	-0.72828
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	0.250644	0.165068	1.52	0.137	-0.08324	0.584526
Cexphd (Baseline educational expectation was PhD---ref BA)	cexphd	0.19915	0.158047	1.26	0.215	-0.12053	0.518831
Cothrac (Race was not Hispanic, Black, or White—ref Black)	cothrac	0.421884	0.268605	1.57	0.124	-0.12142	0.965189
Chisp (Hispanic—ref Black)	chisp	-0.15843	0.244249	-0.65	0.52	-0.65247	0.335611
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.25651	0.169433	-1.51	0.138	-0.59922	0.086198
Cfemale (Female)	cfemale	0.662424	0.125721	5.27	0	0.40813	0.916719
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.116322	0.110851	1.05	0.3	-0.10789	0.340539
_cons	_cons	-2.22556	0.185144	12.02	0	-2.60005	-1.85107
NOTE: Results of this table appear in table 10. 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. See also table B-4a for results using an alternative variable for EHSGY estimation. Number of strata (v5no69st) = 27; Number of PSU (wprojid) = 66; poststratified longitudinal baseline weight (v5bwgtp1).							
SOURCE: Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; federal Student Financial Aid (SFA) files 1994-95 to 2003-04; and National Student Clearinghouse Data 1995-2004							

Table B-4a Intent to Treat (ITT), <i>excludes project 69</i> , logistic regression results for dependent variable of having evidence of attaining a BA degree in +7 of expected high school graduation year (EHSGY) from the any applicable survey, SFA Files, or National Student Clearinghouse (NSC): National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04 (uses alternative grade variable for standardization)							
pr-T = 18.0 pr-C = 14.5 Difference = 3.5**** knba7	Variable name	Coef.	Linearized Std. Err.	t	P> t	95% Confidence Interval	
FFUTC (random assigned to treatment)	ffutc	0.330297	0.082796	3.99	0	0.162826	0.497767
Ffgr9 (grade 10 ref)	ffgr9	-0.22032	0.215644	-1.02	0.313	-0.6565	0.215859
Ffgr11 (grade 10 ref)	ffgr11	-0.00821	0.246862	-0.03	0.974	-0.50754	0.491114
Ffgr12 (grade 10 ref)	ffgr12	-1.20172	0.617098	-1.95	0.059	-2.44992	0.046477
Clowoy (Low income only)	clowoy	0.684771	0.264543	2.59	0.013	0.149683	1.219859
Cfgenoy (First generation only)	cfgenoy	0.577483	0.155877	3.7	0.001	0.262192	0.892773
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	0.341965	0.384344	0.89	0.379	-0.43544	1.119374
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	0.040414	0.169063	0.24	0.812	-0.30155	0.382376
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-0.28152	0.224888	-1.25	0.218	-0.7364	0.173358
Cexdk (Baseline educational expectation was “don’t know”---ref BA)	cexdk	-0.59992	0.208694	-2.87	0.007	-1.02204	-0.1778
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	-2.19774	0.883071	-2.49	0.017	-3.98391	-0.41156
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-0.53217	0.287618	-1.85	0.072	-1.11393	0.049596
cex14aa (Baseline educational expectation was two-year---ref BA)	cex14aa	-1.26668	0.275738	-4.59	0	-1.82441	-0.70895
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	0.35107	0.197926	1.77	0.084	-0.04927	0.751413
Cexphd (Baseline educational expectation was PhD---ref BA)	cexphd	0.213193	0.157779	1.35	0.184	-0.10595	0.532331
Cothrac (Race was not Hispanic, Black, or White—ref Black)	cothrac	0.382197	0.25811	1.48	0.147	-0.13988	0.904273
Chisp (Hispanic—ref Black)	chisp	-0.12792	0.225029	-0.57	0.573	-0.58309	0.327239
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.31609	0.154333	-2.05	0.047	-0.62826	-0.00392
Cfemale (Female)	cfemale	0.67513	0.117384	5.75	0	0.437698	0.912561
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.198067	0.097729	2.03	0.05	0.000391	0.395742
_cons	_cons	-2.10691	0.195875	-10.76	0	-2.50311	-1.71072
NOTE: Model uses an alternative variable from the first follow up (A3) on which to standardize grade. Model uses an alternative variable from the First Follow-up Survey (A3) instead of variable B1 from the Baseline Survey on which to standardize EHSGY (See table B-4). SFA = Student Financial Aid files; Ref = left out reference in dummy variable sequence. See table 5 in body of report for additional note information; Number of strata (v5no69st) = 27; Number of PSU (wprojid) = 66; poststratified longitudinal baseline weight (v5bwgtp1). <b>SOURCE:</b> Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; federal Student Financial Aid (SFA) files 1994-95 to 2003-04; and National Student Clearinghouse Data 1995-2004.							

**Table B-5. Intent to Treat (ITT) logistic regression results for higher academic risk subgroup (bottom 20 percent on 9<sup>th</sup> grade**

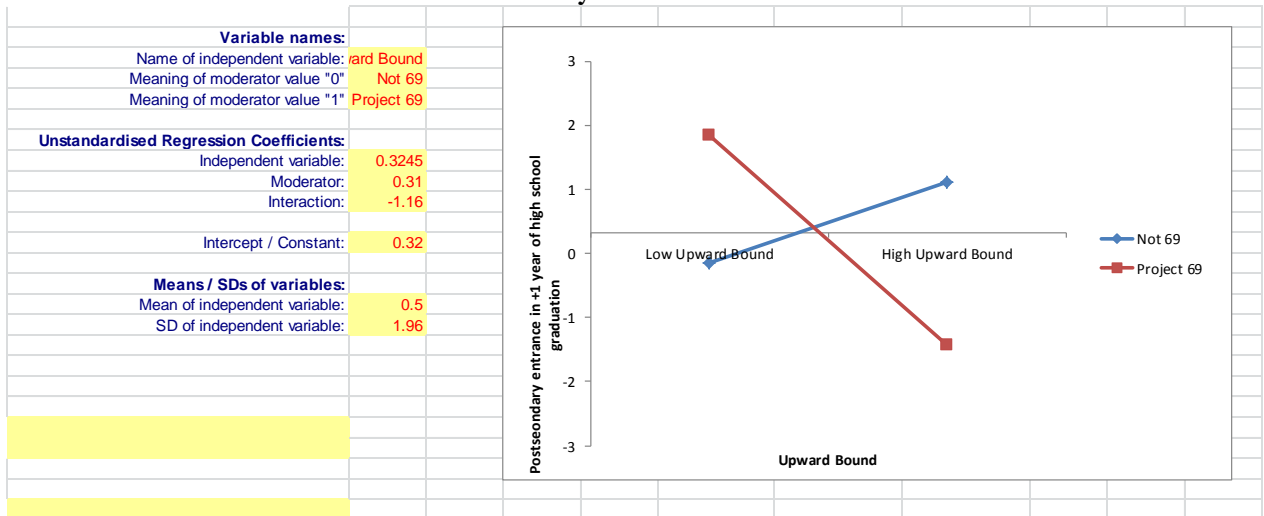


academic indicators) modeling dependent variable of having evidence of entering postsecondary within +1 (18 months) of expected high school graduation year (EHSYG): National Evaluation of Upward Bound, study conducted 1992-93 to 2003-2004 (estimate in table 12 in text)								
Pr---Treatment- 60.1 Control- 41.0 Difference- 19.1*** n3ar20h== 1) 521 cases-- kenye2	Variable name	Coef.	Linearized Std. Err.	T	P>  t	95% Confidence Interval		
Ffutc (random assignment to treatment)	ffutc	0.68927	0.243843	2.83	0.007	0.196051		1.182489
Gr79293 (Grade 7 in 1992-93 ref grade 9)	gr79293	2.221241	1.032169	2.15	0.038	0.133482		4.308999
Gr89293 (Grade 8 in 1992-93 ref grade 9)	gr89293	-0.25723	0.640009	-0.4	0.69	-1.55177		1.037314
Gr109293 (Grade 10 in 1992-93 ref grade 9)	gr109293	0.870869	0.988209	0.88	0.384	-1.12797		2.869711
Gr119293 (Grade 11 in 1992-93 ref grade 9)								
Clowoy (Low income only)	clowoy	0.677017	0.435089	1.56	0.128	-0.20303		1.557068
Cfgenoy (First generation only)	cfgenoy	0.543434	0.497777	1.09	0.282	-0.46342		1.550283
C11gssf (Grade was 11 on student selection form---ref grade 9)	c11gssf	-1.34637	1.380078	-0.98	0.335	-4.13784		1.445102
C10gssf (Grade was 10 on student selection form---ref grade 9)	c10gssf	0.06583	0.682602	0.1	0.924	-1.31486		1.446522
C8gssfm (Grade was 8 on student selection form---ref grade 9)	c8gssfm	-2.55885	1.08985	-2.35	0.024	-4.76328		-0.35442
Cexdk (Baseline educational expectation was "don't know"---ref BA)	cexdk	-0.65612	0.522212	-1.26	0.216	-1.71239		0.400155
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	-2.14998	1.264569	-1.7	0.097	-4.70782		0.407849
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-0.98	0.453137	-2.16	0.037	-1.89656		-0.06345
cex14aa (Baseline educational expectation was two-year---ref BA)	cex14aa	-0.50108	0.412805	-1.21	0.232	-1.33606		0.333896
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	-0.11761	0.344832	-0.34	0.735	-0.8151		0.579874
Cexphd (Baseline educational expectation was PhD---ref BA)	cexphd	-0.80588	0.469222	-1.72	0.094	-1.75497		0.143209
Cothrac (Race was not Hispanic, Black, or White---ref Black)	cothrac	-0.54198	0.572089	-0.95	0.349	-1.69914		0.615182
Chisp (Hispanic---ref Black)	chisp	-0.4907	0.643455	-0.76	0.45	-1.79221		0.810811
Cwhite (Race was White, not Hispanic---ref Black)	cwhite	-0.647	0.348201	-1.86	0.071	-1.3513		0.057307
Cfemale (Female)	cfemale	0.534927	0.181975	2.94	0.005	0.166847		0.903007
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.411237	0.334681	1.23	0.227	-0.26572		1.088194
_cons	_cons	0.080847	0.53191	0.15	0.88	-0.99504		1.156736
NOTE: Results of this table appear in figure 9 and table 12. 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 postratified longitudinal baseline weight (v5bwgtp1). <b>SOURCE:</b> Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study 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 1992-93 to 2003-04; federal Student Financial Aid (SFA) files 1994-95 to 2003-04.								

<b>Table B-6. Intent to Treat (ITT) logistic regression results for sample members with lower educational expectations for modeling of dependent variables of attainment of any postsecondary credential using survey data only adjusted for non-response: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04</b>							
Pr--- Treatment- 50.3 Control- 35.0 Difference - 15.3** bahexp == 0	Variable name	Coef.	Linearized Std. Err.	t	P>  t	[95% Conf.	Interval]
Ffutc (random assignment to treatment)	ffutc	0.781074	0.356626	2.19	0.035	0.059731	1.502417
Gr79293 (Grade 7 in 1992-93 ref grade 9)	gr79293	15.74681	0.737344	21.36	0	14.25539	17.23823
Gr89293 (Grade 8 in 1992-93 ref grade 9)	gr89293	-1.76353	0.856123	-2.06	0.046	-3.4952	-0.03186
Gr109293 (Grade 10 in 1992-93 ref grade 9)	gr109293	-0.04512	0.81914	-0.06	0.956	-1.70199	1.611743
Gr119293 (Grade 11 in 1992-93 ref grade 9)							
Clowoy (Low income only)	clowoy	1.004854	0.97003	1.04	0.307	-0.95722	2.966925
Cfgenoy (First generation only)	cfgenoy	0.316873	0.610377	0.52	0.607	-0.91773	1.551476
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	-1.56826	1.093469	-1.43	0.159	-3.78	0.643495
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	-1.38135	0.637629	-2.17	0.036	-2.67107	-0.09162
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-17.8409	.	.	.	.	.
Cexdk (Baseline educational expectation was don't know---ref BA)	cexdk	-0.45648	0.571858	-0.8	0.43	-1.61317	0.700211
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	1.464931	0.468557	3.13	0.003	0.517185	2.412677
Cex13v (Baseline educational expectation was vocational---ref BA)							
cex14aa (Baseline educational expectation was 2-year---ref BA)	cex14aa	-0.71024	0.872378	-0.81	0.421	-2.47479	1.054315
Cexma (Baseline educational expectation was Masters Degree---refer BA)							
Cexphd (Baseline educational expectation was Ph.D---ref BA)							
Cothrac (Race was not Hispanic or Black or White—ref Black)	cothrac	0.188601	0.435626	0.43	0.667	-0.69254	1.069738
Chisp (Hispanic—ref Black)	chisp	0.65849	0.522057	1.26	0.215	-0.39747	1.714449
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.35067	0.413444	-0.85	0.402	-1.18694	0.485598
Cfemale (Female)	cfemale	0.769137	0.544524	1.41	0.166	-0.33227	1.87054
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.135167	0.207005	0.65	0.518	-0.28354	0.553873
_cons	_cons	0.604448	0.873176	0.69	0.493	-1.16172	2.370613
Note: Ref = left out reference in dummy variable sequence. See table 5 in text for complete note information; Number of strata = 28; Number of PSU = 67; Note results using survey data only subject to non-response bias and sub-group results subject to unequal weighting.							

## Appendix Q

Plots of two-way interaction effects of binary moderator project 69 with dependent variable of entering postsecondary within +1 year of expected high school graduation year



For further information see [www.jeremydawson.co.uk/slopes.htm](http://www.jeremydawson.co.uk/slopes.htm).

## Appendix R

### Examples of Back-up Documentation for impact estimates in Exhibit 13

Table R-1. Instrumental variables regression for Treated on Treated (TOT) longitudinal file of all sample members modeling dependent variable of evidence from any applicable follow-up survey (3 to 5), Pell award files, or National Student Clearinghouse (NSC) of BA in +6 years of expected high school graduation year (EHSGY) without project 69—sample members with no evidence from any sources coded as not having a BA in +6: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04 (estimate reported in Exhibit 13)							
Dependent Variable kbahs6 (PPSS derived) T = 21.1 C = 14.1 Difference = 7.0 ***	Name	Coef.	Std. Err.	t	P>  t	[95% Conf. Interval]	
xnewgp (TOT Participated in UB/UBMS)	xnewgp	0.069152	0.013644	5.07	.0000	0.041554	0.09675
Gr79293 (Grade 7 in 1992-93 ref grade 9)	gr79293	-0.04759	0.060383	-0.79	0.435	-0.16973	0.074542
Gr89293 (Grade 8 in 1992-93 ref grade 9)	gr89293	-0.0514	0.041941	-1.23	0.228	-0.13623	0.033435
Gr109293 (Grade 10 in 1992-93 ref grade 9)	gr109293	0.022341	0.086835	0.26	0.798	-0.1533	0.197982
Gr119293 (Grade 11 in 1992-93 ref grade 9)	gr119293	-0.07402	0.142088	-0.52	0.605	-0.36142	0.213377
Clowoy (Low income only)	clowoy	0.108971	0.049914	2.18	0.035	0.008011	0.209931
Cfgenoy (First generation only)	cfgenoy	0.075356	0.030511	2.47	0.018	0.013643	0.13707
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	-0.06668	0.084937	-0.79	0.437	-0.23848	0.10512
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	-0.04652	0.034757	-1.34	0.189	-0.11682	0.023784
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-0.06141	0.057518	-1.07	0.292	-0.17775	0.054929
Cexdk (Baseline educational expectation was don't know---ref BA)	cexdk	-0.07932	0.024042	-3.3	0.002	-0.12795	-0.03069
Cexhs (Baseline educational expectation was high school only---ref BA)	cex13v	-0.07215	0.030303	-2.38	0.022	-0.13344	-0.01086
Cex13v (Baseline educational expectation was vocational---ref BA)	cexhs	-0.13626	0.018748	-7.27	0	-0.17418	-0.09833
cex14aa (Baseline educational expectation was 2-year--ref BA)	cex14aa	-0.12079	0.020431	-5.91	0	-0.16212	-0.07947
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	0.037389	0.02649	1.41	0.166	-0.01619	0.090971
Cexphd (Baseline educational expectation was Ph.D---ref BA)	cexphd	0.031588	0.024559	1.29	0.206	-0.01809	0.081264
Cothrac (Race was not Hispanic or Black or White—ref Black)	cothrac	0.060315	0.045014	1.34	0.188	-0.03073	0.151364
Chisp (Hispanic—ref Black)	chisp	-0.01927	0.028652	-0.67	0.505	-0.07722	0.038688
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.02981	0.020524	-1.45	0.154	-0.07132	0.011703
Cfemale (Female)	cfemale	0.082776	0.011627	7.12	0	0.059258	0.106294
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.012161	0.01472	0.83	0.414	-0.01761	0.041935
_cons	_cons	0.14058	0.044379	3.17	0.003	0.050814	0.230346
<b>SOURCE:</b> Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Program Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), U.S. Department of Education							

<b>Table R-2. Instrumental variables regression for Treated on Treated (TOT) modeling dependent variable of evidence) of BA by the end of the study period, fifth follow-up responders only with weights adjusted for on-response without project 69: National Evaluation of Upward Bound, study conducted 1992-93 to 2003-04 (estimate reported in Exhibit 13)</b>							
v5dh4_1a pr (Mathematica derived) T = 28.8 C = 21.2 Difference = 7.6 *** v5dh4_1a	Variable name	Coef.	Linearized Std. Err.	t	P> t	[95% Conf.	Interval]
xnewgp (Participated in UB/UBMS)	xnewgp	0.085469	0.030882	2.77	0.009	0.023004	0.147934
Gr79293 (Grade 7 in 1992-93 ref grade 9)	gr79293	-0.18677	0.093043	-2.01	0.052	-0.37497	0.001427
Gr89293 (Grade 8 in 1992-93 ref grade 9)	gr89293	-0.14379	0.078617	-1.83	0.075	-0.30281	0.015226
Gr109293 (Grade 10 in 1992-93 ref grade 9)	gr109293	-0.04484	0.079576	-0.56	0.576	-0.20579	0.116119
Gr119293 (Grade 11 in 1992-93 ref grade 9)	gr119293	-0.0644	0.09729	-0.66	0.512	-0.26119	0.132389
Clowoy (Low income only)	clowoy	0.211413	0.069107	3.06	0.004	0.071632	0.351194
Cfgenoy (First generation only)	cfgenoy	0.110083	0.039279	2.8	0.008	0.030635	0.189532
C11gssf (Grade was 11 on student selection form—ref grade 9)	c11gssf	-0.11876	0.091304	-1.3	0.201	-0.30344	0.065922
C10gssf (Grade was 10 on student selection form—ref grade 9)	c10gssf	-0.14278	0.07331	-1.95	0.059	-0.29106	0.005501
C8gssfm (Grade was 8 on student selection form—ref grade 9)	c8gssfm	-0.04235	0.083008	-0.51	0.613	-0.21025	0.125546
Cexdk (Baseline educational expectation was don't know---ref BA)	cexdk	-0.07725	0.032766	-2.36	0.024	-0.14353	-0.01098
Cexhs (Baseline educational expectation was high school only---ref BA)	cexhs	-0.25599	0.035839	-7.14	0	-0.32848	-0.1835
Cex13v (Baseline educational expectation was vocational---ref BA)	cex13v	-0.13914	0.061902	-2.25	0.03	-0.26435	-0.01393
cex14aa (Baseline educational expectation was 2-year---ref BA)	cex14aa	-0.19172	0.026773	-7.16	0	-0.24588	-0.13757
Cexma (Baseline educational expectation was Masters Degree---refer BA)	cexma	0.007844	0.044606	0.18	0.861	-0.08238	0.098069
Cexphd (Baseline educational expectation was Ph.D---ref BA)	cexphd	-0.06207	0.028258	-2.2	0.034	-0.11922	-0.00491
Cothrac (Race was not Hispanic or Black or White—ref Black)	cothrac	0.112082	0.071994	1.56	0.128	-0.03354	0.257704
Chisp (Hispanic—ref Black)	chisp	0.008647	0.041422	0.21	0.836	-0.07514	0.09243
Cwhite (Race was White, not Hispanic—ref Black)	cwhite	-0.05605	0.032871	-1.71	0.096	-0.12254	0.010437
Cfemale (Female)	cfemale	0.063783	0.027912	2.29	0.028	0.007325	0.12024
Parbefor (Reported participated in other pre-college supplemental services before random assignment)	parbefor	0.022381	0.024845	0.9	0.373	-0.02787	0.072634
_cons	_cons	0.337508	0.089428	3.77	0.001	0.156622	0.518394
<b>SOURCE:</b> Data tabulated January 2008 using: National Evaluation of Upward Bound data files, study sponsored by the Policy and Program Studies Services (PPSS), of the Office of Planning, Evaluation and Policy Development (OPEPD), U.S. Department of Education							