

In 1980 Proceedings of the Social Statistics Section of the
American Statistical Association, 98-107, 1981.

OCCUPATIONAL STATISTICS AND VOCATIONAL ANALYSIS

Linda S. Gottfredson, Eleanor Simonsick, and
Francine Voorstad, Johns Hopkins University

Governments at local, state and federal levels spend a large amount of money and effort periodically collecting occupational data, such as in the recent U.S. census. The government and private foundations also sponsor large surveys of career development by researchers. Both of these kinds of data are often subsequently made available to the general research community for secondary analysis. This report examines how these public occupational data are used and how well they meet the needs of one group of current or potential users--vocational researchers.

Vocational Research

"Vocational research" refers to research done in the tradition of counseling psychology and career guidance. The development of vocational interest inventories such as the Strong-Campbell Interest Inventory (Campbell, 1971) and the Vocational Preference Inventory (Holland, 1975) and person-job matching theories (e.g. Holland, 1973) are part of this tradition, as are the career development theories of Roe (1956), Super (1957), Crites (1976), and Krumboltz (Krumboltz, Mitchell and Jones, 1978). Vocational researchers are frequently associated with Division 17 (Counseling Psychology) of the American Psychological Association, or with the American Personnel and Guidance Association. Specifically, this report examines the use of public data by individuals who have published in either the Journal of Vocational Behavior or the Vocational Guidance Quarterly. Holcomb and Anderson (1977) have shown that these two journals publish the largest number of articles in vocational research. The fields of research not being focused on therefore include industrial and personnel psychology, sociology, economics, as well as other non-psychological fields that study employment, occupations, or career development.

A Model of Data Selection

This study was begun with two assumptions: that public occupational data are seldom used in vocational research (vocational researchers generally using data they or their colleagues collected), and that these data nevertheless could be quite useful in such research. To help investigate why public data are or are not used, the following model of how researchers select data for their research was used. Although the model is applied specifically to vocational research, it can be applied to other types of research as well. Three stages outline how a researcher comes to use a particular data set: being familiar with the data set, seriously looking into using it, and actually using the data in research.

(1) Being familiar with the data. Researchers can only choose to use data sets that they know about, so the first step in using a particular data set is that the researcher know it exists. It is of course true that researchers may actively search for or collect new data sets that will enable them to investigate the research questions

they have posed. However, it is likely that many researchers will first scan the available data they are already aware of before engaging in a search for new possibilities. Searching for new information involves costs and researchers may settle for data which "satisfices" their goals, that is, settle for data that are perhaps not exactly what they want but that are satisfactory. If there is an active search for new data, it will probably occur first in immediate or familiar surroundings and continue only until one or a few promising possibilities are located. The search is not likely to be extensive also because research questions have a way of shifting to better match information that is readily available.

In sifting through various possibilities, researchers will do a preliminary evaluation of different data sets (existing or collectable) and weigh their costs against their benefits. Screening questions such as the following would be used to assess potential benefits. What are the age, race, sex, or geographic areas included in the data? Does the data set have the type of occupational data I need? How recent are the data? A quick evaluation of costs would also be made. How difficult or costly would it be to obtain or process the data? Would I need special personnel or resources? Do I know anyone who has found these data useful or usable?

(2) Seriously looking into using the data. If a data set has passed the preliminary evaluation and seems worth looking into, the next stage is for the researcher to actually try to obtain and set up the data set so it can be used. If the researcher wishes to look into an already existing data set, this involves figuring out exactly who to contact or where to go to obtain the data, how much it will actually cost and in what form it will be obtained, and what resources need to be lined up to receive, set up, and process the data. After the data set or its documentation is received, the researcher can become more intimately acquainted with both the data set's potential benefits (e.g., the specific variables contained and the number of cases available for different analyses) and its costs or problems (e.g., inconsistency of questions, troublesome file structure, missing data, cost of computer time or personnel to recode or process the data).

(3) Using the data. The final step is the decision about the extent to which the data will be used. The data may be abandoned for many reasons--disappointment with the quality of its information, excessive cost in using the data, finding a more suitable source of information, and so on. If the researcher decides to use the data, that use may be minor or major depending on the researcher's needs and other sources of information.

One objective of this study was to discover at what stage in the data-selection process, and why, different public data sets tend to be eliminated from consideration. Another objective

was to provide suggestions about how to improve the value or availability of public data sets to vocational researchers and how to increase their use of these data.

Method

This report is based on data gathered in two studies: a content analysis of journal articles and a mail survey of authors of those journal articles. Both studies involve classifying occupational data as either public or non-public. Because the line between public and non-public is not always clear, the definition used in this report is detailed below.

Definition of public occupational data

Some data are clearly public, such as the volumes of census reports and the computer tapes for sale by the Data Users Division of the Census Bureau. Other examples would be some of the large surveys prepared for general sale by the research organizations funded to collect the data. Such data are available to almost any customer who wishes to purchase them and there are set procedures for obtaining them. There are also other data sets which we might all agree are private property. But there is also a large amount of data which might be called quasi-public or potentially-public. Investigators or institutions may voluntarily share their data with interested researchers, but they may be under no obligation to do so and may have no established procedures for distributing or archiving the data. Data which are considered useful to the research community eventually may be deposited with data archives such as the National Archives and Records Service or the Inter-University Consortium for Political and Social Research, after which time they can truly be considered public. In addition, data which may have at one time been available to almost anyone who asked for them may no longer be public if no provision was made to archive them. The criterion used in this report for considering a data set public was that it be available from a library, data archive, or data user's service.

A data set was classified as occupational if it contained information on occupations, work settings, or career development. Two types of data were included: periodic data and occasional surveys. Periodic data are employment and occupational statistics which are produced regularly --e.g., the census of population, the Current Population Survey, and manpower projections. Occasional surveys are special surveys often funded by, but not necessarily carried out by, the government and which are usually only one-time occurrences--e.g., Youth in Transition. Because much research on careers is done in some fields with these occasional surveys, it was important to include such data sets in this survey of vocational research.

Content Analysis of Journal Articles

A content analysis was done of all articles published in the Journal of Vocational Behavior and of all papers in the "Articles" section of the Vocational Guidance Quarterly (with the exception of 7 articles reprinted from other sources) from 1975 through 1979. Articles were classified according to several criteria, in-

cluding source of data (types of public and non-public data) and content area (e.g., vocational interests, vocational maturity, job satisfaction, treatment effects). A total of 511 articles were analyzed, 331 from the Journal of Vocational Behavior and 180 from the Vocational Guidance Quarterly.

Mail Survey of Journal Authors

Questionnaires were sent to most of the first authors of the journal articles for which a content analysis was done. The 511 journal articles represent 407 first authors. Questionnaires were not sent to the 43 foreign authors, the 3 corporate authors, or the principal investigator of this study, leaving a total of 360 authors to whom questionnaires were mailed. Because this study deals with the evaluation of U.S. public data, it was considered appropriate to exclude foreign authors from the mail survey.

Of the total 360 people in the sample, about 68% returned usable questionnaires; another 3% replied that they do not use public data and so did not fill out the questionnaire. One person in the sample was deceased and the post office returned another 5% of the questionnaires because of no forwarding address. Most of the remaining 24% of the sample presumably received the questionnaires but chose not to respond because the third followup was a certified letter containing another copy of the questionnaire and it was returned by neither the respondent nor the post office.

Five general types of information were sought with the mail survey of authors: (a) their involvement in producing public occupational data, (b) their extent of involvement in using public data, (c) strengths and weaknesses of the different data sets for vocational research, (d) the priorities they would give to obtaining different types of information if it were available, and (e) information about the authors (e.g., journals read) to identify channels by which researchers could be reached to solicit their advice or provide them with information about public data. This report focuses on (b) and (d) above. Detailed results for (a), (c), and (e) are described elsewhere (Gottfredson, Voorstad, & Simonsick, in preparation).

Respondents were requested to identify their degree of involvement in using three dozen specific public occupational and career data sets. The response categories correspond to the model of data selection discussed earlier:

- (a) "not know"--"you have never heard of it or you know almost nothing about it"
- (b) "do know"--"you are familiar with it, but you have not seriously looked into or tried using it in your research"
- (c) "looked into"--"you seriously looked into using it, but you were not able to use it in your research"
- (d) "minor use"--"you have successfully used it a little in your research"
- (e) "major use"--"you have made major use of it in one or more of your research projects"

The list of three dozen public data sets includes many of the largest or best-known sets of public occupational data and they are primarily national in scope. Although the list is hardly a

complete list of public occupational data that might be of interest to vocational researchers, it does include some of the most important and widely-used sets of occupational data. Almost all the data sets on the list were funded by the government, although many were collected by universities or research organizations. The list was compiled by consulting individuals experienced in the use of public data and by consulting a catalog of occupational data compiled by Ashley (1977), the catalog of holdings by the Inter-University Consortium (1977, 1978, 1979), and various government catalogs and descriptions of data (e.g., U.S. Bureau of Census, 1977).

Results

Content Analysis

The content analysis of journal articles led to the following conclusions. It should be noted that the last three conclusions are based on only 15 cases.

Public data are seldom used in vocational research. Of the 511 articles analyzed, 398 were empirical investigations rather than theoretical articles or literature reviews. Of those 398 articles, 17 used public occupational data in

some way. Of the 359 articles by U.S. authors, 15 used public data. Thus, only about 4% of the empirical research articles published in the two journals used public occupational data in either a minor or major way.

Data traditional in vocational research are typically used to supplement non-public sources of data; articles which use non-traditional public data tend to rely exclusively on that type of data. Table 1 shows the types of public data that were used by the 15 U.S.-authored articles (13 first authors) using public data. Results are presented separately according to whether or not that type of public data is traditional in the field and according to whether or not non-public data (e.g., interview data collected by the authors) were also used in the study. Fifteen public data sources were used, 4 of which are traditional in vocational research. Of the articles which used public data to supplement non-public data, all 6 used traditional data and 4 of the 6 used only traditional sources of data. In contrast, 8 of the 9 articles using only public data used only non-traditional data.

Table 1

Sources of Data and Content Area for the 15 U.S.-Authored Articles Using Public Data

Type of Public Data	No. Articles Using Various Public Data ^a		Articles Classified According to Main Type of Public Data Used ^b	
	Used public only	Also used non-public	No. articles	Content area
<u>Traditional^c public data</u>				
Project Talent		2		
Dictionary of Occupational Titles	1	4	5	evaluation of vocational assessments (3), perceptions (1), career change (1)
Occupational Outlook Handbook		2		
American Council on Education Survey		1		
<u>Non-traditional public data:</u>				
<u>"occasional surveys"</u>				
National longitudinal survey, boys	2			
National longitudinal survey, men	1			
National longitudinal survey, girls	1		5	vocational interests (2), evaluation of vocational assessments (1), abilities (1), evaluation of education and training programs (1)
Youth in transition	1			
Proprietary and non-proprietary training programs	1			
<u>Non-traditional public data:</u>				
<u>"periodic data"</u>				
1970 census public use tapes	2			
1970 census reports	2	2		
1960 census reports	1		5	employment patterns (5)
Manpower projections	1			
City-county data book	1			
Membership directories	1			
<u>Number of articles</u>	<u>9</u>	<u>6</u>	<u>15</u>	

^aThe same article may be listed more than once.

^bEach article is listed only once.

^cTraditional within vocational and counseling psychology.

Articles using traditional public data sets tended to focus on the more traditional topics in vocational psychology. Those using non-traditional data tended to study the less often explored topics. Table 2 provides an indication of what the most common topics are in the two journals. Results are provided separately for empirical and non-empirical articles. Vocational interests, vocational maturity (the readiness to make choices), the perception and classification of jobs, job satisfaction and adjustment, and vocational guidance, treatments, and counseling processes accounted for over 75% of both the empirical and the non-empirical articles. Topics related to obtaining and keeping jobs--education, abilities, job search, unemployment, placement and training programs, and employment patterns--were central to only a few articles each. These results are consistent with those of Holcomb and Anderson (1977).

Table 2

Substantive Focus of Journal Articles
(Both U.S. and Foreign Authors Included)

Substantive Area	Empirical	Not Empirical
<u>Individual differences</u>		
Interests, aspirations	64	3
Abilities	5	0
Vocational maturity	41	4
Perceptions of jobs & sex roles	34	1
Family background	15	1
<u>Employment problems of individuals</u>		
Education & training	2	1
Job search	3	0
Career commitment	10	1
Occupational socialization	3	1
Satisfaction	22	4
Adjustment	16	12
Career changes	10	3
Career achievement	21	0
Military service	0	0
Retired	2	0
Unemployment	0	2
<u>Labor market characteristics</u>		
Classifications of jobs	30	3
Employer practices	15	2
Employment patterns	5	1
<u>Design & evaluation of treatments</u>		
Vocational assessments & treatments	71	16
Training programs	6	2
Placement programs	0	3
Counseling profession & practices	21	20
Comprehensive guidance systems	1	24
<u>No primary emphasis</u>	1	9
<u>Total</u>	398	113

The right half of Table 1 shows the major focus of articles using public data. Of the 5 articles using primarily traditional rather than non-traditional public data, 3 evaluated vocational assessments or treatments, one examined perceptions of jobs, and one examined career changes. Incidentally, the former four articles were all concerned with sex bias, a topic of great concern in the field during the late 1970's.

Turning to the 10 articles using primarily non-traditional sources of public data, 3 examined popular topics: 2 looked at vocational interests and one at vocational assessments. A fourth looked at abilities, a topic which some might consider traditional to the field but which receives little attention. The remaining six examined more non-traditional topics: one was devoted to evaluating education and training programs, and five to patterns of employment. It should also be noted that these latter five articles are the only ones in the last five years in the two journals devoted primarily to research on employment patterns and job availability.

All authors using non-traditional sources of public data have been affiliated with research centers, usually with the centers that collected the public data. No systematic analysis of institutional affiliations was done in the content analysis, but the following pattern appeared when users of non-traditional data were examined more closely. First, all users of the "occasional surveys" (5 authors) have worked at the research centers that collected the survey data, 3 being from the same center. All users of the "periodic data" (3 authors) work at a single other research center. Thus, over half the articles using public data originated in only two places.

Author Survey

The results of the author survey are consistent with and extend the results of the content analysis.

Public occupational data are used extensively by a small proportion of vocational researchers, these data playing little or no role in the work of most vocational researchers. Table 3 shows that 70% of authors surveyed say that public data are "somewhat" or "not at all" useful. About 11% say that public data have been "essential" in their work. Table 4 is consistent with this pattern. Three-quarters (179) of the respondents reported doing research on occupations or careers in the past five years. Of these vocational researchers, 58% reported not using any of the

Table 3

Usefulness of Public Occupational and Career Data
In Respondents' Work (N=235)

Rating of Usefulness	% of Respondents
Not at all useful	17
Somewhat useful	53
Very useful	19
Essential	11

the 36 specified data sets in the past five years; another 22% reported that only 1 to 10% of their research was conducted with these data. About 13% of the researchers (being about 10% of all respondents) reported moderate to heavy use of public data (i.e., 21 to 100% of their research being done with public data).

Table 4

Percentage of Research Done With Public Occupational Data ^a (N=179)	
% of Research	% of Researchers Reporting Each Level of Use
0	58
1-10	22
11-20	7
21-50	8
51-100	5

^aPublic data refers here only to the 36 data sets listed in the survey.

Most public occupational data sets are not well known. But published reports and government data are better known and more often used than are data tapes and non-governmental data. This conclusion about levels of contact is derived from Table 5. This table shows the percentage of all respondents reporting different levels of contact with each of the 36 data sets. The median case is underlined to highlight the major levels of contact with each data source.

Only one data source, the Dictionary of Occupational Titles, is used by most of the respondents. The high rate of use of the DOT is not surprising because it is a tool in vocational guidance and counseling. (Had the Occupational Outlook Handbook or other books of job descriptions been included in the list, these counseling tools would also have been marked as highly used.) Many of the respondents may have used the DOT for non-research purposes and not for research, but the results do show that almost all respondents are familiar with it and most use it in some manner. In contrast, most respondents knew little or nothing about the computer tapes of DOT data and very few had used them.

The next most widely known and used data are the various published reports from the Department of Labor and the Census Bureau. It is likely that the relatively high rates of familiarity with the wage surveys, special labor force reports, CPS and census reports, the 1972 Manpower Survey, and the 1976 survey of income and education are overestimates. For example, respondents may examine income data occasionally but may not accurately remember the source of that data. Therefore, they may say they are familiar with any source--e.g., wage surveys--which sounds appropriate. Nevertheless, whether or not the particular sources are often used, it appears that Census Bureau and Department of Labor reports are familiar to most vocational researchers and are more widely used than other public data listed in the questionnaire. This same high level of familiarity and use does not extend to the tape versions of such data, however,

as was also the case with the DOT data.

The two data sets from the National Center for Educational Statistics, another government agency charged with collecting and disseminating data, were as well known and used as often as any of the data sets collected by non-governmental organizations. In contrast, the data set from the Civil Service Commission, an agency whose mission is not data collection, was not at all familiar to the respondents.

Project Talent, from the American Institutes for Research, was the only non-governmental data set known and used by a relatively large proportion of respondents. It is among the older data sets, but it has also been one of the most widely advertised in vocational psychology. Other data sets from the same organization are comparatively unknown.

The non-governmental data sets (most of which were funded by the government or involved participation by government agencies such as the Census Bureau) can be divided into two groups. The first group are those organizations which have collected several occupational and career data sets (and which generally do so fairly routinely). The second group are those listed under "other non-governmental," and which were produced by organizations perhaps less involved in the collection of occupational and career data. The latter group was the least well known--generally only 10 to 20% of respondents knew very much about these data sets. The National Assessment of Educational Progress is an exception here, but it has been fairly well advertised and it is of particular interest to the National Center for Educational Statistics.

Familiarity with and use of public occupational data vary according to the characteristics of respondents and their work environments. The content analysis suggested that people who are involved in producing occupational data and people affiliated with research centers are likely to use public data more. Table 6 is consistent with these results. It shows that from among respondents who have done research on occupations or careers in the past five years, those who have been closely affiliated with research units or research organizations or who have been involved in planning or evaluating government data are familiar with and use more of the 36 public data sets than do the other respondents.

Table 6 also shows that familiarity and the number of data sets used in either a minor or major way are associated with disciplinary identification and years since receiving one's highest degree. Finally, the mean number of data sets known and used is related to respondents' ratings of the usefulness of public data in their work. As with several of the other characteristics (e.g., involvement in producing public data), however, the direction of causation is not clear.

Many public occupational data sets are more relevant to the concerns of vocational researchers than their low level of familiarity and use of would suggest. Cost, convenience, and quality of data are all important concerns in selecting and being able to use different data sets. But relevance (coverage of relevant populations and

Table 5
 Percentage of Respondents at Different Levels of Involvement with Selected Public Data Sets
 (A="Not know about," B="Do know about," C="Tried to use," D="Minor use," E="Major use;" Median is underlined; N=243)

Data Set	Level of Involvement					Data Set	Level of Involvement						
	A	B	C	D	E		A	B	C	D	E		
<u>Department of Labor</u>													
1 Published Dictionary of Occupational Titles codes for jobs, worker traits & conditions	3	28	10	33	26	20	Proprietary & non-proprietary vocational training programs, student survey	85	9	3	2	1	
2 Computer tapes of Dictionary of Occupational Titles job & worker codes	58	32	6	3	3	Center for Human Resource Research, Ohio							
3 Wage surveys, e.g. area wage surveys	27	44	12	13	4	21 National longitudinal survey of girls 14-24	71	17	5	5	2		
4 Special labor force reports	26	32	12	22	7	22 National longitudinal survey of boys 14-24	72	16	5	5	2		
<u>Census Bureau</u>						23 National longitudinal survey of women 30-44	72	16	5	5	2		
5 1970 public use sample computer tapes	51	39	3	3	3	24 National longitudinal survey of men 45-59	71	17	4	6	2		
6 1960 public use sample computer tapes	54	42	3	1	0	<u>Institute for Social Research, Michigan</u>							
7 Published reports from the 1960 or 1970 censuses, e.g., Occupational Statistics subject report	21	42	7	20	9	25 Youth in transition	62	24	7	5	2		
8 Current population survey computer tapes	64	32	2	1	1	26 Survey of working conditions	64	22	5	7	2		
9 Current population survey published reports, e.g., Current Population Reports	40	40	4	12	5	27 Quality of employment survey, 1972-73	63	19	6	10	2		
10 1972 professional, technical & scientific manpower survey	47	36	6	9	3	28 Quality of employment survey, 1977	63	21	7	7	2		
11 1976 survey of income & education	50	33	7	11	1	29 Panel study of income dynamics	83	12	3	2	0		
<u>Civil Service Commission</u>						Other non-governmental							
12 Central personnel data file (CPDF)	93	6	0	0	0	30 Occupational changes in a generation (OCG) I or II (Wisconsin)	89	7	2	1	1		
<u>National Center for Educational Statistics</u>						31 Community colleges & vocational technical education centers, Graduate questionnaire (Bureau of Social Science Research)	89	7	1	2	1		
13 High school dropouts, 1960-1970	66	22	5	6	1	32 Career plans & experiences of 1961 college graduates (National Opinion Research Center)	78	16	2	3	2		
14 National longitudinal study of the high school class of 1972	59	26	5	7	4	33 Retrospective life history data (Johns Hopkins)	86	10	2	2	1		
<u>American Council on Education</u>						34 School to college: Opportunities for post-high school education (SCOPE) (Survey Research Center, Berkeley)	80	14	3	3	0		
15 Followups of 1961 college freshman class	71	23	2	4	0	35 Career & occupational-development survey of the National Assessment of Educational Progress (Education Commission of the States)	71	17	4	5	4		
16 Followups of 1966 college freshman class	70	23	3	4	0	36 Explorations in equality of opportunity: A 15-year followup of sophomores (North Carolina)	95	4	0	0	0		
17 Followups of 1968 college freshman class	71	22	3	4	0								
<u>American Institutes for Research</u>													
18 Project Talent	24	43	13	16	5								
19 Proprietary & non-proprietary vocational training programs, alumni survey	84	10	2	4	1								

Table 6
Mean Number of Data Sets Researchers are
Familiar With and Have Used

	Mean Number of Sets Known	Mean Number of Sets Used	(N)
<u>Affiliated with research unit</u>			
No	11**	3**	(112)
Yes	16	5	(83)
<u>Have been involved in planning or evaluating government occupational data</u>			
No	13*	3**	(173)
Yes	18	8	(18)
<u>Disciplinary identification</u>			
Psychologist	13*	3	(112)
Sociologist	17	4	(14)
Economist	19	6	(5)
Counselor	14	4	(16)
Educator	13	4	(31)
<u>Years since highest degree</u>			
0-5	11**	3*	(73)
6-10	14	4	(66)
11+	16	4	(55)
<u>Usefulness of public occupational data</u>			
Not at all useful	9**	1**	(34)
Somewhat useful	13	3	(97)
Very useful	16	4	(37)
Essential	20	7	(18)

Note. Table includes only respondents who have done research on occupations or careers in the past five years.

*Significance level of differences $\leq .05$ (Chi-squared test)

**Significance level of differences $\leq .01$ (Chi-squared test)

variables) is a primary consideration. If the data sets respondents were questioned about are not relevant to vocational researchers, it makes little sense to improve the three other aspects of the data--cost, convenience, and quality--specifically for vocational researchers.

Table 7 shows the priority respondents would give to obtaining different types of data if they were available. (Only respondents who plan to do research on occupations and careers in the next few years are included.) The results are consistent with the profile of topics exhibited in the journal articles analyzed (Table 2). Data on personal values, adjustment, and interests are given the highest priority. Several of the public data sets used in the published articles examined issues central to the discipline--vocational interests and vocational treatments. And

over half of the public data sets listed in the mail survey provide information for examining core topics such as vocational interests, preferences, values, and satisfaction. Although the coverage of some relevant variables may be poorer in public data than that which the respondents may collect, the public data sets do have the advantage of being large and more representative.

Table 7
Priority that would be Given in Obtaining
Different Types of Information if it were
Available (Percentage)

Variable	Priority			
	None	Low	Mod	High
Work related values and attitudes	2	8	26	64
Job satisfaction and adjustment	1	11	27	62
Vocational interests and aspirations	5	9	28	58
Characteristics of occupations and work environments	4	16	34	47
Perceptions and knowledge of occupations	7	18	34	41
Personal abilities and aptitudes	7	17	39	38
Job performance, achievement, and income	3	24	35	38
Socioeconomic and cultural background	4	20	40	36
Job histories	10	27	30	33
Education and training histories	7	18	39	36
Job search	10	26	33	31
Childbearing plans and sex role attitudes	15	29	25	31
Labor market conditions and job availability	11	28	33	28
Characteristics of employers and firms	11	33	32	24
Parental values and child-rearing practices	21	35	26	19
Characteristics of schools and training programs	14	39	26	21
Characteristics of spouse and own children	18	37	28	17
Community characteristics	13	52	22	13

Note. Table includes only respondents who plan to do research on occupations and careers in the next few years. N = 168.

In some cases the public data sets probably provide more relevant data than respondents may be able to collect on their own. For example, some of the data sets provide detailed job histories which could be examined to test the still largely untested theories of career development within the field. About 36% of the respondents place a high priority and 30% place a moderate priority on obtaining job history data. Nevertheless, only about a quarter of the respon-

dents were familiar with some of the major longitudinal studies that collect extensive job history information--e.g., the four National Longitudinal Surveys by the Ohio Center for Human Resource Research and the ISR Panel Study of Income Dynamics. In addition, three other ISR surveys would probably be of particular interest to vocational researchers (the Survey of Working Conditions and the two Quality of Employment Surveys) because of their focus on job satisfaction and adjustment, but only about a third of the respondents were familiar with these data sets. These results suggest that lack of familiarity is a major reason for the non-use of relevant public data.

Discussion

The mail survey of authors revealed that public occupational data are essential or a major source of data for no more than 11% of the respondents. The analysis of journal articles showed that less than 4% of the articles during a five-year period actually used any public occupational data, even in a minor way. Of the 36 data sources in the survey, the average number ever used was 3 or 4. Only a few of the 36 had been used in a major way by more than 2% of the respondents, published reports and the DOT being the most widely used. The modal response was that public occupational data are "somewhat useful" although none was used in research in the last five years. Respondents had heard about a third of the data sources on the average, but the modal response was "not know about" for 31 of the 36 data sources. In short, public occupational data are important to a small minority of vocational researchers, but they are of only peripheral and occasional use to most of the rest.

There did seem to be a potential interest in using more public data, however, because over half the respondents said that improving the availability and quality of public data would be "probably a big help," and many requested additional information about the data sources in the survey.

Explaining the Use and Non-use of Public Data

The infrequent use of public data could be explained if the data were not relevant to the concerns of vocational researchers, but an earlier discussion argued that this is not an adequate explanation. Other explanations emerge from the results when interpreted in terms of the data selection model presented earlier. Those results suggest that the major current barrier to the use of public data is lack of familiarity--the first stage in the data selection model. With the exception of the DOT and several published census report series, only about half the respondents were aware of major data sets from the Census Bureau and the Department of Labor. And generally less than a quarter to a third of respondents knew about the different non-governmental data sources. As already mentioned, respondents on the average did not know about two thirds of the 36 data sets.

The number of data sets respondents were familiar with was related somewhat to their discipline, with sociologists and economists being more familiar with public data. This is not

surprising because the use of public data is greater in those two disciplines whereas in psychology such data are generally considered "non-psychological" and thus presumably not worth looking into. Having been closely affiliated with a research center and years since one's last degree (tenure in the field) were also associated with a greater familiarity with the data in the survey. Both these factors could be expected to increase the likelihood that a researcher would be exposed to a greater variety of data sources and to learn about them. Looking at the data sets that are most widely known, it appears that sets may be better known when the organizations which sponsor and produce them have as a primary mission collecting and disseminating occupational data.

If we assume that searching for information poses costs in time and money for the researcher, the more the environment exposes researchers to information about particular types of data, the more likely researchers are to become familiar with them. The factors illustrated here--tenure in the field, discipline, being in a research center, and publicity by the data producer--all seem to accomplish this, in this case for public data.

Familiarity is a necessary but not sufficient condition for the use of any particular type of data. As noted before, cost, convenience, and quality of the data become important considerations once researchers begin to seriously consider using any data set. For example, it is generally known that vocational researchers are not familiar with using large data sets, so it is highly unlikely that many of them already have in place the computing skills, facilities, and staff that would be required to process the large data sets. Without technical advice and assistance from some source, the large data sets are overwhelming to the uninitiated. Even for the veteran they often strain patience and resources.

Results from the content analysis suggest that use of large public data sets is more likely where clusters of individuals or resources exist, e.g., in research centers. Anyone who works with colleagues who are already using a particular set of data faces many fewer costs in deciding to use the same data. This was surely the case with the five authors who were affiliated with the research centers that produced the public data they used in their research. In such cases the selection of public data for research probably operates according to the same principles as the selection of non-public data--they are readily available and clearly usable in the immediate environment. Examples of such clusters of data use can also be seen in the articles using non-public data (results not shown here). Colleagues can provide detailed information about the data, and they most likely can also provide the data themselves free or at reduced cost. In addition, research centers provide a pool of personnel and other resources which may be necessary to handle some of the larger, more complicated and costly public data sets which would be impractical for more isolated investigators. In cases where the public data have been acquired by the researcher himself or herself, the research center may play primarily the latter role of providing resources such as specialized staff, computing facilities,

and collective expertise in obtaining and using costly or unwieldy data. Such resources are extremely helpful, if not necessary, for using some of the large national surveys or censuses.

Although the existence of colleagues and collective resources may contribute to the use of particular types of data--particularly the large non-traditional sets of public data--not all types of public data require such resources. For example, some of the journal articles in this study which used non-traditional public data involved coding information from published reports, and surely required no more effort, money, or specialized help to use the data than was expended in many of the several hundred articles which collected new data. So in some cases other aspects of the data could conceivably become the stumbling blocks to greater use--the variables or populations covered in the data, presentation of the data in a form (e.g., in categories) meaningful to the investigator, and so on. This study collected evaluations of these data qualities, and these will be analyzed elsewhere.

Suggestions

Much of the current "problem" with public data in vocational research is simply that vocational researchers know little about it compared to their usual sources of data. At this stage, then, what is needed is to disseminate more information about public data. This is also the suggestion made most often by respondents. The first aim of the information would be to create a greater awareness of the data and the second would be to make the data easier to obtain and use if researchers were to become interested in it. Specific suggestions are given below.

Several types of information could be distributed. One is simply a list of data archives, where to obtain catalogs of the holdings of these archives, and a general idea of the types of data they have. Vocational researchers could also be introduced to the services provided by the government. For example, researchers could be made aware that they can be placed on mailing lists for press releases, announcements, and reports from the Bureau of Labor Statistics, and that there is a Data Users Service at the Census Bureau that both distributes data and runs courses on federal statistics. Another type of information is brief descriptions of specific data sets which might be particularly useful to vocational researchers, and perhaps references to research which has already used the data. Both machine-readable data sets as well as published data in government reports could be described. There need not be many such descriptions, but a few data sets could be publicized in order to show concretely that public data do have benefits for vocational researchers. But the major focus should be on making clear that diverse types of data are available and making clear how to get them. It is very important that all descriptions of data archives, data services, or data sets themselves be accompanied by a name, address, telephone number, or reference by which the researcher can obtain more information or obtain the data itself. Once people know that particular data sets are available,

providing important details increases the convenience and lowers the costs of checking into and using the data.

To whom should this information be distributed? Research directors (at research centers or organizations) and department chairs could be sent a package of materials about government-sponsored public data. Journals and newsletters already reaching vocational researchers could publish information. The mail survey showed that the APA Monitor, the newsletter of the American Psychological Association, was the most commonly read journal; it was read by 68% of the respondents. Advertisements could be placed in the widely read journals for useful sources of information such as the Data User News or the Monthly Labor Review which are read by relatively few (respectively 2 and 18%) of the respondents. If convinced of the potential utility of public data to vocational researchers, journal editors and other gatekeepers in the discipline might actively support this dissemination effort. Their participation would, in turn, lend legitimacy to public data, data which are typically perceived as "non-psychological."

Another suggestion from respondents was that symposia or training sessions be presented at professional meetings such as those of the American Educational Research Association and the American Psychological Association. These sessions could review some of the major data sets, presentations being made by both a producer and a user of each data set. Such sessions would not only familiarize people with the content, source, uses, and problems of the data, but also allow vocational researchers to ask questions and make contacts with other people with whom they might pool resources or exchange information. Such sessions would be a useful first step in stimulating greater use of public occupational data among vocational researchers.

References

- Ashley, W. L. Occupational information resources: A catalog of data bases and classification. Columbus, Ohio: Ohio State University, Center for Vocational Education, 1977.
- Campbell, D. Handbook for the Strong Vocational Interest Blank. Stanford: Stanford University Press, 1971.
- Crites, J.O. A comprehensive model of career development in early adulthood. Journal of Vocational Behavior, 1976, 9, 105-118.
- Gottfredson, L. S., Voorstad, F., and Simonsick, E. An evaluation of public occupational data for vocational researchers. Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools, in preparation.
- Holcomb, W. R. & Anderson, W. P. Vocational guidance research: A five-year overview. Journal of Vocational Behavior, 1977, 10, 341-346.
- Holland, J. L. Making vocational choices: A theory of careers. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.
- Holland, J. L. Manual for the Vocational Preference Inventory. Palo Alto, California: Consulting Psychologists Press, 1975.
- Inter-University Consortium for Political and Social Research. Guide to resources and services, 1976-1977. Ann Arbor, Michigan: University of Michigan, Institute for Social

- Research, 1978.
- Inter-University Consortium for Political and Social Research. Data acquisitions, 1977-1978. Ann Arbor, Michigan: University of Michigan, Institute for Social Research, 1978.
- Inter-University Consortium for Political and Social Research. Data acquisitions, 1978-1979. Ann Arbor, Michigan: University of Michigan, Institute for Social Research, 1979.
- Krumboltz, J. D., Mitchell, A. M., & Jones, G. B. A social learning theory of career selection. In J. M. Whiteley & A. Resnikoff (Eds.), Career counseling. Monterey, California: Brooks/Cole, 1978.
- Roe, A. The psychology of occupations. New York: Wiley, 1956.
- Super, D. E. The psychology of careers. New York: Harper & Row, 1957.
- U.S. Bureau of the Census. Reference manual on population and housing statistics from the Census Bureau. Washington, D.C.: Bureau of the Census, Subscriber Services Section, March, 1977.