

Paperwork Reduction Act Submission Supporting Statement

Annual Mandatory Collection of Elementary and Secondary Education Data through *EDFacts*

Part B. Collections of Information Employing Statistical Methods.

The agency should be prepared to justify its decision not to use statistical methods in any case where such methods might reduce burden or improve accuracy of results. When Item 17 on Form 83-I is checked “yes,” the following documentation should be included in the Supporting Statement to the extent that it applies to the methods proposed:

The SY 2009-10 CRDC will survey a sample of local education agencies (LEAs) and all of the schools in those LEAs. The selection of the sample and the analysis of the data after the collection both employ statistical methods.

- 1. Describe the potential respondent universe (including a numerical estimate) and any sampling or other respondent selection method 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 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 potential respondent universe for the SY 2009-10 CRDC is the universe of elementary and secondary public school agencies in the United States, which comprises approximately 18,000 LEAs. However, for drawing the CRDC sample, ED has used a universe that is limited to certain types of LEAs. The primary source data are collected by *EDFacts* and compiled and maintained by the National Center for Education Statistics (NCES) as the Common Core of Data (CCD) and only CCD types 1, 2, 3, and 4 agencies are included in the CRDC universe¹. Selected charter schools, juvenile justice agencies that provide education services, and state operated programs for special populations of students (such as schools for the deaf and schools for the blind) are added to the universe if they are not already in the CCD list. This CRDC universe includes approximately 14,200 LEAs.

The CRDC, which has collected data from school districts (LEAs) since 1968, has generally included a sample of approximately 6,000 school districts. In 1976 and 2000, data were collected from a universe of school districts. An initial goal for OCR was for the

¹ Type 1 is regular school district, type 2 is supervisory union component, type 3 is administrative center of a supervisory union, and type 4 is Regional Educational Service Agency. Types 3 and 4 are included in the universe only if they operate schools with students that attend for at least half day.

SY 2009-10 CRDC to be a third universal collection. For funding reasons this was not possible. As an alternative, OCR explored the possibility of increasing the sample size.

OCR determined, using the Common Core of Data (CCD), that by increasing the sample size from 6,000 to 7,000 school districts, it would be possible to sample all districts with more than 3,000 students. This will provide a significantly more comprehensive picture of the access to equal educational opportunity for all students. These data will be used extensively by OCR and by other offices in ED for programmatic and policy purposes.

By increasing the sample to 7,000, OCR's enforcement offices will have additional data upon which to decide where to focus enforcement and technical assistance efforts, as well as have additional data on districts for which OCR receives complaints from the public. Having data from a larger number of districts should expedite investigations in those districts.

Increasing the sample size is particularly important at this time. Since it has been three years since the last administration of the CRDC, it is critical to have updated data. By increasing the sample size, it will also be possible to include the 51 state agencies that operate state juvenile justice facilities. These agencies have not previously been included in the CRDC, so adding them will provide a valuable source of data.

Approximate counts of the CRDC universe and the sample for the SY 2009-10 CRDC by strata are shown in the table below.

Enrollment stratum	Total in Universe	Total in Sample
1 – 300*	3,044	610
301 – 3,000	7,763	2,887
3001 – 5,000	1,456	1,456
5,001-25,000	1,701	1,701
25001 +	275	275
Total	14,239	6,929

The expected response rate for the SY 2009-10 CRDC is between 95% and 98% of the sample LEAs, which would be approximately 46% to 48% of the LEAs in the universe.

The response rate on the 2006 CRDC, which was the most recent administration of this survey, was 100% of the eligible LEAs in the sample, which was 42% of the LEAs in the universe.

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

- **Statistical methodology for stratification and sample selection.**

The sample for the SY 2009-10 CRDC will include the sample originally drawn for the planned 2008 CRDC (which was never conducted) plus additional specified LEAs as described below.

The potential respondent universe for the 2008 sample draw was the universe of public elementary and secondary school agencies collected by *EDFacts* and compiled and maintained by the National Center for Education Statistics (NCES) as the Common Core of Data (CCD). Using a pre-release version of the SY 2006-07 CCD, the sampling frame

was developed; it included districts with membership whose agency-type code in the CCD is

- 1—Local school district,
- 2—Local school district component of a supervisory union,
- 3—Administrative center of a supervisory union (only those that operate one or more schools), and
- 4—Regional education service agency.

Districts with no membership or missing membership at the district level were generally excluded, except in 441 special cases, such as where membership data were available for the associated schools. Additionally, some education agencies not in the above types were included in the sampling frame under the “selected with certainty” provision described below.

There were 14,204 LEAs in the sampling frame and OCR specified a target sample size of 6,000.

The sample drawn for 2008 included both LEAs that were “selected with certainty” and LEAs selected by probability through a stratified sampling design.

The following LEAs were “selected with certainty,” that is, all LEAs that meet any one of these criteria were included in the sample:

1. All LEAs in states with 25 or fewer regular public school LEAs.
2. All LEAs with 25,000 or more students.
3. All LEAs currently or recently under Department of Justice court orders.
4. Selected state operated educational programs (such as schools for the deaf, schools for the blind).
5. All LEAs that were granted a deferral for the most recent previous CRDC.
6. Other specified LEAs identified by ED, including selected LEAs identified by ED’s Office for Civil Rights and selected juvenile justice education entities, chosen in collaboration with the Department of Justice.

In the initial sample, 978 distinct LEAs were selected with certainty. Many of them met more than one of the above criteria.

An additional 5,020 LEAs were selected by probability using a multi-state rolling stratified sample design. This approach used strata divided by size of district, and sub-strata of high/low proportion of minority students. The percentage of LEAs selected varied by state, from a low of 34% selection to a high of 48.5% selection, inversely related to a factor based on the number of LEAs and the enrollment (*i.e.*, states with fewer LEAs and students generally had a larger percentage selected to ensure adequate representation for statistical reliability). Procedures were applied for minimizing the overlap with LEAs that had been in the sample for the previous (2006) CRDC.

The certainty districts were removed from the sampling frame and the sample was drawn according to the above design from the reduced sampling frame.

The following table shows the composition of the universe and sample planned for 2008.

Enrollment stratum	Total in Universe	SAMPLE					
		Total in Sample	Regular LEA (CCD Type 1, 2, or 3)		RESA (CCD Type 4)		Other (CCD Type 5 or 7)
			Selected with Certainty	Probability sample	Selected with Certainty	Probability sample	Selected with Certainty
1 – 300*	3,044	600	8	505	3	21	63
301 – 3,000	7,763	2,846	223	2,553	0	61	9
3001 – 5,000	1,450	842	111	722	0	9	0
5,001-25,000	1,672	1,435	284	1,418	0	1	2
25001 +	275	275	275	0	0	0	0
Total	14,204	5,998	901	4,928	3	92	74

*Includes LEAs with zero or missing membership

Since the above sample was drawn, but not used, it will be the nucleus of the sample for the SY 2009-10 CRDC. To be more inclusive, and yet limit burden, ED will add about 1,000 LEAs to the 2008 sample by the following two expansions of the group of LEAs selected with certainty:

- Expand from all LEAs with 25,000 or more students to all LEAs with 3,000 or more students.
- Expand from selected juvenile justice agencies to all state-level juvenile justice agencies.

Additionally, the 2008 sample will be updated from the most recently available CCD to modify for closed, merged, or split LEAs. The exact counts for the sample will be determined as the sample goes through this updating process.

▪ **Estimation procedure.**

Respondents are asked to submit exact counts, not estimates.

▪ **Degree of accuracy needed for the purpose described in the justification.**

The Department’s Office for Civil Rights (OCR) is responsible for implementing civil rights laws, and the primary purpose of the Civil Rights Data Collection (CRDC) is to provide data to support OCR’s compliance and enforcement activities through use of the data from the surveyed LEAs. These data meet the degree of accuracy necessary. The methodology used for selecting the sample and in-depth analysis of non-respondent LEAs and/or schools and missing items within surveys ensure that the state and national projections yield statistically sound conclusions for the universe.

▪ **Unusual problems requiring specialized sampling procedures, and**

The Department has not found unusual problems requiring specialized sampling procedures beyond those described above.

▪ **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

The Department has historically conducted this survey biennially to reduce burden. This

SY 2009-10 survey is the first in three years because the planned 2008 survey was not conducted because of a delay in approval of the Department's FY 2009 budget.

- 3. Describe methods to maximize response 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.**

Historically the civil rights survey has had a very high response rate. In 2000, the predecessor Elementary and Secondary (E&S) Survey was sent to a universe of all school districts and schools in the United States. The overall response rates were 97% of all school districts and 99% of all schools. The overall response rates for the 2002 E&S Survey were 98% for school districts and 98% for schools. For the 2004 Civil Rights Data Collection, the response rates, including partial respondents to the data collection, were approximately 97% of all districts, and 97% of all schools. The 2006 Civil Rights Data Collection achieved an unprecedented 100% response rate for school districts and a 99.6 % response rate for schools. If school districts fail to respond in a timely manner, the contractor for the data collection, with assistance from OCR as necessary, provides extensive outreach and assistance to the greatest extent possible until the districts respond, or the final deadline for accepting data has passed.

As has been done in the past, after the SY 2009-10 collection, statistical projections for most items will be calculated for the nation and for each state, basing the calculations on the methodology used for selecting the sample and in-depth analysis of non-respondent LEAs and/or schools and missing items within surveys to yield statistically sound conclusions for the universe. Data will be weighted to compensate for schools that did not provide usable data and values will be imputed on an item-by-item basis to compensate for item non-response. Such analyses for the 2004 and 2006 CRDCs are available on ED's Web site at: <http://www.ed.gov/about/offices/list/ocr/data.html?src=rt>

- 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 tests may be submitted for approval separately or in combination with the main collection of information.**

Content. The Department conducted five record keeping site visits. The respondents were selected to encompass a group that varies by urbanity, region, size, and level of reporting (LEA vs. state agency). Respondents were asked to provide feedback on the definitions and on any general issues they wished to raise about the CRDC or their recordkeeping practices. They were also provided a mock-up of the survey and asked whether they kept the type of requested data.

Specifically, for each question on the survey, respondents were asked: 1) Can you provide this information using your existing recordkeeping systems? 2) If yes, can you easily provide this data, or will you be required to perform manual or other non-routine operations in order to provide it? Please describe the difficulty of those alternative methods; and 3) If you cannot provide the information requested, why not?

Respondents reported no difficulty providing data for the large majority of the current and proposed CRDC questions using their existing recordkeeping systems. They also reported that the administration of the survey in two parts would not cause difficulties.

To the extent respondents expressed difficulty with any of the requested data elements, it was with 1) responding to SAT and ACT test participation questions, since this information must be gathered from data reported to LEAs by outside agencies; and 2) reporting on harassment and bullying, for which some LEAs do not have reporting mechanisms in place.

Sample. No specific tests of the sampling procedures will be undertaken because of the successful history of these procedures. The rolling stratified sample design with certainty selections was used to select the samples for the 1998, 2002, 2004, and 2006 civil rights surveys. (The 2000 CRD was a universal collection and thus no sample was drawn.)

Survey Tool. A new more user-friendly online survey tool is being developed to collect the 2009-10 CRDC data. The 2009-10 CRDC will be conducted in two phases. Piloting will be undertaken for both phases. ED plans to undertake two pilot tests of up to 50 LEAs as part of the testing of the online survey tool (one pilot test for each part of the survey). The Department will identify LEAs for participation. The pilot tests would be conducted by the software development contractor as follows. The pilot LEAs would be asked in advance to review each section of the survey within the first five days it is open and to advise ED of issues and provide ED with suggested improvements. The contractor will provide all communications and instructions, including a mechanism for testers to report their questions and issues. At the completion of the pilot testing period, the contractor will compile all suggestions, annotated for level of criticality and feasibility and provide these to ED. From this information, ED will identify feasible corrections and the contractor will make these corrections within the first month that the survey is open, thus providing for rapid turnaround of improvements. ED will maintain the list of suggestions that are not feasible in the short term for possible implementation in future surveys.

ED is also planning to invite 10-15 school districts to participate in an initial pilot prior to the official opening date of the Part 1 website. A similar initial pilot will also be conducted for Part 2. The procedures for these initial pilots will be similar to the larger pilots. Feedback received from users during the conduct of the Part 1 of the survey will also inform the development of the Part 2 survey tool.

Single File Submission. A new single file submission (SFS) is being developed for the 2009-10 CRDC. We are planning to post the specifications for the Part 1 SFS on the CRDC website approximately two weeks before the data collection begins. We will also

arrange for a small group of school districts to pilot the SFS and provide feedback to ED. A similar process will be followed for Part 2. Input from Part 1 SFS users will also inform the development of the Part 2 SFS.

5. 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 persons who will actually collect and/or analyze the information for the agency.

The following persons prepared the 2008 sample, under contract to the Department:

Dr. Adam Chu, Statistician, Westat, Inc., Rockville, Maryland: (301) 251-4326

Yong Lee, Westat, Inc., Rockville, Maryland: (301) 251-4326

The SY 2009-10 CRDC is conducted in collaboration between the Office for Civil Rights, Russlynn Ali, Assistant Secretary; and the Performance Information Management Service, Ross Santy, Director: (202) 401-1259. The data will be collected through *EDFacts*, managed by the Performance Information Management Service.

Technical services for this collection and the analysis of data will be provided by the Department's contractor for *EDFacts*: 2020 Company, Inc., LLC, Haresh Bhungalia, Co-Founder, 800-327-9015.