127919

UNITED STATES GENERAL ACCOUNTING OFFICE
Washington, D.C. 20548

FOR RELEASE ON DELIVERY
EXPECTED AT 10:00 a.m.
THURSDAY, JULY 25, 1985

STATEMENT OF
GENE L. DODARO
ASSOCIATE DIRECTOR, GENERAL GOVERNMENT DIVISION
BEFORE THE
SUBCOMMITTEE ON CENSUS AND POPULATION
COMMITTEE ON POST OFFICE AND CIVIL SERVICE
HOUSE OF REPRESENTATIVES
ON
THE CENSUS BUREAU'S PREPARATIONS FOR THE
1990 DECENNIAL CENSUS

033208

Mr. Chairman and Members of the Subcommittee:

I am pleased to appear today to discuss the Census Bureau's preparations for the 1990 Decennial Census. I am accompanied by Mr. Jack Kaufman, who is responsible for our audits at the Census Bureau. My comments will focus on the Bureau's automation activities and pretests preparatory for the 1990 census. I also will comment on the proposed residency rules for the 1990 census and briefly discuss the recommendations made by the Commerce's Office of Inspector General on the Bureau's organization and automation plans.

The Bureau's preparations, including tests to date and planned milestone dates for important decisions, are not conducive to a cost efficient 1990 Decennial Census. We share your concern that the 1990 census not become a census of lost opportunities. On the bright side, however, some options are available for the Bureau to get back on track, but time is quickly running out.

Although it may appear that there is ample time to plan wisely for the 1990 census, in fact there is only a limited time to thoroughly test significant changes or modifications to census forms, equipment, and procedures. For all practical purposes the Bureau must complete these evaluations well before its 1988 dress rehearsal. Developing census forms and procedures for the 1988 rehearsal will take about one year or longer. Thus, in early 1987 the Bureau must decide almost exactly how it will conduct the 1990 Decennial Census. The importance of the 1985 and 1986 pretests and special purpose

tests cannot be overstated. These tests will be the last ones completed in time to precipitate major changes in the 1990 census.

With this timetable in mind, we will highlight our observations and suggestions to help expedite the Bureau's planning and improve its preparatory efforts.

AUTOMATION OF QUESTIONNAIRE PROCESSING

Early indications are that some concepts tested in the 1985 pretests have potential for 1990 census application. These successes include accounting for questionnaires as they are received, early capture of data into computer files, and automated review of questionnaires for determining the completeness and consistency of responses. However, as discussed in our April 18, 1985, testimony before this subcommittee, we remain concerned that the Bureau is waiting until September 1986 to make an equipment decision. An early 1986 decision would be more realistic, given the Bureau's procurement experience. Historically, the Bureau has taken 4 to 5 years to obtain automated equipment after its need has been identified, yet the decision on the equipment to be used for the 1990 census is not scheduled until late 1986.

Equipment alternatives

The Bureau is actively considering three types of data entry equipment and has incorporated them in its pretests. These three types are:

--A modified version of the film optical sensing device/FACT 80 used in the last census.

--Large optical mark reader/scanners.

--Data entry keying.

The FACT 80 and the optical mark reader will be tested in the 1986 Los Angeles pretest, and data entry keying will be tested in the 1986 Mississippi pretest. Both optical mark reader equipment and data keying were used in the 1985 pretests. Data keying will also be used in Los Angeles for name capture for the post enumeration survey and for entering data from some of the long forms. The existing optical mark reader does not have page turner capability needed to capture the data from the currently designed long forms.

FACT 80

The FACT system, developed jointly by the Census Bureau and the Bureau of Standards, has been used in every decennial census since the 1950's. The system incorporates an automated camera for microfilming, a device for turning pages, and a film optical sensing device for input to computers (FOSDIC). The current cost estimate for producing an upgraded version of the 1980 FACT system is $75,000 per camera and $75,000 per FOSDIC. The number of complete, modified FACT 80s needed would depend on the deployment and turn around speed needed. Under one current FACT 80 deployment scenario, the Bureau would need about 60 automated cameras and 36 FOSDICs for the 1990 census. Under this scenario, acquiring the equipment through either in-house assembly or contractor fabrication would cost about $5.9 million and would take several years.

Optical mark reader

The optical mark reader was tested in Jeffersonville, Indiana, as part of the 1985 Tampa pretest. The Jeffersonville personnel found it easy to use, and the raw data from the test shows that the machine records marks on the questionnaire accurately. Despite six breakdowns requiring minor repair, the optical mark reader performed well.

The machine operation does, however, require that the questionnaires be stored in a climate-controlled environment and not be exposed to high levels of humidity. On several occasions during testing, the reader failed to read properly because of improper questionnaire storage. This problem was resolved when the questionnaires were placed in the proper environment for several hours.

Another problem noted was the higher-than-usual nonresponse rates for some questions. Because the optical mark reader was designed to process a much smaller size form than that used in 1980, the questionnaire for the Tampa 1985 pretest was physically reduced in size while still containing all the 1980 short form questions. This reduction in form size provided smaller spaces for the questions and responses. Apparently, respondents were confused as to where to place the answer to a question or just overlooked questions. This problem was confirmed by comparing the percentage of nonresponses for the compressed Tampa form versus the more spacious Jersey City form--even for such basic questions as sex and birth which traditionally have very low nonresponse rates. For example,

almost 12 percent of the Tampa respondents did not answer the question on sex, compared to 2 percent of Jersey City respondents. About 17 percent in Tampa did not answer the question on century of birth compared to about 4 percent in Jersey City.

The optical mark reader pretest also did not resolve some Bureau concerns. The Bureau needed to test the reader under heavy workload conditions to determine its reliability in a census environment, but the Tampa pretest did not provide a sufficient volume.

In order to test the reader under a heavier workload, the Bureau devised an optical mark reader "load test" that took place between May 20 and June 4, 1985. During this test, previously processed Tampa questionnaires were run continuously through the reader for 4 to 8 hours daily. On some days, as many as 30,000 forms were reprocessed; during the entire "load test" a total of about 158,000 forms were processed. The optical mark reader again operated very well. It did, however, have two breakdowns; one was corrected with a minor repair and the other required the vendor to dispatch a technician from the Iowa office. What caused this latter breakdown is still uncertain.

Because the optical mark reader used in the Tampa pretest was not designed for decennial census work, the Tampa pretest did not provide definitive information. Therefore, the Bureau plans to test a modified version of the optical mark reader in the 1986 Los Angeles pretest. The modifications being made, such as a change in the equipment's ability to accept a large

size questionnaire, are substantial enough to require an almost total reevaluation.

Bureau-required modifications to the optical mark reader tested in 1985 would allow it to process 11" x 17" forms as contrasted with the 8 1/2" x 11" forms. This appears necessary based on the problems the respondents experienced with the 1985 Tampa pretest forms. The Bureau estimates the cost of each modified optical mark reader unit at about $150,000, after an initial research and development cost of about $2 million. Therefore equipping 18 processing offices with 36 units would cost about $7.4 million. The unit that may be used in the 1986 pretest should be considered only a prototype model.

On June 20, 1985, the Bureau published a notice of intention to acquire a modified version of this optical mark reader from its vendor for testing in the 1986 Los Angeles pretest. Early indications are that several companies may wish to submit proposals. If this occurs, the Bureau will need to evaluate the proposals, which could slow down the planned data processing experiments for the 1986 Los Angeles pretest.

Data entry keying

Data entry keying is the slowest, most error-prone, and least automated of the three types of technologies. It is also the most expensive. In fact, when the Bureau developed the forerunner to the FACT 80 in the 1950's, it recognized that keying was too slow for the massive amount of data collected in a decennial census.

Nevertheless, the Bureau is strongly considering keying machines as the "automated" data capture equipment for 1990. Bureau officials are considering keying because it offers support to several aspects of decennial processing such as name capture of multifamily dwellings to assist in follow-up enumeration of questionnaire nonrespondents and for assessing the accuracy and completeness of the population counts (coverage evaluation activities). (Optical mark reader technology does not provide this capability). According to one Bureau plan, as many as 35,000 machines would be purchased at a unit cost, including software and hardware needed to operate the keying stations, of about $5,000 per machine or a total estimated cost of $175 million. This amounts to at least $165 million more than the cost of equipment for the other current data entry options. It should be noted that in recent years, according to a Bureau official, the industry has only produced an average of 20,000 keying machines in a year.

Another point to keep in mind is that the use of keying equipment requires the employment of many operators. Bureau officials estimate that during the census tabulations, the Bureau would have to hire 140,000 keyers to operate the 35,000 machines on two shifts and allow for the expected personnel turnover. The payroll costs of these operators could approximate over $200 million. Additionally, according to Bureau experts such a large number of keyers for short-term temporary work would not be available.

Inadequate time to properly evaluate equipment alternatives

The Bureau is likely to encounter difficulties in completing proper evaluations of each type of equipment in pretest performance prior to the planned selection date. Census day for the 1986 pretests is planned for late March 1986. The Bureau is scheduled to select the 1990 data capture equipment in September 1986. Assuming that the Los Angeles and Mississippi mail response rates are similar to the 56-percent rate in Tampa and 38-percent rate in Jersey City, the pretests are unlikely to be completed prior to mid-June 1986 because of the at least 2-month period needed to complete the field work. As in the case of the Tampa pretest, the Bureau would probably not begin analyzing the raw data from the testing until July 1986. This would leave the Bureau with just 2 to 3 months to complete its analysis of the three types of equipment. Such a time constraint could have an adverse impact on evaluating the optical mark reader, since it is the only equipment that the Bureau has not used extensively.

Bureau decisions on decennial offices and evaluation/adjustment plans could affect equipment decision

The Bureau's choice of equipment will be influenced by the number and locations of its 1990 processing offices and by its decision on coverage evaluation and possible population adjustment. If the Bureau chooses to have a centralized structure, it is likely to choose either the FACT 80 or the

optical mark reader to process the large volume of questionnaires. On the other hand, if the Bureau chooses a more decentralized structure, it becomes more feasible to use the slower data entry keying for the smaller volumes processed at each office. Additionally, the Bureau's current thinking on coverage evaluation/adjustment favors key entry. That entry technique allows the Bureau to enter names on computer files which is important for automated matching (comparing census results to other survey results) procedures. This technique will be tested as part of the 1985 Tampa pretest.

Alternatives not pursued by the Bureau

The Bureau has eliminated some automation alternatives without testing them. For example, optical mark readers can accommodate a multipaged questionnaire (i.e., the long form questionnaire), provided that the form has perforated pages that can be separated for processing. This alternative has not been actively considered by the Bureau. In the Tampa pretest, only the short form was processed using the optical mark reader equipment.

In addition, the desktop optical mark reader, an inexpensive ($15,000 system), easy-to-operate scanner, has been tested for other Bureau applications but not for the decennial census. The Bureau could use the desktop optical mark reader if it reduced the number of questions on the short form and thereby decreased the size of that form. Naturally, the required population questions would remain on the short form. Questions

removed from the short form and considered necessary could be included on the long form. The desktop optical mark readers are fast enough to process large volumes of forms quickly (about 400 per hour) for decentralized data entry, yet inexpensive enough to be acquired in large quantities, and could have a diversity of uses after the census.

Impact of life cycle cost on equipment decision

Bureau officials have indicated to us that they will consider some aspects of life cycle cost during the equipment selection decision process. The purchasing of equipment that has a once-every-10-years application, however, requires unique life-cycle-cost considerations. Bureau officials have indicated that they plan to consider the equipment's

--unit cost,

--overall cost,

--purchase-versus-leasing cost,

--processing capacity, and

--maintainability and reliability.

The Bureau's equipment acquisition objective of processing over 100 million questionnaires in a short time frame must be balanced by the need to consider the long-term cost and potential use of acquired equipment.

Equipment purchased for decennial processing should not end up in storage for 9 years, as was the case with much of FACT 80 equipment used in the last census. The Bureau conducts, on a

continuing basis, various surveys and censuses that could be tailored to make use of equipment acquired for the decennial. Given the rate of technological advancements, it is unlikely that equipment acquired for the 1990 decennial will be the most efficient equipment for the 2000 decennial. Therefore, it becomes very important that any equipment acquired for 1990 census have other Bureau uses during the interim nondecennial years. In addition, by identifying other uses for the equipment, equipment costs can be prorated among the various surveys and censuses.

PRETESTS

Jersey City

A two-stage census approach for administering questionnaires using long and short forms was tested in Jersey City. One half of the Jersey City households received only short forms (first stage) and about 6 weeks later 20 percent of them were sent a long form (second stage) to obtain additional information. In the other half of the city, (non-test portion) 80 percent of the households received a short form and 20 percent received a long form at the same time, similar to the 1980 census.

The two-stage was tested at the urgings of GAO and others to determine if simplifying the basic short form might encourage greater public cooperation and thereby improve the accuracy of the population counts, the primary purpose of the census. Moreover, a shorter, more simplified form would allow quicker

processing and thus more time for Bureau and local officials to review preliminary counts.

As we mentioned in our previous testimonies before your subcommittee in June 1984 and April 1985, we have strong reservations about the size and content of the short form. We believe the short form should be limited to the basic questions needed to obtain an accurate population count. For example, we believe that the questions about plumbing and the value and rent of housing units increase the questionnaire's complexity and thus tend to discourage responses.

Despite the fact that the short form was not as short as we recommended in our report Programs to Reduce the Decennial Census Undercount (GGD-76-72, May 5, 1976), the response rate for the short form was better than the long form--39 percent versus 31 percent. This differential in the mail response rate for the short and long forms was consistent with the experience in the hard to enumerate areas in the 1980 census. The Tampa pretest results were similar. The short form response was 58 percent and the long form response was 48 percent. This differential is important to keep in mind considering the Bureau's latest estimate that it could save $5 to $6 million in the decennial census for every 1 percent of increased questionnaire mail response which would therefore preclude followup activity.

The 1985 Jersey City pretest indicates that there is a greater productivity in the followup enumeration for the short forms than for long forms. Preliminary data shows that

enumerator productivity was about 37 percent higher for short forms than for long forms. This is an important factor not only because of cost consideration but also because of the difficulty in obtaining a sufficient number of competent enumerators in the last census and in the Jersey City pretest.

Another factor favoring a short form is the amount of follow-up needed for unacceptable mailed-back long forms as compared to the short forms. In the 1980 census 36 percent of the mailed-back long forms were considered unacceptable (failed edit) and required follow-up. This contrasted with only 13 percent failed edit for the short form.

The mail-back response for the second stage June 10, 1985, long form was 16 percent, which is considerably lower than the 31-percent response for long forms mailed back in the non-test portion of Jersey City. However, the results of the Jersey City pretests are inconclusive and the test was of limited value for a number of reasons.

The second stage long form repeated 10 questions which the respondents had been asked in the first stage short form. The form also repeated seven questions for each household member. In addition, most of the nonrespondents to the first stage short form were visited by enumerators to obtain the same information requested by the second stage long form. This probably discouraged many of the potential second stage respondents.

The Census Bureau did not attempt to publicize the second stage. The June 10 outreach was limited to a booth at a festival in the city, and the Bureau prepared a press release

which appeared in two local newspapers. The person responsible for outreach said basically there was "nothing out there" in terms of outreach for the second stage, no literature distribution, nothing targeted. Moreover, there was limited upfront publicity about the second stage advising the Jersey City residents that some would be receiving a second questionnaire.

The Bureau did not determine why the first stage response rate of 38 percent was far less than the 60 percent anticipated. A limited test of interviewing the nonrespondents was hurriedly planned for both the Jersey City and Tampa pretests. The sample size goal of 200 for each location was too limited to derive valid statistical results. In fact, the number of actual interviews was 109 in Jersey City and 158 in Tampa.

In addition, the Census Bureau will never really know the reason why the two-stage test failed because it does not intend to find out why the stage-two respondents did not mail back their questionnaires. The Bureau does not plan to do any evaluations; and no interviewing of the stage-two nonrespondents was planned.

Tampa

As previously discussed, a major objective of the 1985 pretest in Tampa was to test the use of optical mark reader equipment. In addition, the Tampa pretest included other evaluations such as the use of reminder cards.

In our report A $4 Billion Census in 1990? Timely Decisions on Alternatives to 1980 Procedures Can Save Millions (GAO-82-13, February 22, 1982), we recommended that the Bureau test the

feasibility of using mail reminder cards and follow-up mailings for nonrespondents. This recommendation was intended to reduce the need for personal visit interviews for the 1990 census.

We are somewhat concerned that the Bureau's mail card follow-up testing in Tampa did not include a multiphased follow-up approach. Response to the initial questionnaire mail-out was 56 percent in Tampa. The Bureau sent reminder cards to about half of the nonrespondents. The cards generated a net 3.8 percent increase in responses. Even with the increase in responses, enumerators still had to make door-to-door collections for about 40 percent of the Tampa questionnaires. In view of the Bureau's estimate that each 1-percent increase in the response rate will save $5-6 million in 1990, we believe the Bureau missed an opportunity to test the impact of a multiphased mail follow-up designed to achieve a greater overall mail response rate. A multiphased follow-up would involve two to three mailouts designed to encourage nonrespondents to return their questionnaires, prior to the Bureau beginning door-to-door collections. At least one of the reminders should include another questionnaire in case the original questionnaire was discarded.

In conjunction with the Tampa pretest, the Bureau is currently testing procedures to assist it in determining the feasibility of adjusting the raw census counts. The main features of the test include a post enumeration survey and an attempt to match the survey population results to the pretest population results using automated matching techniques.

Matching, as discussed in our report Procedures to Adjust 1980 Census Counts Have Limitations (GGD 81-28, December 24, 1980), was a major problem in prior attempts at coverage evaluation/adjustment.

Because of the importance of these tests to possible population adjustments, we plan to closely monitor the Bureau's activities.

1986 tests

During 1986, the Bureau will conduct several tests with different procedures and activities being stressed. Two pretests are planned--one in Los Angeles and the other in eight rural counties in Mississippi, including an Indian reservation. Both pretests will stress the use of decentralized data processing and combined collection/processing offices. I have already commented about these tests as they relate to automation.

Other activities to be tested include the delivery of questionnaires (Mississippi), ways to structure temporary jobs to reduce turnover (Los Angeles), improved methods for enumerating an Indian reservation (Mississippi), and continued work on testing procedures to adjust the raw census counts (Los Angeles).

We have also noted that the Bureau has made some changes in its proposed short form questionnaire for the 1986 pretests. For example, the question on plumbing was deleted and placed on the long form, some other questions were combined, other

questions expanded, and some were modified. Overall, the size of the form or content was not reduced.

Another test scheduled for 1986 is the Bureau's national content test. In that test various types of questions or versions of questions will be tested. We plan to closely monitor these activities in the ensuing year because of their possible influence on the decisions to be made for the next census.

RESIDENCY RULES

Where should persons who are counted be tabulated, and who should be included in the totals for apportionment purposes are basic issues in a decennial census. Traditionally, residency rules have been relatively constant and have followed the basic rules laid down in the First Census Act of 1790. The concept of usual residence has been fundamental in all past censuses. This is generally construed to mean the place where the person lives and sleeps most of the time. On the basis of current proposals, the Bureau will retain its basic residency rules.

The usual place of residence is not necessarily the same as a person's legal residence, voting residence, or the place where he or she happens to be staying on Census Day. For example, individuals from the United States who are abroad for an extended period of time are not included in the counts for apportionment purposes. Thus, a member of the Armed Forces who is assigned abroad and who may maintain a permanent legal address in this country and vote using an absentee ballot, is not counted for apportionment purposes. Moreover, a member of

the Armed Forces assigned to a domestic base or port is counted at that location regardless of his preservice residence or voting residence. A college student has traditionally been counted in the locality in which he resides while attending school.

Some states and locations can gain an advantage or be at a disadvantage because of these rules. For example, a state which has large military bases or ports, such as Virginia would gain from the Bureau's rules. Some states which have more college age students than college enrollment opportunities within that State (net college student migration) will be at a disadvantage for apportionment purposes.

On the other hand, decennial census counts are used for purposes other than apportionment, such as fund distributions. The larger the populations, the greater burden on the state or community for services needed. On that basis, the usual home rule has merit regardless of legal or voting residence.

There are other factors to consider in pondering the appropriateness of the rules. These include the ability to obtain accurate information with a reasonable cost and the relative size of the population groups under consideration. For example, there is no good source of data for the number of U.S. citizens overseas who are not affiliated with the federal government. Locating them would be difficult.

There is no simple formula for residency rules. The Bureau has followed our forefathers' resolve as laid down in the 1790 Act, and it has taken into consideration the practical

implications. Congress has not legislated residency rules in recent censuses; it has delegated that authority to the Secretary of Commerce, and it has permitted the Secretary to delegate further to the Bureau of the Census. Although the residency rules can be debated, the Bureau needs to develop its questionnaire and instructions to accommodate the rules decided on. The residency rules should be established within the next 2 years. Therefore, if the Congress wishes to involve itself in developing the rules, now is an appropriate time to do so.

RECOMMENDATIONS OF THE OFFICE OF INSPECTOR GENERAL ON ADP

In its report dated September 30, 1984, the Office of Inspector General (OIG), Department of Commerce, recommended that Bureau management support an ongoing ADP future systems design staff that would coordinate its efforts with decennial planning staffs. The OIG also recommended that the Bureau establish a formal ADP planning process which would include the development of a long-term life-cycle development plan to identify systems to be automated by 1990 and set a timetable for automation upgrade.

We support the OIG's recommendations. However, we do not currently believe that the recommendations, even if fully adopted, would expedite the Bureau's planning cycle for automation of the 1990 census. We believe that the incorporation of a life-cycle development plan, as I noted earlier, could affect the choice of equipment for data entry of the 1990 census because of the factors affecting the disposition of the equipment after the census.

To update the status of the Bureau's actions, we note that the Bureau designated a chief of the ADP Planning and Acquisition staff effective March 10, 1985. The Bureau specified that the functions and staff of the former long-range ADP planning staff, which had been disbanded in 1983, were reassigned to the ADP Planning and Acquisition staff.

The former long-range ADP staff reported to the Bureau's Deputy Director. The new staff will report to the Assistant Director for ADP, which is two levels below the Office of the Deputy Director. Because of this lower position in the organization, the current staff's influence and independence may be reduced. In that environment it will be difficult for the new staff to influence the planning of the automation of the 1990 census. This is discouraging if the Bureau is to achieve an integrated ADP operation.

OBSERVATIONS AND SUGGESTIONS

The Census Bureau's efforts in the 1985 pretests were useful in studying the automation of clerical activities associated with collection operations and determining the feasibility of early data capture. However, because the Bureau did not start vigorous planning and research early, as we have continually suggested, it has not maximized its opportunities in the pretests. It did not

--incorporate into its pretest planning the long lead time needed for automation acquisition,

--adequately design the census questionnaire to encourage item response in its Tampa pretest of the optical mark reader,

--formally evaluate the use of a long form that could be separated to permit data entry using the optical mark reader,

--properly test the two-stage census using a streamlined short form and a long form which does not repeat, except for needed linkage, questions in the short form,

--adequately evaluate the reasons for nonresponse in the 1985 pretests, and

--test the effectiveness of a multiphased mail followup scheme, including the sending of another questionnaire, to questionnaire nonrespondents.

Additionally, the Bureau is seriously considering the use of data keying, the most expensive, slowest, error prone, and labor intensive of the data entry options. On the other hand, it is not seriously evaluating the use of desktop optical mark reader equipment, which could have multi-purpose use after the census.

In order to develop the best 1990 census, we believe the processing technology, the collection methodology, and the questionnaire content and design must all be compatible and synergistic; and that significant changes in equipment, procedures, and forms should be adequately tested and evaluated. However, in view of the short planning time remaining, we believe the Bureau cannot afford the luxury of testing all the potential improvements in its formal pretests. Pretests require about a year for preparation, are expensive, and involve enumerator followup activities which are not a

requisite for all testing. We believe the Bureau should use special purpose tests to adequately evaluate (1) the potential of different types of data entry equipment (2) questionnaires, specifically a short simplified form and (3) procedures, including a two-stage census. Such tests could be completed months ahead of the 1986 pretests (field activities would not be required) and would provide the Bureau an opportunity to make up for some lost time in its decision making process.

Mr. Chairman, this concludes my prepared statement. We will be happy to respond to any questions.