THE FAFSA COMPLETION STUDY: SUPPORTING STATEMENT FOR OMB CLEARANCE

PART B: COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

SEPTEMBER 12, 2012

The Institute of Education Sciences (IES) at the U.S. Department of Education (ED) requests Office of Management and Budget (OMB) approval to conduct data collection for a rigorous study of the FAFSA Completion Project. The project will provide 80 Local Educational Agencies (LEAs) or school districts with access to data on whether specific students have completed the Free Application for Student Aid (FAFSA); this information is intended to help schools implement targeted outreach to seniors and their families who have not yet submitted a FAFSA, or who submitted a FAFSA that may be incomplete. The evaluation of the project is being conducted by IES staff in the National Center for Education Evaluation. The study will use a delayed-treatment control group design, and will examine the impact of access to the data on students' application for and receipt of federal student aid and a proxy for college enrollment. This is a first collaboration between IES and ED's Federal Student Aid (FSA) office on an evaluation relating to financial aid.

Overview of the Project and Evaluation Design

In January 2012, ED issued a nationwide invitation to LEAs to apply to participate in the project. In addition to meeting specified technical requirements to gain access to the FAFSA data, LEAs were also required to have two or more high schools, since the study design calls for random assignment of schools within districts, and to agree to be part of an evaluation. Out of 149 applicants, 80 multiple-district districts from 27 states were randomly selected to participate.

The participating school districts and schools are responsible for developing and implementing their plans for targeted outreach and counseling of students. In October 2012, FSA will use funds from the Lumina Foundation to convene a conference of participating districts to provide technical assistance in developing an approach to using the FAFSA data to increase FAFSA completion rates for seniors. In addition, the participating school districts are responsible for establishing the technical capacity (including hardware and software) to receive the FAFSA completion data through an FSA secure portal. No federal funds are available to districts for any part of the demonstration project.

Also in fall 2012, each participating school district will submit a list of high schools to IES that it wishes to include in the study, as well as a roster of all the seniors in each school. Within districts, designated schools will be randomly assigned to the treatment group or the control group.[1] Treatment school staff will have access to students' FAFSA completion data beginning in the 2012-2013 academic year, and control school staff will have access beginning in the following year, making 2012-2013 the experimental study period.

We will compare the outcomes of seniors in the two groups of schools to estimate the effects of schools' access to individual student FAFSA applications and any follow up that the schools undertake. Specifically, the evaluation will address the following research questions:

- ***What is the impact of access to the data on students' application for federal student aid?*** With completion of a FAFSA a necessary prerequisite for obtaining student aid, the most direct goal of the project is to increase rates of FAFSA form completions; the evaluation will examine whether that is the case.

[1] The study will probably "block" or match schools within districts prior to conducting the random assignment. Blocks would be based on a combination of school characteristics including prior year's FAFSA completion rate, and possibly other variables, including enrollment size, the percentage of students eligible for free and reduced price lunches, and average test scores.

- ***What is the impact on students' receipt of federal student aid?*** Receipt of financial aid is the key gateway to college enrollment. In addition, it is possible that schools' project efforts could not only increase FAFSA application completion rates but also the accuracy and quality of the information provided, thereby making students more likely to *receive* aid,

- ***What is the impact on college enrollment?*** Ultimately, ED hopes that providing schools with access to individual student FAFSA data will – through increased receipt of financial aid – also increase college enrollment.

The data collection to address these research questions will create minimal burden for the respondents and will have limited cost to the government. This package is requesting permission for IES to obtain lists of high schools and student rosters from the participating districts or their high schools. Other data for the study -- completion of a FAFSA, receipt of Pell Grant, and a proxy for college enrollment (whether an institution of higher education has drawn down the Pell Grant funds for individual students) -- will come from existing ED administrative data that will not generate any new burden because they are already collected for other purposes. The analyses will be conducted internally by IES staff on data that is stripped of personally identifiable information. The results will be summarized in an internal memo.

The study is authorized by the Education Sciences Reform Act (ESRA) of 2002, which created IES and authorizes it to conduct research in areas of demonstrated national need (20 US Code Chapter 76 9501-9564).

B.1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. *Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.*

The study is designed to collect information on public high school seniors in 80 multiple-high school LEAs. This number was based on FSA's calculations of the costs of providing on-demand access to student-level FAFSA completion data and the need to allow some smaller (single-high school) LEA to participate as well. Applications from multiple-high school LEAs totaled 149; IES randomly sorted the applicants such that the first 80 could provide the project sample and FSA could go down the list if an LEA ultimately was unable to participate. Thus, the district sample is representative of the universe of multiple-high school LEAs who volunteered to participate.

LEAs will select which of their high schools will participate in the project and therefore the evaluation. We expect most LEAs to include all their high schools. However, LEAs may choose to exclude certain high schools, such as alternative or special education schools or schools undergoing other changes that might make implementing the targeted outreach to and counseling of students challenging. A preliminary investigation of the 80 participating districts using the 2009-2010 Common Core of Data (CCD) indicates that each LEA, on average, includes 10 schools of any configuration containing grade 12. If LEAs include all their high schools in the study, this will yield an estimated total of 800 schools in the study as a whole.

The study universe consists of all seniors in the selected high schools, estimated to be 179,200 students (Table B.1). The study will gather data for the entire universe rather than for a sample of students. This will increase the precision of estimates at little additional cost to the study since the only original data to be collected are student rosters. The study anticipates obtaining complete rosters from each study school.

Table B.1 Population Sizes

Number of LEAs	80
Estimated number of high schools per LEA	10
Estimated total number of high schools	800
Estimated number of seniors per high school	224
Estimated total number of seniors	179,200

Note: Estimates based on 2009-2010 Common Core of Data.

B.2. Describe the procedures for the collection of information including:

- *Statistical methodology for stratification and sample selection*,
- *Estimation procedure*,
- *Degree of accuracy needed for the purpose described in the justification*,
- *Unusual problems requiring specialized sampling procedures, and*
- *Any use of periodic (less frequent than annual) data collection cycles to reduce burden.*

The goal of the study is to estimate experimental impacts of school and LEA access to student FAFSA completion data on student outcomes. Allowing schools and LEAs to check whether their seniors have submitted complete FAFSAs may help them target completion assistance and counseling more effectively to high school seniors who could benefit from help with applying for financial aid. The outcomes of interest include whether students apply for federal financial aid, the amount of financial aid received, and a proxy for college enrollment – all from internal ED databases.

a. Selection of Districts and Random Assignment Procedures

ED's Federal Student Aid (FSA) Office issued an invitational letter to multiple-high school LEAs that described the study and specified requirements for: (1) computer hardware and software to ensure compatibility with FAFSA and ED systems, and (2) participation in the evaluation. To respond, LEAs merely had to submit letters of interest. After receiving such letters from 149 LEAs, LEAs were sorted in random order and the first 80 were selected. However, if any of these LEAs prove unable to meet study requirements, the next district on the list will be selected to obtain the target of 80.

As the treatment is a school-level intervention, the study design calls for random assignment of schools to experimental conditions. Through district personnel, treatment school staff will have access to FAFSA completion data for students beginning in the 2012-2013 academic year, and control schools will not have access to these data until the following year. Thus, the experiment will take place in 2012-2013. Random assignment will take place within districts, with half of the district's schools being assigned to the treatment group and half to the control group. Where possible, the study will conduct block random assignment within districts. For larger LEAs with a large number of high schools, blocks

will be defined by the school's prior year's FAFSA completion rate, and possibly variables from the Common Core of Data, including enrollment and the percentage of students eligible for free and reduced price lunches. For LEAs with a small number of high schools, the study will only use the prior year FAFSA completion rate, or (in the case of LEAs with only two high schools) not block at all. This will help ensure that these important characteristics, which may be related to the types of counseling schools offer and the level of interest in college among students, are balanced between treatment and control groups.

b. Data Sources and Collection Schedule

The key information necessary for data collection consists of lists of participating schools from each of the LEAs and rosters of seniors at each school. Rosters include student directory information: name, date of birth, and address. This directory information will be used to obtain outcome data for each student from internal ED data sources. IES will collect lists of schools in early fall 2012, after OMB approval is received.

IES will collect senior rosters from LEAs in November 2012. This timing balances two conflicting goals. Ideally, the study would obtain a complete list of study seniors *before* or immediately after random assignment to avoid potential selection bias that may arise if the treatment affects dropout rates. However, school enrollment typically is not finalized until about a month after the school year begins due to transfers and late entrants. Gathering rosters in November will preserve the internal validity of the impact estimates since the important part of the treatment (outreach and counseling) will not occur until students begin submitting FAFSAs in January.

Outcome data will be obtained by matching the roster data to internal ED data sources. FSA records of FAFSA completion and financial aid receipt will be extracted using the student directory information in both June 2013 and December 2013. FSA data on whether an institution of higher education has drawn a student's Pell Grant funds (a proxy for college enrollment) will be extracted in December 2013.

Information on name, date of birth, and address should be sufficient to correctly match most students on the rosters to their outcome data. However, matching may be challenging for students who move and thus change their address. Name and date of birth may not be unique identifiers among a large population of students who are very close in age. For movers who do not change school, high school identifiers may be helpful in the matching process, but the study may be unable to match movers who change residence and school.

The issue of matching is closely related to the issue of sample attrition. There is likely to be some flux in and out of study schools by students throughout the school year. Students who move into study schools after November will not be part of the analysis sample, although movers into treatment schools will still receive the treatment. Students who leave study schools—drop-outs and movers, whom we will call "attriters"—are more problematic. We have no expectation that the treatment – outreach to and counseling of students and their families about the benefits of completing the FAFSA – will affect attrition so we believe that the study results will not be biased. However, we anticipate taking some steps to both assess and adjust for any potential problems.

The study's first step in addressing this potential issue will be to compare attrition rates between the treatment and control schools. One way to mitigate potential bias if rates differ is to construct analysis weights. Such weights let the respondents within each treatment group who look most like

attriters carry a greater weight so that they can stand in for their missing counterparts. Given the study's limited information on students, weights will be adjusted based on school, so that students in schools with relatively more attriters will carry relatively greater weights. Another way to address the issue is to try to bound the size of the impact based on different extreme assumptions. For instance, the assumption that treatment attriters all enroll in college and that no control attriters do produces an upper bound, while the converse assumption that all control attriters and no treatment attriters enroll in college produces a lower bound.

Impacts of the FAFSA Completion Study will be estimated using hierarchical linear models (HLM). The models will analyze student-level outcomes, where the outcomes are a function of treatment status; student, school and district random effects; and student, school, and district-level control variables. A number of sources will furnish data that will be used to define control variables. Student-level control variables may include age categories (such as over or under 18; or above or below typical age range for grade 12), as these are the only student data the study will have besides outcome data. School-level control variables will include prior year's FAFSA completion rate from FSA. Additional school-level control variables will be drawn from the CCD using school NCES identifiers, and may include enrollment size, grade configuration, special status (e.g., charter or magnet school), racial and ethnic distribution, and percentage of students who are eligible for free or reduced price lunch. District level controls from the CCD may include urbanicity, number of high schools, revenue per student, and percentage of students who are limited English proficient (LEP). We may also include local geographic controls such as annual unemployment rates and poverty rates will be included using zip code-level data from the Bureau of Labor Statistics and the Area Resource File.

c. Precision and Minimum Detectable Effects

To demonstrate the precision that can be achieved by the study, Table B.2 provides the half-widths of the 95 percent confidence intervals for two means of the outcome variables. Since the outcome variables are binary (e.g., complete/not complete the FAFSA, obtain/did not obtain financial aid), means can be interpreted as proportions. The table computes half-widths under two assumptions about the school-level intraclass correlation (ICC), reflecting the extent to which mean outcomes differ across schools. Research has shown that the school-level ICCs often range from 0.1 to 0.2 for standardized test score outcomes, and from 0.01 to 0.09 for social and risky behavioral outcomes (Schochet 2005). The study's outcomes of FAFSA completion, financial aid receipt, and postsecondary enrollment do not fall into either category. Thus, the study considers first the assumption that the ICC falls in between the ICCs found for each type of outcome, at 0.1, and second a more pessimistic assumption that the ICC is 0.15.

The first row of Table B.2 examines precision for the entire sample. The half-width of the confidence interval ranges from 0.021 to 0.025 when half of the population has a particular outcome. When the proportion is only 0.10, the half-width falls to 0.012-0.015. These figures indicate that the study will produce relatively precise estimates of the outcome variables. Moreover, the study can also produce precise estimates of outcomes for subgroups comprised of half the total sample, as shown in the second row.

Table B.2. Student Sample Size and Precision

	Sample Size	Half-Width of Confidence Interval of a Proportion of 0.50		Half-Width of Confidence Interval of a Proportion of 0.10	
		ICC=0.10	ICC=0.15	ICC=0.10	ICC=0.15
Total sample	183,624	0.021	0.025	0.012	0.015
50 percent subgroup	91,812	0.029	0.036	0.018	0.022

Note: The confidence intervals are based on a 95 percent probability level. The confidence intervals are based on the effective sample size, which is equal to the sample size divided by the design effect.

Table B.3 presents estimated minimum detectable effects (MDEs) under a variety of assumptions. The study computes MDEs using the same assumptions about ICCs as above. The study also examines two assumptions about the amount of between-school variance that is explained by the school-level control variables in the HLM model. The optimistic assumption is 20 percent (R^2=0.2), as is often found in educational evaluations (see Schochet 2005), while the pessimistic assumption is 0 percent.

The MDE estimates suggest the study will be able to detect differences in the receipt of Federal financial aid, the amount of aid, and post-secondary enrollment of between 5 and 8 percent of the standard deviation of each outcome. These estimates are small relative to the precision standard of 10 to 20 percent of a standard deviation that is frequently used for studies of student achievements.

Table B.3. Minimum Detectable Effects

Assumptions		Minimum Detectable Effects (MDEs)
ICC = 0.1	R^2=0.2	0.057
ICC = 0.1	R^2=0	0.063
ICC = 0.15	R^2=0.2	0.069
ICC = 0.15	R^2=0	0.077

Note: MDEs are reported in standard deviation units, and are computed for a two-tailed test with alpha of 0.05 and 80 percent power, assuming sample sizes of 818 schools per district and 224 seniors per school. The table specifies assumptions where ICC represents the school-level intraclass correlation coefficient and R^2 represents the proportion of the between-school variance that is explained by the regression model.

MDE = Minimum detectable effect.

Minimum detectable impacts offer another perspective on the precision of estimates that the study can produce. Table B.4 presents the minimum detectable impacts (MDIs) of the estimation procedure, using the same assumptions as in Table B.3. The mean of FAFSA completion among students is based on 2004 ED data as reported by Finaid.org. The mean of the receipt of Federal financial aid by high school seniors is estimated as the percent of 2006-2007 high school graduates enrolled in a 4-year institution times the percent of undergraduates who receive any Federal financial aid (Digest of Education Statistics 2010). The estimate assumes that all those who receive an offer of Federal financial aid graduate high school and attend a 4-year institution and that rates of Federal aid receipt are constant over time. The mean of enrollment in postsecondary education is based on the percent of 2007-2008 high school graduates attending degree-granting institutions in 2008 (Digest of Education Statistics 2010). As all outcomes are binary, standard deviations can be estimated as $\sqrt{p(1-p)}$ where p denotes the mean.

The power calculations in Table B.4 show that the estimation procedure can detect meaningful differences between the treatment and control groups. Depending on the specific assumptions used, the procedure is powered to detect a difference of between 2.8-3.8 percentage points in the probability of FAFSA completion, 2.0-2.7 percentage points in the probability of receiving Federal financial aid, and 2.7-3.7 percentage points in the probability of enrollment in postsecondary education. In case the estimates of mean outcomes prove not to be accurate, the table also presents MDIs for an outcome with a mean of 50 percent. An outcome with this mean produces an upper bound on estimated MDIs. These upper bound MDIs range from 2.8-3.8 percentage points.

Table B.4. Minimum Detectable Impacts

		N	Minimum Detectable Impact			
Outcome	Mean	Standard Deviation	ICC=0.1, R^2=0.2	ICC=0.1, R^2=0.0	ICC=0.15, R^2=0.2	ICC=0.15, R^2=0.0
Complete FAFSA	0.570	0.495	0.028	0.031	0.034	0.038
Receive Federal student aid	0.190	0.393	0.022	0.025	0.027	0.030
Enroll in postsecondary education	0.638	0.481	0.027	0.030	0.033	0.037
Any 50 percent outcome	0.500	0.500	0.028	0.032	0.034	0.038

Notes: MDIs are computed for a two-tailed test with alpha of 0.05 and 80 percent power, assuming sample sizes of 818 schools per district and 224 seniors per school. The table specifies assumptions where ICC represents the school-level intraclass correlation coefficient and R^2 represents the proportion of the between-school variance that is explained by the regression model. The MDIs are for differences between the treatment and control groups, where each group is one-half of the sample.

B.3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

The study aims to maximize the completeness of the student rosters using a variety of methods. At the outset, ED required LEAs to meet certain technology standards in order to participate in the study, which will facilitate the transmission of data. Once the study begins, IES will notify LEAs of the request for rosters in advance to allow for more time to identify and address any issues. In addition, provision of the rosters is required in order for LEAs to participate in the FAFSA completion project; failure to submit the rosters could result in the districts' being denied access to student-level FAFSA completion data. As specified in Supporting Statement Part A, we anticipate a response rate for this collection of 100 percent, or from all 80 school districts.

B.4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

This study does not include a survey component or new procedures, so no tests will be conducted.

B.5. B5 Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

The following individuals were consulted on the statistical aspects of the design:

Marsha Silverberg, Institute of Education Sciences, 202-208-7178

Ann Gord0n, Mathematica Policy Research, (609) 275-2318

The following people will be responsible for the data collection and analysis:

Marsha Silverberg, Institute of Education Sciences, 202-208-7178

References

Schochet, Peter Z. Statistical Power for Random Assignment Evaluations of Education Programs, 2009.