

Implementing the Standard Occupational Classification in the OES Survey

Kristine Heffel and Laurie Salmon

The Occupational Employment Statistics (OES) program at the Bureau of Labor Statistics (BLS) began in 1971, with 50,000 questionnaires being sent to manufacturing establishments across the country. In 1973, nonmanufacturing firms were added to the survey. The original occupational classification system included mainly manufacturing workers because the survey originally was designed to collect data in manufacturing industries. Additional occupational groups were added as more industries were added to the survey. In 1980, the Standard Occupational Classification (SOC) system was developed by an interagency committee authorized by the U.S. Office of Management and Budget (OMB), and the occupational structure of the OES program was modified to reflect the SOC when possible.

When changes in the structure of the U.S. workforce over time made portions of the SOC outdated, the OES program used its data on workforce composition to update its own occupational structure to reflect these changes. The OES program used its own occupational structure through the 1998 survey. Because other sources of occupational information had different occupational structures that best suited the needs of their data users, the occupational data from different sources in the Federal Government were not comparable.

In order to ensure comparability of data among government agencies, the OMB established the SOC Revision Policy Committee (SOCRPC) in 1994 to create a classification system that would reflect the current structure of the U.S. workforce and be flexible enough to be used by all agencies collecting occupational data. Universal use of a single classification system would allow different types of data on the same workers to be compared. The SOCRPC conducted a study on how they should proceed, and decided to use the OES classification system as the starting point for the revised SOC, because the regular updating of the OES classification system had made it one of the most comprehensive and modern systems in existence.

For purposes of the new SOC, the OES classification system was modified to incorporate the needs of statistical agencies and to better reflect the changing economy and current occupational structure of the Nation. The new SOC reflects the shift toward a more service-oriented economy, advances in automation and technology, and increasing concerns about the environment. In comparison with the OES system, it

contains more detail in the areas of computer, engineering, health, and science occupations. In addition, many of the previous production occupations from the OES system have been consolidated, combining machine setters, operators, and tenders of a single machine, for example, into a single occupation in the SOC structure.

The SOC classification is based on occupations rather than individual jobs, where an occupation can be considered an aggregation of jobs requiring similar skills and tasks. Occupations in the SOC focus on the work performed and on required skills, education, and training. A definition is provided for each occupation, listing tasks performed by all workers in the occupation and, occasionally, duties that some workers may perform. The SOC is grouped by occupation into a hierarchy. There are 23 major groups, 98 minor groups, about 450 broad occupations, and approximately 800 detailed occupations, a structure that permits the study of different levels of occupational detail through the use of different levels in the SOC structure.

Federal agencies currently using the SOC coding system include BLS, the U.S. Census Bureau, the Employment and Training Administration (ETA), the National Science Foundation, the National Center for Education Statistics, and the Defense Manpower Data Center. The OES program is one of the BLS programs that uses the SOC for collecting occupational data, beginning with its 1999 survey. The OES program uses a 3-year survey cycle to collect the full sample of 1.2 million establishments, with one-third of the sample collected each year. Starting in 1998, data from 1996, 1997, and 1998 were combined to form the first set of OES estimates using the full sample. The introduction of the new SOC coding system for the 1999 data provided a technical challenge in combining the requisite 3 years of data. The remainder of this article describes changes from the old OES occupational classification to the new SOC classification system, and then explains some of the procedures used by the OES program in the transition to the SOC.

SOC-based survey forms

The OES program began using the new SOC codes, titles, and definitions on its 1999 survey forms, with the resulting employment and wage estimates included in this publication. Data for the majority of the detailed SOC occupations are collected on the OES survey forms. However, in eight

cases, the OES program used the SOC broad occupation instead of the detailed occupations. For example, the OES survey forms contain the broad occupation “dentists” rather than the five more-specific dental specialty occupations. The specific combination of detailed and broad SOC occupations used by the OES program is referred to as the OES 99 structure.

The SOC-based survey forms contain both the occupational titles and definitions; however, in many cases, the definitions have been shortened to include only the key duties of the occupation, due to space limitations on the survey forms. In addition, many occupations have an alternate title listed at the beginning of the definition to assist in coding specific job titles. No form contains all 800 SOC occupations; instead, each survey form contains occupations specific to the industry, or industries, being surveyed with that form. This reduces the burden on responding employers by limiting the number of occupations that appear on each survey form. If an occupation is not specifically listed on a survey form, the respondent can write in additional occupations on a supplemental page.

The occupations for each form were chosen based on employment for related occupations appearing on the old OES forms. Supplemental data from the BLS Office of Employment Projections matrix were used to decide which additional occupations should be surveyed for a particular industry. Based on these sources, the most common occupations for an industry or group of industries appear on the relevant form. Some occupations are common to all industries and appear on nearly every survey form, including general and operations managers; accountants and auditors; executive secretaries and administrative assistants; bookkeeping, accounting, and auditing clerks; and janitors and cleaners, except maids and housekeeping cleaners.

There are three types of forms used for the OES survey: Long forms, short forms, and unstructured forms. The long forms are sent to larger establishments, generally those with more than 50 employees. Most long forms contain between 100 and 200 occupations, with an average of 162 occupations per form. In addition to the detailed occupations, the long forms contain residual categories for those workers in occupations not specifically listed on the form. At the end of the form is a supplemental page on which employers are requested to provide more-detailed occupational information for those workers who have been categorized in the residual categories.

Establishments with 10 to 49 employees usually receive a shorter version of the form. The number of occupations on the short forms ranges from 13 to 95, with an average of 54 occupations per form. There are no residual categories on the short forms. Instead, there is a supplemental page at the end of the form that allows employers to write in additional detailed occupations not included on the form. These supplemental sheets provide information about occupations that may be new and emerging in the labor market or that may be appearing in new industries.

The unstructured forms generally are sent to establishments with fewer than 10 employees and do not list any detailed occupations. Instead, the employer provides the title and job description for each employee in the establishment, in a procedure similar to that used to fill out the supplemental sheets on the other forms. State classification specialists code each worker listed on an unstructured form to an SOC occupation based on the job duties.

SOC training and support materials

To assist the States in learning the new SOC coding system, the OES program developed a computer-based training (CBT) system. This system presents principles and rules of occupational coding in the SOC within an individualized interactive format, allowing the user to work at his or her own pace and review sections as necessary. Sections in the CBT system cover background information, SOC structure, SOC definitions, coding guidelines, SOC highlights, and coding exercises. The SOC-CBT is primarily designed to introduce new users to the SOC system and the rules of occupational coding, but can also serve as a coding review for the more experienced SOC user. An SOC coding course also has been offered to provide the State agency staff with additional information about the SOC system and coding principles.

Data validation

The OES program uses a series of computer programs to validate the employment and wage data. Many of these validation programs have been modified to accommodate the new SOC codes. The validation system includes programs that list:

- Occupational wages by area that are significantly different from data for the previous year.
- Occupational pairs for cases in which a typically low-paying occupation earns more than a typically high-paying occupation (such as nurses’ aides reported as earning more than nurses).
- Occupations in unusual wage intervals (such as lawyers reported in the lowest wage intervals).
- Occupations in unusual or invalid industries (such as Postal Service mail carriers reported in industries other than the Postal Service).
- Dependent occupations without an independent occupation (such as supervisors with no subordinate workers, or carpenter helpers in an establishment that reports no carpenters).
- Questionnaires on which all employment appears in one occupation or one wage range.

Related programs also identify the specific establishments reporting the anomalies identified by the programs above. Thus, when potential problems are located, respondents can be contacted for verification, clarification, or correction of the data provided. These related programs are designed to screen data for potential coding or data entry errors, as well as for respondent errors, and to assist in data review at both the national, regional, and State office levels.

Conversion of 1997 and 1998 OES data

Because the OES sample design requires combining data collected over a 3-year period, data from 1997 and 1998 had to be converted to the SOC format so that they could be combined with 1999 SOC-based data. In order to convert OES codes to SOC codes, a list of occupations comparable between the current OES structure and the new SOC structure was created. About one-half of the 769 occupations permit combining OES-coded data from 1997 and 1998 with the SOC-coded data from 1999. These matches include occupations that “crosswalk” from one OES occupation to one SOC occupation and cases in which several OES occupations aggregate cleanly into a single SOC occupation. The one-to-one matches have similar titles and definitions in the OES and SOC classification schemes. For example, the OES occupation credit analysts (21105) has the same title and definition as the SOC occupation credit analysts (13-2041). Other examples of one-to-one matches include tax preparers, clergy, interior designers, and ship engineers. An example of a single SOC occupation that consists of a complete aggregation of multiple OES occupations is welders, cutters, solderers, and brazers (51-4121), which is made up of the two OES occupations welders and cutters (93914) and solderers and brazers (93917).

When an SOC occupation could be constructed exactly from one or a combination of several OES occupations, the OES data were converted to SOC-based data. The 3 years of combined data (1997, 1998, and 1999) were used in estimating wages, while employment estimates were generated from

the single year of 1999 SOC data. If data for an OES occupation could be converted to an SOC occupation, but the SOC occupation could not be fully constructed from OES data, none of the OES data for that occupation was used in SOC estimates. In the 2000 survey estimates, to be released in early 2002, all survey data used in producing the estimates will be derived from source data originally coded to the SOC-based classification.

Testing

Using the OES-to-SOC occupational crosswalk, the 1997 and 1998 data were converted from the OES occupations to the SOC occupations. For testing purposes, occupations classified as nonmatches were coded to the SOC major group. Employment and wage estimates using 1999 data only at the SOC major group level were compared with SOC major group estimates produced using 3 years of data. This major group validation was conducted at the national and State levels. Detailed estimates produced with the recoded 1997 and 1998 data were compared with 1998 estimates for the occupations that are one-to-one matches, and with SOC estimates from 1999 data alone. This validation process was used to ensure that the occupational data were converted correctly from the OES codes to the SOC codes.

After the 1997 and 1998 master files were converted to SOC occupational codes, and after procedures were validated, the data were combined with SOC data collected in 1999 to create wage estimates based on 3 years of sample data. Occupational wage estimates using a combination of 3 years of data are footnoted in table 1 of this publication.

Generation of SOC estimates

The 1999 wage estimates were generated using 3 years of data wherever possible to improve estimates for small geographic areas. For the occupations that could not be recoded from OES to SOC, wage estimates were generated from the single year of 1999 data. Further information on the estimation procedure can be found in appendix B.