Request for OMB Clearance

Evaluating Student Need for Developmental or Remedial Courses

at Postsecondary Education Institutions

(Formerly titled Survey of Placement Tests and Cut-Scores in Higher Education Institutions).

Supporting Statement for Paperwork Reduction Act Submissions

Supporting Statement B: Statistical Methodology for Pilot Test and Full-Scale Survey

B.1.   Respondent Universe for Pilot Study and Full-Scale Survey

The sampling frame or respondent universe, from which the sample of institutions for both the pilot test and the full-scale survey will be drawn, will be constructed from the 2008-09 Institutional Characteristics (IC) component of the Integrated Postsecondary Education data System (IPEDS) maintained by the National Center for Education Statistics (NCES). The 2008-09 IC file contains over 7,100 Title IV postsecondary institutions. However, only 2-year and 4-year degree granting institutions located in the United States with a bachelors' program or lower are eligible for the study. Thus, post graduate institutions offering only a first professional degree will be excluded from the sampling frame. Table B-1 summarizes how the frame of institutions will be constructed. As indicated in the last row of the table, 4,148 institutions meeting the eligibility criteria for the study will be included in the sampling frame, of which 2,458 will be 4-year institutions and 1,690 will be 2-year institutions. Table B-2 summarizes the distribution of eligible institutions by level and control.

Table B-1. Summary of frame construction for pilot test and full-scale survey

Description of institutions in 2008-09 IPEDS IC File	No. of institutions	4-year	2-year	Other
All institutions	7,126	2,906	2,299	1,921
Institutions in United States (excluding outlying territories)	6,961	2,837	2,274	1,850
Institutions in United States offering undergraduate classes	6,649	2,541	2,274	1,834
Non-degree granting	2,377	6	537	1,834
Degree granting	4,272	2,535	1,737	n/a
Non Title IV	45	32	13	n/a
Title IV*	4,227	2,503	1,724	n/a
Administrative unit†	79	45	34	n/a
Non-administrative unit **	4,148	2,458	1,690	n/a

* Includes 4 US Service academies.

† Pending review, all or most of these will be deleted from the sampling frame.

Table B-2. **Distribution of eligible postsecondary institutions to be included in the sampling frame, by level and control**

Level	Control	Number of institutions in sampling frame[*]
4-year	Public	636
	Private, not-for-profit	1,311
	Private, for-profit	472
2-year	Public	1,022
	Private, not-for-profit	92
	Private, for-profit	563
Total institutions (with non missing enrollment)		4,096

[*] The counts given in this table are based on initial tabulations from the 2008-09 IPEDS Institutional Characteristics (IC) file. Fifty-two eligible institutions in the 2008-09 IC file have missing values for enrollment and are not included in this table. However, such institutions will be included in the final sampling frame and given appropriate chances of selection.

B.2. Statistical Methodology

Sample Design and Selection for the Pilot Test

Since a primary purpose of the pilot test is to identify potential data collection problems during the full-scale survey, the goal of the sample design will be to ensure that (a) a diverse cross-section of postsecondary instititions in the United States are included in the sample, and (b) the sample sizes for major subgroups defined by type of control and level are sufficiently large to provide a range of responses that can usefully inform the design of the questionnaire for the full-scale survey. Thus, in addition to institutional control and level, the sample will be stratified, to the extent feasible, by characteristics such as size of institution and highest level of degree offering. Within a sampling stratum, institutions will be selected at rates that are approximately proportional to the square root of their enrollment. The proposed sample size of 120 institutions for the pilot test will yield about 100 respondents assuming a response rate of 85 percent. Although measures of selectivity will not be used as a stratifier in sample selection, the sample of 4-year institutions to be included in the pilot test is expected to cover a broad range of institutions with respect to selectivity. Table B-3 summarizes the proposed sample sizes by major subgroups defined by institutional level and control.

Table B-3. **Proposed target sample sizes for the pilot test, by control and level**

Subgroup	Number sampled	Respondents[*]	
Total sample	*120*	*102*	
Public	*60*	*51*	
4-year		30	26
2-year		30	26
Private, not-for-profit	*40*	*34*	
4-year		30	26
2-year		10	9
Private, for-profit	*20*	*17*	
4-year		10	9
2-year		10	9

Note: Details may not sum to total due to rounding.

Sample Design and Selection for the Full-Scale Survey

Sample Allocation for the Full-Scale Survey

One of the goals of the sample design is to ensure that the sample sizes are sufficiently large to enable the detection of significant differences of 0.2 standard deviations in assessment score scales between major subgroups to the extent feasible. Table B-4 summarizes the minimum detectable effect size (i.e., the difference expressed as a multiple of the standard deviation) between subgroups of various sizes and a range of design effects. The design effects will arise primarily because of the varying probabilities of selection with which the institutions will be drawn into the sample. Depending on the magnitude of the design effect, it can be seen that each subgroup would have to include between 400-500 respondents to be able to detect an effect size of 0.20 (i.e., 0.20 standard deviations). Larger effect sizes of 0.40 can be detected with subgroups sample sizes of 100-200. Only very large effect sizes of 0.60 or greater can be detected with subgroup sample sizes as small as 50 to 60.

Table B-4.   Minimum detectable effect sizes[*] between subgroups for selected sample sizes and design effects

	Design effect		
Subgroup sample size[†]	**1.10**	**1.20**	**1.30**
800	0.15	0.15	0.16
700	0.16	0.16	0.17
600	0.17	0.18	0.18
500	0.19	0.19	0.20
400	0.21	0.22	0.23
300	0.24	0.25	0.26
200	0.29	0.31	0.32
150	0.34	0.35	0.37
100	0.42	0.43	0.45
60	0.54	0.56	0.58

* Entries are detectable effect sizes for a two-sided test with alpha = 0.05 and power = 0.80.

† Sample size per subgroup.

In addition to the specified precision goals, disaggregation of the sample into subgroups defined by control and level will also be important for analysis. However, this will not be possible for very small subgroups without unduly reducing the overall efficiency of the sample. For example, it can be seen in Table B-2 that most of the private 2-year institutions are for-profit institutions. Since the private 2-year institutions tend to be small (fewer than 500 students), it does not seem useful to separate the 2-year private for-profit and not-for-profit institutions from either a statistical or policy perspective. Among the private 4-year institutions, on the other hand, there are several extremely large for-profit institutions (e.g., DeVry, University of Phoenix) that are inherently different from the traditional private not-for-profit institutions. In this case, separating the for-profit institutions from the others would be desirable to provide meaningful comparisons.

In view of the above considerations, we propose to select the numbers of institutions indicated in Table B-5. The proposed sample sizes assume an overall 85 percent response rate and are designed to permit separate analysis of most major subgroups, while giving appropriate representation in the sample of the rarer subgroups. For example, under the proposed design, it will be possible to detect an effect size of 0.20 for comparisons between public vs. private not-for-profit institutions, between public 4-year and public 2-year institutions, and between public 4-year and private not-for-profit 4-year institutions. Larger effect sizes of 0.40 can be detected between smaller subgroups; e.g., comparisons between private not-for-profit and private for-profit institutions or between private 4-year not-for-profit institutions and private 4-year for-profit institutions.

Table B-5.   Proposed target sample sizes for the full-scale survey, by control and level

Subgroup	Number sampled	Respondents[*]
Total sample	1,668	1,418
Public	*936*	*796*
4-year	468	398
2-year	468	398
Private, not-for-profit	*490*	*417*
4-year	468	398
2-year	22	19
Private, for-profit	*242*	*206*
4-year	160	136
2-year	82	70

[*] Assumes 85 percent response rate.

Note: Details may not sum to total due to rounding.

Sample Selection Procedures for the Full-Scale Survey

To permit the efficient allocation of the sample to meet the goals stated above, the sampling frame will be stratified by size of institution, level, and type-of-control, along with other institutional characteristics such as highest level of degree offering. Implicit stratification within major strata by geography and minority status of institution will also be employed to ensure that all regions and types of institutions are appropriately represented in the sample. Within a stratum we will sample institutions at rates that are approximately proportional to the square root of their enrollment. Table B-6 shows the distribution of the proposed sample by each of the 39 sampling strata using this method. A similar design has been used to select institutions for the Postsecondary Quick Information System (PEQIS) and is

expected to be efficient for the proposed survey on the use of placement tests and cut scores. Note that institutions with enrollments of 10,000 or more are selected with certainty under this design.

Table B-6. Distribution of postsecondary institutions to be included in the sampling frame

Stratum	Level	Type of Control	Highest level of offering	Enrollment size class	Number of institutions in frame*	Number of institutions in sample
1	4-year	Public	Doctorate	<3,000	20	7
2				3,000 to 9,999	63	42
3				10,000 to 19,999	94	94
4				20,000+	106	106
5			Masters	<3,000	57	19
6				3,000 to 4,999	51	34
7				5,000 to 9,999	87	65
8				10,000+	42	42
9			Bachelors	<1,000	19	5
10				1,000 to 2,999	45	15
11				3,000 to 9,999	38	25
12				10,000+	14	14
13	4-year	Private (not-for-profit)	Doctorate	<500	47	5
14				500 to 2,999		33
					100	
15				3,000 to 9,999	152	91
16				10,000+	53	53
17			Masters	<500	81	9
18				500 to 999	92	31
19				1,000 to 2,999	269	90
20				3,000 to 9,999	92	55
21				10,000+	8	8
22			Bachelors	<500	151	17
23				500 to 999	107	21
24				1,000 to 2,999	146	49
25				3,000+	13	7
26	4-year	Private (for-profit)	—	<100	28	6
27				100 to 499	163	41
28				500+	281	112
29	2-year	Public	—	<1,000	113	16
30				1,000 to 2,999	288	72
31				3,000 to 4,999	201	80
32				5,000 to 9,999	241	121
33				10,000+	179	179

See notes at end of table.

Table B-6. Distribution of postsecondary institutions to be included in the sampling frame (continued)

Stratum	Level	Type of Control	Highest level of offering	Enrollment size class	Number of institutions in frame*	Number of institutions in sample
34	2 year	Private (nonprofit)	—	<500	67	13
35				500 to 999	20	7
36				1,000+	5	3
37	2 year	Private (for profit)	—	<500	369	41
38				500 to 999	154	31
39				1,000+	40	10
Total					4,096	1,668

* The population counts given in this table are based on initial tabulations from the 2008-09 IPEDS Institutional Characteristics (IC) file. Fifty-two eligible institutions in the 2008-09 IC file have missing values for enrollment and are not included in this table. However, such institutions will be included in the final sampling frame and given appropriate chances of selection.

Note: Details may not sum to total due to rounding.

For the 4-year institutions, measures of selectivity are key factors in the proper interpretation of the results from the survey. Thus, it is important that the sample include institutions covering a broad range of selectiveness. Applying the algorithm developed by Cunningham (2005), [1] we have estimated the numbers of institutions to be included in the sample by various categories of selectivity. The numbers of institutions in the frame and the corresponding expected sample sizes are shown in Table B-7. As can be seen in this table, a wide range of institutions with respect to selectivity is expected to be included in the sample. Note that other methods of defining selectivity can also be employed. The results in table B-7 are intended to provide rough orders of magnitude of the sample sizes to be expected under the proposed sample design.

[1] Cunningham, A.F. (2005). *Changes in Patterns of Prices and Financial Aid* (NCES 2006-153). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Table B-7. Number of institutions in the frame and expected sample sizes, by selectivity among 4-year institutions

Selectivity (based on Cunningham method)	Level	Number of institutions in frame	Number of institutions in sample
Most/Very Selective	Public	106	93
	Private	249	129
Moderately Selective	Public	316	248
	Private	560	213
Minimally Selective	Public	96	58
	Private	523	155
Open Admissions	Public	118	69
	Private	451	131
Total		2,419	1,095

Note: Details may not sum to total due to rounding.

Expected Levels of Precision for the Full-Scale Survey

Table B-8 summarizes the approximate sample sizes and standard errors to be expected under the proposed design for selected subgroups. Since the sample sizes in Table B-8 are based on preliminary tabulations of the 2008-09 IC file, the actual sample sizes to be achieved may differ from those shown. Also, note that the sample sizes represent the expected numbers of completed questionnaires with eligible institutions, and not the initial numbers of institutions to be sampled. The standard errors in Table B-8 reflect an overall design effect of 1.3. The design effect arises primarily from the use of variable sampling fractions across the major sampling strata and differential nonresponse weighting adjustments (see description under **Estimation and Calculation of Sampling Errors**). In particular, the design effect reflects the fact that under the proposed stratified design, large institutions will be sampled at relatively higher rates (i.e., have smaller sampling weights) than small institutions. In fact, some very large institutions will be selected with certainty or near certainty. Subgroups consisting solely of such institutions will have negligible standard errors because the finite population correction (fpc) for these subgroups will be 0 or close to 0. Hence, the entries in the table should be viewed as rough upper bounds on the standard errors to be expected from the survey.

The standard errors in Table B-8 can be converted to 95 percent confidence bounds by multiplying the entries by 2. For example, an estimated proportion of the order of 20 percent (P = 0.20) for public 4-year institutions will be subject to a margin of error of ±4.6 percent at the 95 percent confidence level. Similarly, an estimated proportion of the order of 50 percent (P = 0.50) for 4-year institutions requiring test scores for admission will be subject to a margin of error of ±4.4 percent at the 95 percent confidence level.

Table B-8. Expected standard error of an estimated proportion under proposed design for selected analytic domains, by control and level

Domain (subset)	Expected sample size*	Standard error† of an estimated proportion equal to ...		
		P = 0.20	P = .33	P = .50
Total	1,418	0.012	0.014	0.015
Public	*796*	*0.016*	*0.019*	*0.020*
4-year	398	0.023	0.027	0.029
2-year	398	0.023	0.027	0.029
Private, 4-year	*534*	*0.020*	*0.023*	*0.025*
Not-for-profit	399	0.023	0.027	0.029
For-profit	135	0.039	0.046	0.049
Private, 2-year	89	0.048	0.057	0.060
4-year schools	*931*	*0.015*	*0.018*	*0.019*
Requires test scores	620	0.018	0.022	0.023
Has open admissions	144	0.038	0.045	0.048
Most/Very Selective**	188	0.023	0.027	0.028
Moderately Selective	392	0.023	0.027	0.029
Minimally Selective	181	0.034	0.040	0.042
Open Admissions	170	0.035	0.041	0.044

* Expected number of responding eligible institutions, assuming response rate of 85 percent. The standard errors given in this table are given for illustration. Actual standard errors may differ from those shown.

† Assumes unequal weighting design effect of 1.3. For subgroups consisting of institutions selected with certainty, the standard errors will be smaller than those shown.

** Standard errors include an approximate finite population correction to reflect the fact these institutions will be selected at relatively high rates.

Estimation and Calculation of Sampling Errors for the Full-Scale Survey

For estimation purposes, sampling weights reflecting the overall probabilities of selection and adjustments for nonresponse will be attached to each data record. To properly reflect the complex features of the sample design, standard errors of the survey-based estimates will be calculated using jackknife replication. Under the jackknife replication approach, 50 subsamples or "replicates" will be formed in a way that preserves the basic features of the full sample design. A set of estimation weights (referred to as "replicate weights") will then be constructed for each jackknife replicate. Using the full sample weights and the replicate weights, estimates of any survey statistic can be calculated for the full sample and each of the 50 jackknife replicates. The variability of the replicate estimates is used to obtain a measure of the variance (standard error) of the survey statistic. Previous surveys, using similar sample designs, have yielded relative standard errors (i.e., coefficients of variation) in the range of 2 to 10 percent for most national estimates. Similar results are expected for this survey.

B.3.   Methods for Maximizing the Response Rate for the Pilot Test and Full-Scale Survey

For both the pilot test and full-scale survey, the President's office of each sampled postsecondary institution will be contacted to solicit the institution's participation in the survey and to identify the appropriate survey respondent. Survey respondents will have the option of completing the survey on paper or on the web. The paper version of the questionnaire is limited to four pages of questions. Telephone followup for nonresponse, which will be conducted by Westat staff, will begin about 3 weeks after mailout for each type of collection. Experienced telephone interviewers will be trained to conduct followup and will be monitored by Westat supervisory personnel during all interviewing hours. Collection procedures will follow standard methods developed by Westat for previous surveys on postsecondary institutions.

B.4.   Tests of Procedures and Methods

A pretest of the survey was conducted with seven respondents to determine what problems respondents might have in providing the requested information and to make appropriate changes to the questionnaire. Responses and comments on the questionnaire were collected by fax and telephone during the pretest, and the results were summarized as part of the documentation for the study.

A pilot test will be conducted with a sample of 120 postsecondary institutions. Respondents will have the option of completing a web version that will be accessed through the Internet or a traditional paper and pencil questionnaire. The pilot test data will be analyzed to identify potential problems that may emerge from data collection during the full-scale survey. In addition, findings from the pilot test will provide insights into the extent to which institutions use various tests, either independently or in combination with other evaluation criteria to identify student need for remediation in mathematics and reading. Results of the pilot test, including survey management procedures for contacting respondents and obtaining cooperation, will be summarized as part of the documentation for the study.

B.5.   Reviewing Statisticians for the Pilot Test and Full-Scale Survey

Westat's statisticians, Adam Chu (telephone: 301-251-4326) and Greg Norman were consulted about the statistical aspects of the design.