SOME RECENT DECENNIAL CENSUS OCCUPATIONAL EXPERIMENTAL WORK

Stanley Greene, Bureau of Census

The Decennial Census data on occupation are receiving a great deal of attention of late. This very meeting on occupational classification is but one manifestation of this concern. Others are demand for special tabulations, special studies, articles in learned journals, and Bureau of the Budget research subcommittees. Although it is nice to be recognized, not all this attention has been flattering. In fact, some of it has been downright critical. "Why do you lump professional athletes with chemists?" "One-third of the labor force is ill-defined." "Response variation makes the data little better than nothing."

Some of these criticisms are valid. Valid in the sense that there is a satisfactory and at least theoretical alternative. Response variation can theoretically be reduced. If we ask the proper question(s) of the proper respondent we should get the correct answer.

Underlying the other critical comments is basically a dissatisfaction with the classification scheme. This is so because for many operational and administrative programs, projection work, training needs, educational facilities, the data just are not sufficiently refined. This is a broader concern and one for which no working alternative is posed and simply reflects a legitimate dissatisfaction with the keenness of the only available comprehensive tool. At any rate, any classification scheme is no easy task. For example, in a purer science than that of job classification, that of biology, various characteristics of an organism must be taken collectively into account in attempting even to distinguish between animal or vegetable. In some cases, especially of lowly organized forms, the distinction is difficult or uncertain.

So we see that implementing a classification scheme is difficult. Some schemes are impossible, for example classifying jobs by skill level, though desirable, is not directly possible. What common denominator is there, besides dollar earnings, to measure the comparable skills between the television repairman and the economist? The economist is completely baffled by the maze of wires and parts constituting the innards of the simplest electronic circuits and I daresay the repairman would not find Samuelson comprehensible.

Despite the difficulties the Census Bureau is in the job of classifying occupations and has done so since 1820. During that time a great many changes have been introduced into the system, varying from an industrial frame of reference to an extensive listing comprising 600 occupational groups. I mention this to indicate the flexibility of the Census approach to these problems.

This brings me to the subject of the paper, the concrete steps we are now engaged in, or seriously considering, to solve some of our occupational classification and reporting problems.

Within our existing classification system, that is, our job grouping system, there are three basic areas of concern: The not reported group, those cases for which we get no reports or the job is described in words that cannot be intelligently interpreted; n.e.c. or not elsewhere classified groups, such as "clerks n.e.c." where the category may be so broad as to provide little information; and the reliability of the data--is the distribution accurate?

Our experimental work in the not reported area has taken the form of a field test to determine if one of the elements comprising our not reported cases could be eliminated. The not reported category is one of our major concerns and this problem has been aggravated by the increases in the rates in the 1960 Census. In the hope of minimizing, to some extent, this problem, we examined a sample of occupation returns from the 1960 Census. These returns represented written entries that could not readily be coded and had to be reviewed by expert classifiers. These are called "referral cases." We noted that about 15 percent of such referral cases represented written entries falling into a single category. They are the "department" or "area of work" type of returns, which are actually subdivisions of industry returns rather than occupations. Some examples of this type of entry are "shipping department," "office work," "stock room." Since these responses provide no clue regarding any particular type of work activity, and, theoretically, they may reflect a broad spectrum of job activities, we had no alternative but to classify them as "not reported." A rough estimate indicated that they may account for around 6 percent of the total not reported figure.

A case could be made in support of the hypothesis that most of these "department" type entries do not cover a broad spectrum of activities within the area, but do, in fact, refer to specific types of work. That is, a "department" type entry may be strongly correlated with one occupation, the person doing work of a general nature in that area. If such is the case, and the groups of "department" type occupational entries are strongly correlated with a given occupation or fixed occupational proportions, then a system of assigning an occupational classification would be feasible. Such a system, in addition to reducing the number of occupation not reported cases, would also result in savings on operational costs by reducing the number of referral cases for the experts to resolve.

To test this hypothesis, a sample of the Cleveland Special Census (basically a methodological test of the mail out - mail back technique) was drawn from among such "department" entries. To this group we mailed a form noting the job entry provided in the Special Census and asking for further details. The additional probing questions related primarily to "job title" and priority order of activities and duties.

The returns from this test support the original hypothesis. As may be noted in the handout of Table A, 71 percent reporting "shipping department" were determined through the additional probes to be shipping and receiving clerks; 42

percent of the "stock room" entries could be classified to "stock clerks." Although these relationships are promising and do provide evidence to support the original premise that a sound allocation system might be devised, the most fruitful feature is that the responses to the additional probing items permitted the classification of occupation. Almost all (approximately 95 percent) of these cases formerly assigned to the not reported category, could, after examination of the entries to the additional questions, be assigned to an occupational group. However, there may be an adverse effect on the other component of the not reported, i.e., complete blanks. The possibility exists that by adding to the reporting burden we discourage response completely. I shall return to that point later.

The Census Bureau's Occupational Classification System's "Not Elsewhere Classified" categories are also a source of attention and concern by both users and the Bureau.

The Bureau of the Budget has noted in Working Paper No. 66-2: "These nonspecific categories included close to one-third of total employment in 1960 and their size has been the occasion of considerable complaint on the part of users of Census data." A similar observation was noted by Dr. Scoville in his paper on the relevance of occupational data where he states: "Further evidence of the present loss of analytical value of the existing classification scheme appears in its treatment of the 'not elsewhere classified' categories. It is probable that many of the key jobs for analysis of technological change are incorporated in these groups...Not only do the different groups grow at markedly different rates, but it is impossible to assert that the placing of one-third of the labor force into 'not elsewhere classified' groups does not affect its analytical value."

To put these views in proper perspective, it may be advisable to discuss briefly what these n.e.c. groups are and what they are not.

There are over 30,000 different job titles to codify, or classify. Because of limits of page space and tabulation, any statistical display of such detail is patently impossible. Thus the problem of combining and grouping. What criteria should be used in the decisions to determine the titles to be combined into a group? It is not done arbitrarily, nor by simple arithmetic division. Within the limits of our resources we provide the maximum detail of job families. The criterion used to determine the detailed categories is significance--significance in regard to analytic usefulness, policy need, and number of workers represented. What remains and constitutes one part of our residual n.e.c. categories are those job groups in which relatively few persons are employed and fail to have basic analytical or policy significance. Of course, combining many such groups will and does result in sizable employment figures.

One other element enters into the residual groups; they comprise that set of job entries provided by the respondents not sufficiently detailed to enter more than a generalized code. This may be necessitated by the fact that some jobs require a variety of duties to be performed. Thus our n.e.c.'s comprise two elements--a very detailed job entry--but not significant, and a generalized respondent entry. We may conclude from this that the n.e.c. categories are not simply "catch-all groups." Moreover, some of the distinct 31 occupation n.e.c. groups are quite limited in scope. As an example, our "Natural scientists (n.e.c.)" line is only n.e.c. by virtue of the fact that data for "Chemists" are shown separately. Had chemists been subsumed in the "Natural scientists," leading ironically to a broader category, there would be no n.e.c. designation required. Furthermore, the n.e.c. groups are in and of themselves of intrinsic value, certainly the "Natural scientists n.e.c." group is clearly distinguished from "Service workers, n.e.c."

One other matter concerning the n.e.c. groups--they are not fixed and static. In past decades the n.e.c. groups have been examined for specific job titles which occur frequently enough to warrant setting up new occupation categories. For example, after an examination of the "Clerical and kindred workers (n.e.c.)" group in 1950, five new specified clerical occupation groups were set up for 1960 (File clerks; Payroll and timekeeping clerks; Postal Clerks; Receptionists; and Stock clerks and storekeepers.)

Nonetheless, despite the logic of the n.e.c. groups and periodic reviews resulting in their streamlining, their composition suggested further approaches to the problem. You recall that the groups theoretically comprise two parts--a general part and a specialized part. If each of these parts contributes a sizable proportion to the whole, then an approach of splitting the n.e.c.'s into their two basic components would be advisable since it would result in a much better analytical understanding of their makeup, for we could then define and present separate data for the two components in our Census tabulations. For as Dr. Ann Miller has noted: "Even a simple separation of the n.e.c. categories into two component parts, 'miscellaneous' and 'not specified,' for example (analogous to the way such categories are handled in the industry tables of the Census of Population) would make some contribution to a reduction of the problem. For instance, analysts working with the cross tabulation of occupation by industry would probably be aware of the particular 'miscellaneous' occupations, as listed in the Classified Index, that are important in specific industries and would be able to interpret the significance of the 'miscellaneous' category in this context."

To determine whether the two components were numerically significant, it was necessary to review schedules. This was so, since a single 3-digit code applies to both components. Therefore, we analyzed each job title comprising our major n.e.c. groups--distinguishing between the general titles and the specialized titles. We used both a sample of Census returns for this analysis and a sample from our monthly CPS returns. As you may note from Table B, each of the two components contribute a sizable proportion in each particular n.e.c. group. The lowest proportion of general titles amounts to 19 percent for the operatives n.e.c. group. The Table also shows that these significant levels are true for both Census and CPS data and have held up for a period of more than a decade.

Barring any unforeseen operational difficulties, we propose a sizable reduction in our n.e.c. groups for 1970 by the device of splitting them into two distinct homogeneous categories - specific and general - resulting in much better analytical usefulness. Of course, in addition we shall perform our normal procedure of investigating each job title comprising our n.e.c. groups to determine if further subdivision or allocation to other occupation categories is warranted.

In regard to the still sizable general component, it would be very helpful if we could reduce this portion, resulting from vague, broad spectrum descriptions. It is certainly conceivable that these broad descriptions result from vaguely defined job tasks such as are associated with new workers where there are a multitude of secondary tasks. If this is the case, then little more can be done to more sharply depict this real job situation. There is no doubt, however, that some of this component is contributed to by communications failure. It is this phase of the problem that played a role in our experimental questionnaire design, adopted for our First Content Pretest for the 1970 Census. I shall return to this point shortly.

The third area of major analytical concern that I mentioned earlier had to do with the reliability of the data--is the distribution accurate? Much solid and valuable analytical work has been accomplished over the years with the existing body of occupation statistics. Nonetheless some of our evaluation work indicates that there is much room for improvement in the reduction of net and gross error when related to independent distributions.

The main thrust of our thinking on this matter is in terms of <u>additional</u> questions on occupation. This is a departure from our usual approach on these matters of question wording. For usually when an item is reported poorly we can trace the cause to a communications misunderstanding and a revision of the wording of the question will usually remedy the problem. But for occupation, we felt our basic question on "What kind of work were you doing?", with a series of examples--of any single approach was doing the best possible communications job. But what to do about the too vague idefinite responses? the upgrading problem? Perhaps if we elicited <u>more</u> information, other facets, more pieces in the jig-saw puzzle--the job picture would come into clearer focus. With this idea in mind, we formulated a series of supplementary questions to follow the basic item. These relate to job activities in priority order and to the employer's title. This new formulation of questions, among other items, was tested in the Bureau's First Content Pretest which went into the field in two test sites around the middle of May of this year. Field work was completed near the end of June and though it is too early for a definitive report on the results, we do have enough evidence to support some tentative conclusions.

My general impressions from field interview observations and a scanning of schedule returns are favorable. I was encouraged by the responses. The additional parts of the item on "Most important activity," Other important activities," and "Employer's title for this job" seemed to help clarify the concepts. It also may break down the difficult concept of occupation into more easily understandable parts, in addition to providing more information for our coders to use. In any case, many of our old <u>general</u> titles that were of necessity coded not reported or n.e.c. could now be more clearly defined. For example, occupation entries of office work had supplementary entries of "bookkeeper," "cashier," "clerk" was followed by cashier.

More quantitative measures of the success of this item are revealed in the handout Table C, which compares the NA and n.e.c. rates for the pretest against like measures from the 1960 Census which used only the basic questions.

As explained earlier, the n.e.c.'s contain a large component of generalized responses. If the level of this component is reduced, it would be a clear indication of more precision in our returns from the additional probes. As you can note in the table prepared especially for this paper from a sample of the St. Louis Park portion of the pretest, all but one of the major n.e.c. categories for the pretest showed lower proportions than for the 1960 Census. Some of the reductions were fairly substantial and overall the n.e.c. category was reduced close to 15 percent. The improvement may be even greater than this, for some of the generalized responses though clarified may have shifted to the specific component of the n.e.c. category, by the very nature of the n.e.c. makeup.

A problem noted with the wording is that the employer's title for the job is sometimes misconstrued to mean the title of the employer's job. So we get curious inconsistencies, such as housekeeper reporting the employer's title for her job as "owner of house," or "stock clerk" and "sweeper" followed by "manager." Therefore, we are going to modify the wording in our next Pretest and ask for the <u>formal</u> job title.

Another problem we face in implementing this item expansion is one faced by other Census items. The difficulty has to do with the methodological changes in census taking. Since we now operate on a self-enumeration, with follow-up of failures-basis, we require compatible item wording and format, that is, the schedule must be both respondent readable and easily adaptable for direct interview, if we are to avoid the problem of dual forms. However, the present wording, though appearing to satisfy the self-enumeration phase, is somewhat awkward in many situations of direct interviewing. How can an interviewer matter-of-factly ask someone who reports to the first part of the item "plumber" or "registered nurse" the additional probes on "activities" and "employer's title?" Problems of this nature can usually be resolved, most readily, by special training and instructions to interviewers.

There are two other questions about these supplementary probes before we can give them our unqualified endorsement. The first one has to do with the fact that we are getting <u>more</u> information. This introduces the problem of abstracting the relevant information from the additional entries which may be beyond the capability of the temporary relatively inexperienced Census coders. We will check on this feature by having such

coders (Census style) attempt to do the job, and then comparing it to our expert coders' results shown in the Table C.

The other question has to do with the earlier fear mentioned that increasing the occupational reporting burden might harm the overall return rate. This fear may be groundless, for a comparison of the not reported rates as shown in the last line of Table C clearly shows an improvement in the not reported rates for the Pretest relative to the Census. Moreover, the Cleveland follow-up test of "department" type entries had a higher return rate than many mail surveys conducted in the same manner.

You may observe that all three parts of our problem - the "department type entries" portion of our nonresponses; the "general" component of our "not elsewhere classified;" and the reliability phase - all have a common thread intertwining them. This linkage has to do with the acquiring of more detailed responses. As noted in our First Content Pretest, the additional probing items used to solicit these detailed responses reduced our n.e.c.'s. Similar style questions in our Cleveland follow-up test clarified many of our department type entries. Such results surely point to improved reliability.

This completes the subject topic on census experimentation in occupation classification. However, there are some other basic considerations I would like to present to this forum.

The first concerns the allocation of not reported occupation cases. In the planning of the 1960 Census, the basic change in enumeration procedures and the extensive use of computers led us to adopt a system of allocation of nonreports for many items. This allocation is basically a system of assigning an entry for an unknown characteristic on the basis of other reported characteristics. Presumably this is a help to the analyst, and superior to simply distributing the unknowns in accordance with reported distributions. This was not proposed for occupation because of the variety and complexity of the distribution. We are investigating techniques for such an occupation allocation and attempting to determine if a system of allocation according to important socioeconomic characteristics can be developed within a reasonable budget.

Concerning the reliability of the data, one view holds that respondents cannot report adequately because they are too limited in technical knowledge. This view concludes that we may have progressed as far as possible with the present approach. What is needed is an approach that secures the occupation information from the employee, through a modified W2 form or some other direct contact with the employer. This technique could be extended to other items, presumably more reliably reported by employers - wage and salary income would immediately come to mind. The basic problem here would be employer and Internal Revenue cooperation and the feasibility of linking these returns with the basic demographic data gathered from the households. At any rate, discussions are being held on these matters.

In regard to the basic consideration of the adequacy of the classification system itself. Do the groupings provide salient data for today's problems? Would another system be more appropriate? Is the worker's relationship to the machine a significant characteristic of job determination?

The only point I would want to make in this regard is to scotch the underlying assumption of such proposals that there is one ideal basis of classification. There is no such thing. No one system can supply the statistics necessary to meet the multitude of needs of users and analysts. Although a statement of this kind, at one time, would mean the end of discussion after agreement on the one best compromise system, in this day of the computer it is not impractical to think of a variety of systems, each tailored to a specific set of problems, and our thinking is exactly along these lines.

TABLE A.--OCCUPATION DISTRIBUTION OF "DEPARTMENT" TYPE RETURNS BASED ON CLEVELAND FOLLOW-UP PROBES

Occupation by Original Department Entry	Percent of Employed
Shipping Department	100.0
Shipping and Receiving Clerks	71.4
Laborers (n.e.c.)	14.3
Operatives (n.e.c.)	14.3
Stock Room	100.0
Stock Clerks	41.7
Purchasing Agents and Buyers (n.e.c.)	8.3
Operatives (n.e.c.)	8.3
Laborers (n.e.c.)	25.0
Clerical and Kindred Workers (n.e.c.)	16.7
Kitchen Help	100.0
Kitchen Workers (n.e.c.)	73.3
Cooks exc. pr. household	10.0
Housekeepers and Stewards	3.3
Managers (n.e.c.)	6.7
Prof. Tech. and Kind. (n.e.c.)	3.3
Service Workers exc. Pvt. Household	3.3
Office Workers 1/	100.0
Specified clerical occupations	60.9
Clerical and kindred workers (n.e.c.)	39.1
Specific duties	30.4
Broad duties	8.7
Miscellaneous Dept. Entries	100.0
Operatives and Kindred Workers (n.e.c.)	15.6
Clerical and Kindred Workers (n.e.c.)	9.4
Tech., Medical and Dental	9.4
Laborers (n.e.c.)	12.5
Furnacemen, Smelters and Pourers	6.3
Office Machine Operators	6.3
All Others	40.6

1/ Entries of "Office Workers" were classified as "Clerical and Kindred Workers (n.e.c.)" in the 1960 Census.

TABLE B.--MAJOR NOT ELSEWHERE CLASSIFIED (n.e.c.) OCCUPATION CATEGORIES BY GENERAL AND SPECIFIC COMPONENT FROM THE 1950 CENSUS AND 1965 CPS

(Percent)

Major n.e.c. category	1950 Census		1965 CPS	
	General Title	Specific Title	General Title	Specific Title
Both Sexes				
Managers, officials and Proprietors (n.e.c.)	61	39	69	31
Clerical and Kindred workers (n.e.c.)	68	32	49	51
Salesmen and sales clerks (n.e.c.)	70	30	60	40
Mechanics and repairmen (n.e.c.)	82	18	56	44
Operatives and kindred workers (n.e.c.)	26	74	19	81
Laborers (n.e.c.)	69	31	45	55
Males				
Managers, officials and Proprietors (n.e.c.)	61	39	69	31
Clerical and kindred workers (n.e.c.)	55	45	39	61
Salesmen and sales clerks (n.e.c.)	62	38	31	68
Mechanics and repairmen (n.e.c.)	NA	NA	56	44
Operatives and kindred workers (n.e.c.)	NA	NA	19	81
Laborers (n.e.c.)	NA	NA	45	55
Females				
Managers, officials and Proprietors (n.e.c.)	59	41	75	25
Clerical and kindred workers (n.e.c.)	75	25	55	45
Salesmen and sales clerks (n.e.c.)	82	18	95	5
Mechanics and repairmen (n.e.c.)	NA	NA	NA	NA
Operatives and kindred workers (n.e.c.)	NA	NA	19	81
Laborers (n.e.c.)	NA	NA	27	73

TABLE C.--EMPLOYED IN MAJOR N.E.C. OCCUPATION CATEGORIES FOR ST. LOUIS PARK, MINN., 1966, AND HENNEPIN COUNTY, MINN., 1960

(Percent)

Major n.e.c. category and not reported	St. Louis Park 1966 1/	Hennepin County 1960	Percent change
Major n.e.c. group total	26.6	31.1	-14.5
Managers, officials and proprietors (n.e.c.)	6.5	7.6	-14.5
Clerical and kindred workers (n.e.c.)	5.3	6.4	-17.2
Salesmen and sales clerks (n.e.c.)	6.4	7.7	-16.9
Mechanics and repairmen (n.e.c.)	0.3	1.6	-81.2
Operatives and kindred workers (n.e.c.)	5.3	5.0	+ 6.0
Laborers (n.e.c.)	2.8	2.8	-
Occupation not reported	1.8	4.6	-60.9

1/ Adjusted for differences in basic distribution of major groups.