APPENDIX G

Using Computers to Match Workers and Jobs: A Preliminary Assessment of the U.S. Employment Service's Automated Matching System

CHARLES F. TURNER

The task of matching workers and employment opportunities requires extensive record-keeping. Files of job openings must be continually updated; jobs that have been filled must be removed from the files and new listings must be added. A similar process occurs for files containing descriptions of available workers. To be kept current, these files must also be regularly updated.

Most important, there must be regular searches of the files to find suitable openings for each worker and suitable workers for each opening. The matching of workers with available jobs is, after all, the reason for the existence of a government-funded employment service. One ideal for such service is the following:

1. Each time a worker comes to the Employment Service, a search will be made for an appropriate opening currently on file with the Employment Service.
2. On a daily basis each worker on file at the Employment Service will

---

Since this work was requested late in the lifetime of the committee and since it required detailed familiarization with the operation of a complicated computer system, this undertaking was assigned to a member of the committee's staff. While the following analyses have been discussed by the committee, the responsibility for this review resides with the staff member who undertook this work and authored this appendix.

be compared with all available employment openings to determine if any available jobs are suitable for him or her.

To appreciate the magnitude of the record-keeping and file-searching involved in the operation of the Employment Service, let us consider the situation in one city we visited. During the month of January 1978 the local Employment Service office in Houston, Texas, listed 2,829 job openings and 3,884 workers registered with the office.

In a fully comprehensive, unstructured search of the file of job openings, one would have to make nearly 11 million individual comparisons to determine whether there were any appropriate openings for each of the 3,884 workers registered with the Houston Employment Service. The enormous magnitude of this task dictates that comprehensive and unstructured file searches should be avoided. Clearly, a more efficient strategy of searching for appropriate matches of workers and jobs is required.

For this reason the Employment Service has traditionally restricted the scope of the file searches that are routinely performed. Workers are never compared, one by one, against the entire file of available jobs. Instead, each worker and each job order are assigned to a narrow occupational category. This traditionally has been done by selecting one of the approximately 12,000 base titles from the DOT to describe each worker and each job. Ordinarily workers are matched against only the file of available job openings in the particular occupation in which they were coded.[1] The use of this DOT-based search strategy dramatically reduces the amount of file-searching required to match workers and jobs.

The efficiency of the DOT matching strategy can be seen in our example. If the 3,884 workers and 2,839 job openings in Houston were uniformly spread across 300 occupations (an improbable event, we admit), the number of individual comparisons between job orders and worker descriptions required for a complete file search would be reduced from approximately 11 million to 37,000. If the workers and job openings were spread over a larger number of occupational titles, the number of comparisons required decreases, assuming a uniform distribution of workers and jobs across occupations. To the extent that the distributions become uneven (i.e., some occupations having more workers or jobs than others), the number of comparisons will increase. At one extreme one has the case of one worker and one job per occupational title; in this case the number of comparisons required to perform a DOT search is exactly equal

[1]To accommodate workers who have experience or qualifications for more than one occupation, duplicate occupational assignments are sometimes made.

to the total number of workers in the files. At the other extreme, if all workers and jobs are in the same occupation, a DOT-based search strategy is the same as an unrestricted search (i.e., nearly 11 million individual comparisons would be required in the example).

The DOT file searches are considerably less burdensome than unrestricted or unsystematic searches. File-searching is reduced by simply limiting the number of jobs for which workers will be considered and, likewise, by restricting the number of workers who will be considered for each job opening. Such a strategy in many cases seems appropriate. There is little reason to compare individually a group of records describing workers with experience in the performing arts to job orders for scientists. However, it is not clear that this strategy is always in the best interests of workers and employers. In particular, we note that this strategy assumes that there is no transferability of skills between occupations. Moreover, for occupations that do not have formal entry qualifications (educational requirements or occupational licenses), such a search strategy unnecessarily restricts the opportunities of workers to find employment.

There are, of course, hybrid search strategies that preserve the efficiency inherent in the use of a single classification structure to categorize workers and jobs but avoid locking in workers to the specific occupations in which they have previously worked. Those methods are briefly described in a later section of this appendix.

First we will consider how the Employment Service has automated the operation of its local offices. In considering this program we note that our interviews with the staff of the Department of Labor and local Employment Service offices and the studies conducted by others (e.g., U.S. General Accounting Office, 1978) indicate that DOT applicant and job order files are not frequently searched for matches of applicants with new job orders. Only one third of all referrals resulted from manual file searches done subsequent to an applicant's initial appearance in the Employment Service office. Data such as these were used to support the original proposal for the automation of the Employment Service's placement activities.

AUTOMATION AND JOB-WORKER MATCHING IN THE EMPLOYMENT SERVICE

An initial distinction must be made between the automation of the Employment Service's operations and the particular type of computerized method it uses for matching workers and jobs. The latter system, called keywording, is not the only way of matching workers and job orders by computer. Matching could, for example, be performed using the tradition-

al DOT-based search strategy; the file-searching and record-keeping would be done by the computer rather than manually. Indeed, the present keyword matching system includes an option that permits the automated matching of workers and jobs using DOT codes. The computer searches for an exact match between the DOT code(s) assigned to a particular worker and those on job orders in the job bank.

The automation of the clerical and record-keeping tasks involved in job placement has many obvious benefits. Those benefits, however, accrue to automation, not to the keyword system. In this appendix we do not comment on the effects of automation per se; rather we assess the implications of automation for the Employment Service's occupational analysis program and, indirectly, for the quality of the job-worker matching done by the U.S. Employment Service.

DEVELOPMENT OF SYSTEMS

In the early 1960s the Employment and Training Administration (then the Manpower Administration) began to support studies of automated job-worker matching. Initially, funding was provided to the states of California, New York, Utah, and Wisconsin. Automated matching systems based on aspects of the *Dictionary of Occupational Titles* were developed in New York and Utah.

The experimental system developed by New York matched jobs and workers using DOT codes. Members of the committee and staff viewed the operations of this system during a visit to the occupational analysis field center in New York. The system in its present form serves the City of New York; not only does it make file searches by direct matching of DOT codes, but it also incorporates a hybrid procedure that allows workers with one DOT code to be matched to job orders with different DOT codes that are thought to be sufficiently similar to permit suitable referrals. For example, in the New York matching system the computer first searches its files using an automated version of the manual DOT search strategy; if this search is unsuccessful, it may be augmented by the use of a cross-reference list of similar occupations stored in the computer. For each DOT code the computer refers to this list to determine what DOT codes will provide suitable matches for a particular applicant (or job order). So, for example, when a Batter Mixer (520.685-014) comes to the New York City office seeking appropriate employment, she or he might be considered not only for openings as a Batter Mixer but also for those as Blender Machine Operator (520.685-018), Cake Stripper (520.685-042), Candy Puller (520.685-046), Confectionery Drops Machine Operator (520.685-078), etc.

The initial judgments about what are similar occupations were made by staff of the Employment Service, and they can be continuously revised.

While we cannot evaluate the overall adequacy or comprehensiveness of the judgments of similarity incorporated in the New York system, the New York matching system illustrates how flexibility might be built into a hybrid system based on an explicit and detailed occupational classification such as the *Dictionary of Occupational Titles*.

Work by the Employment Service in other states took different paths. Some attempted to define occupational similarity, and others attempted to develop descriptive vocabularies. A brief catalogue of these attempts is useful in appreciating the history of the current keyword system.

In Wisconsin an experimental system was developed using the worker trait arrangement (GED, SVP, physical demands, etc.) of the DOT to define the similarity of occupations, and thereby to specify the sorts of job-worker matches that might be made. In contrast, an experimental system originally developed by the State of California does not rely on the *Dictionary of Occupational Titles*. The California system, known as the Labor Inventory Network Communication System (LINCS), is similar in many fundamental ways to the keyword system (described below). A distinguishing characteristic of the LINCS system is its attempt to provide a language for describing occupations that includes natural language elements; this system does not rely on an explicit, predetermined classification of all occupations.

The early development of these automated systems for the centralization of information about job openings and available workers reflected the recommendations made by various advisory bodies. For example, the National Commission on Technology and the American Economy recommended in 1966 that the government establish such computerized placement systems to assist workers dislocated by automation.

During the late 1960s and early 1970s, further experimental systems for job-worker matching were developed. One system, Job Analysis Vocabulary (JAV), was developed using the language of job analysis to describe the characteristics of particular jobs and the experience of individual workers. This system explicitly rejected the concept of "occupation" as the bridge for matching workers and jobs. Instead, it attempted to describe the characteristics of particular jobs, without fitting them into a preconceived set of occupational categories. The developers of this system envisioned that Employment Service interviewers would do very abbreviated job analyses of each opening in their files. The presence of common task-related terms (e.g., driving van, collecting money, delivering products, etc.) in the job description and the history of an applicant would produce matches.

The great advantage of such descriptive systems is flexibility. Once we allow the idiosyncrasies of each job to be faithfully mirrored in a matching system, it becomes possible to tailor job specifications to the needs of a particular employer. Moreover, the system ensures that all of the idiosyncrasies of a given job will be used during the process of job-worker matching. In such a system, workers and jobs can be described and matched even if they contain elements of two or, theoretically, hundreds of occupations. For example, a school that needs a guidance counselor who can also teach Greek can have a job order written to specify that experience in both of these occupations is a prerequisite for employment.

A second-generation LINCS system was subsequently produced; it is known as the Detailed Computer Assisted Language (DECAL). While the original LINCS system was restricted to a range of technical and professional occupations, DECAL covers the entire occupational spectrum.

After a series of field studies of the performance of state Employment Service offices using the various systems, the Division of Automated Matching of the Employment Service concluded that the second-generation automated systems (JAV and DECAL) have a demonstrated superiority to manual methods of job-worker matching. The Employment Service subsequently committed itself to an automation program estimated to cost $250 million (U.S. General Accounting Office, 1978).

KEYWORDING: THE EMPLOYMENT SERVICE MATCHING SYSTEM

The matching system resulting from this experimental work was released in 1975 by the Division of Automated Matching of the Employment Service. At present, 23 states have received funding for automation, and the Department of Labor anticipates automating Employment Services across the nation by 1984.

Our interest in automated matching focuses on its characteristics as a tool for facilitating the employment and mobility of labor. A survey by the Bureau of Labor Statistics (U.S. General Accounting Office, 1977:6) indicated that one third of all job seekers use the U.S. Employment Service. A system that increases the opportunities of these workers to find desired jobs would make a significant contribution to the country's well-being.

The present keyword system used by the Employment Service incorporates many features that are a natural concomitant of automation. For example, the system permits automated elimination from matching of jobs

that do not satisfy workers' criteria for hours of work, geographic location, salary, etc. Thus workers who specify that they will not work night shifts are never matched against jobs requiring night work. Similarily, workers who are not willing to travel from their homes to jobs in distant locations are not considered for such openings. These aspects of keywording are an obvious use of computer capabilities to assist in the matching of workers and jobs. As noted earlier, since such features are not unique to keywording, we will not consider them at length. However, we do note one inefficiency of the present keyword system. It arises because Employment Service counselors must code keywords describing applicants with a numerical code. Thus despite the fact that the keyword system uses ordinary English words to describe applicants and jobs, it is necessary for the interviewers to consult a code book to translate these words into a five-digit numerical code. This task is one that could be easily performed by a computer. It is inefficient to require human interviewers to recode the standardized keyword vocabulary into numeric digits. We suspect that this requirement decreases interviewers' efficiency and increases the occurrence of errors in the coding of workers and jobs.

DESCRIPTION

The keyword system is based on a vocabulary that segments all occupations into 36 occupational areas (occupational units). The organization of the occupational units reflects the early concentration of the LINCS system on technical and professional occupations. Thus there are nine separate units describing professional occupations. Indeed, there are separate units for writing, artwork, and the performing arts, even though the Employment Service receives relatively few job openings in these three areas. As Table G-1 shows, approximately .4 of all workers are concentrated in 4 of the 36 keyword occupational units: clerical, service, general labor, and business administration.

The major reference document for the keyword system is the *Handbook of Occupational Keywords* (U.S. Department of Labor, 1972). The *Handbook* contains a complete listing of the descriptive vocabulary used by the keyword system. The description of the 36 occupational units has much in common with traditional classifications of occupations. Indeed, its descriptions of the occupational units contain an explicit list of the occupations for which each unit is appropriate.

This aspect of the keyword system is of particular interest, since the developers of the keyword system claim that it is not an occupational

TABLE G-1 Estimate of the Distribution of the National Labor Force by Keyword Occupational Units

Occupational Units	Percentage of the Labor Force
Scientific/technical	
(1) biological sciences	0.2
(2) health sciences	4.9
(3) physical sciences	0.8
(4) social sciences	1.1
(5) engineering/drafting	3.5
Education and the arts	
(6) education/library	6.3
(7) writing/translating	0.3
(8) artwork	0.8
(9) performing arts	0.2
Marketing and sales	
(10) insurance/real estate	1.8
(11) sales	2.1
Business	
(12) business administration	5.2
(13) electronic data processing	0.6
(14) clerical	14.2
(15) inspecting/quality control/appraising	1.7
Construction, structural and extraction	
(16) construction/metal structural work	5.9
(17) welding	1.1
(18) installation/maintenance/repair	6.5
(19) construction/shop painting	0.9
(20) mining/drilling	0.4
Manufacturing: machine work	
(21) machining/machine operating	5.6
(22) woodworking	0.2
(23) cloth/leather working	2.6
Manufacturing: fabricating	
(24) printing/paperwork	0.8
(25) technical/electrical equipment fabricating	0.1
(26) metal products fabricating	0.4
(27) mixed materials fabricating	0.0
(28) rubber/plastics fabricating	0.0
Manufacturing: processing	
(29) stone/clay/glass/sand working	0.4
(30) processing	0.4
(31) textile preparation/finishing	0.5
Services	
(32) services	6.1
(33) transportation	5.5
(34) utilities/communications	0.7

Continued overleaf

TABLE G-1 (continued)

Occupational Units	Percentage of the Labor Force
General labor and general services	
(35) agriculture/fishing/logging	4.6
(36) general labor/general services	13.7

SOURCE: This tabulation was derived by assigning each of the 441 detailed census categories from the 1970 Census classification of occupations to keyword major groups and then retabulating the occupational distribution for a 4-percent sample of the 1970 Census. Assignment of keyword occupational units to the detailed census codes was done for us by staff of the U.S. Employment Service's Division of Automated Matching Systems. Because we have been forced to use the census occupations as a link in estimating this distribution, this tabulation represents only a rough approximation of the true distribution. There is, however, no feasible way to generate more reliable estimates at the present time.

classification. The fact that the *Handbook* lists almost 600 occupational titles in its description of the occupational units suggests otherwise.[2] Furthermore, the primary keywords listed in the *Handbook* appear in general to be occupational titles that have been grammatically transformed, e.g., sociology work (sociologist), anthropology work (anthropologist), tool making–die making (tool and die maker), motor generator assembling (motor-generator assembler), drama directing (drama director). Indeed, where this is not the case one finds special notes identifying the relevant occupational title, e.g., the occupational keyword: "Death Investigating" Work (56310) contains a special note advising the reader: "identifies work of coroner."

For each of the 36 occupational units the keyword system provides a special vocabulary of terms to describe work in the unit. These terms consist of two types: primary and complementary. The primary terms are meant to describe the major work activities. For example, the unit Biological Science contains primary terms for botany, microbiology, zoology, etc. The complementary keyword terms cover such things as materials, products, equipment, or tools associated with particular areas of work. In addition, for units covering professional occupations, complementary terms may also describe subject areas and languages involved in the work. For example, the Biological Sciences unit contains complemen-

[2]It should be noted that the 1970 Census classification of occupations contains many fewer occupational titles than those listed in the *Handbook of Occupational Keywords.*

tary terms describing educational areas, such as "Environmental Science." By coding this keyword one can specify particular educational characteristics of a job or worker. Similarily, occupational keywords for this unit describe research specialties as well as the various types of plants and animals that might be studied by someone doing agricultural research. Finally, there is a set of complementary terms describing the various locations in which work might be performed, e.g., a medical laboratory, a botanical garden, etc. Complementary keywords are specific to an occupational unit and its associated primary terms; thus selection of a primary term, e.g., biological work, constrains the domain of complementary terms that will be used to describe the work.

To use the keyword system to code a job or worker, an Employment Service interviewer must do the following:

1. The appropriate occupational unit or units that characterize the worker (note that primary terms from more than one occupational unit may be used to describe workers or jobs) must be determined.
2. All primary terms needed to describe the major work activities must be selected.
3. All appropriate complementary terms to define the other characteristics of the work or worker must be selected.
4. He or she must code the information concerning the geographic location of the job and worker's residence, salary levels offered by the employer and acceptable to workers, and certain other specific matching information, e.g., whether the job requires an occupational license and whether the worker holds such a license; shifts worked; job duration; work week; etc.
5. Finally, for job orders, the keywords must be formatted into simple sentences to describe the exact nature of the employer's requirements. The vocabularly permits two types of specification for job qualifications. A characteristic may be "required" or "desired." When the former is specified, workers lacking this characteristic are excluded from consideration. When more than one worker meets all required criteria, the system will select those workers having more of the desired characteristics. A job order may specify the characteristics of jobs in terms of (required or desired) education, experience, knowledge, or skill.

Descriptions of jobs are constructed by first specifying the major field of work and then listing all of the desired and/or required skills, experience, education, or knowledge specified by the employer. For example, one might have a job order coded as follows: (1) primary term: Botany, (2) complementary terms: master's degree in Botany required, 3 years'

experience in microbiology required, skill in research desired, and knowledge of statistical methodology desired, and (3) non-occupational terms: work week, 40 hours; job duration, 150+ days; shift, first shift (normal day); working conditions, inside; public transportation, not available; geographical location, special map code; and salary, $20,000/year. Workers, in turn, are described in terms of their work experience and education. Experience is quantified in terms of the number of months worked, and education is coded into six descriptive categories: special courses, vocational-technical school training, associate's, bachelor's, master's, or doctoral degree.

Matching is performed by searching for instances in which the requirements of a given job are satisfied by a worker's education and experience. In the matching strategy used by the keyword system, the primary terms, which describe major work areas such as biological work, have a logical precedence in matching over the complementary terms. Skill and knowledge requirements in job orders may be satisfied by either the worker's formal education or his or her work experience. The information coded about salary levels, geographic locations, etc. is used to eliminate job-worker matches on non-occupational criteria.[3] If many applicants fulfill all of the job requirements, i.e., have coded on their file all of the keywords that were required on the job order, they are rank-ordered by the number of desired keywords that the applicant and job order have in common. The applicant with the most keywords in common with the job order is the first-ranked candidate for referral.

EVALUATING KEYWORD MATCHING

Many aspects of the keyword matching system can be evaluated conceptually. Evaluation of the impact of keywording on the actual functioning of the Employment Service, however, requires data that were not available to us—data that, in fact, have not been collected.

Although the evaluation studies (U.S. Department of Labor, 1976a) conducted by the Employment Service's Division of Automated Matching concluded that keyword matching was demonstrably superior in terms of efficiency, cost, and quality of placements to manual search methods or alternative automated systems, two important inadequacies in the design of these studies make these conclusions questionable.

[3]We have ignored in this discussion the use of non-occupational criteria in matching; the principles are the same but a number of the criteria involve simple exclusion rules (e.g., exclude night work jobs if applicants want day work) or simple *ad hoc* formulae (e.g., exclude jobs that offer less than 80 percent of an applicant's salary requirement).

First, the studies did not distinguish between the effect of automation per se and the effect of particular matching systems.[4] Since automation could be accomplished in a variety of ways other than keywording, one cannot confidently attribute the observed differences between manual procedures and automation to keywording. It could instead be argued that all of the effects were due to factors such as the automation of record-keeping, changes in the routing of applicants through Employment Service offices, changes in management procedures induced by automation, and so forth. One theoretically could even argue that keywording may have diminished the otherwise salutary effects of automation. The evidence provided by the studies is simply inconclusive on this point.

Second, the studies made no estimates of the effects of temporal fluctuations in the economy on the operations of the Employment Service. Neither experimental controls nor any adjustment procedures were used to allow for effects of changes in the national economy or the local labor markets on the probability of a worker's finding employment.

In our assessment the available evidence (i.e., U.S. Department of Labor, 1976a) is insufficient to support the conclusion that keywording improved efficiency. A recent study by the General Accounting Office (U.S. General Accounting Office, 1978:iii) has independently reached a similar conclusion:

> Although experiments with computerized job matching within the Employment Service have been going on for 10 years, Labor has not demonstrated that this system is an effective way to find jobs for people and people for jobs. . . . Labor's evaluations, past and planned, do not differentiate between improvements due to computerized job matching and those due to other factors, such as changes in the economy or organizational and procedural changes.

We also wish to note, however, that while the evidence is not convincing in the affirmative, neither is it convincingly negative. It is simply not known whether keyword matching produces more and better matches of workers and job openings than manual methods.

[4]Controlled comparisons of alternative automation systems could provide information in this area. Data from the field studies conducted by the Employment Service, however, are inadequate for these purposes because the field site and type of automation system are entirely confounded (sites had one or the other type of system). No explicit comparative data are presented; the authors of the final report observe (U.S. Department of Labor, 1976a:132):

> Throughout the experimentation, attempts were made to separate out the effects the individual vocabularies [i.e., matching systems] had on the performance data presented in the analysis. This proved most difficult considering the multiple number of variables that may affect the data and especially so in view of the major changes in the economy during the periods covered by the experiments.

It should also be noted that site visits by committee members and staff and statements by state Employment Service officials lead us to suspect that keywording procedures require more personnel time for interviewing of applicants and employers. Keywording also requires the purchase of computer hardware. Since it appears that the capital costs (about $250 million nationwide) involved in keywording may not be offset by decreased labor costs, the lack of convincing evidence that keywording produces more or better placements is disconcerting.

It was neither the intention nor charge of this committee to conduct a field evaluation of the costs and benefits of the keyword matching system. The committee's charge was to assess the need for the *Dictionary of Occupational Titles* and the occupational analysis program of the Employment Service and to suggest avenues for future research, particularly in light of the planned automation of placement services. Upon the subsequent request of the Employment Service the committee agreed to consider briefly the conceptual bases of keyword matching, its potential utility to the Employment Service, and the nature, scope, and appropriateness of various technologies of classification.[5] In this vein we offer three general comments and one particular comment on the manner in which the present keyword matching system operates.

Use of Information

Because machines are (or can be, if properly programmed) more efficient bookkeepers than humans, the automation of local Employment Service offices makes possible new methods of matching workers and jobs that would be completely infeasible otherwise. Many of these opportunities arise because complete information on the work histories, referrals, and placements of workers can be used to guide local offices in deciding which workers make "good" referrals for particular job openings. Similarly, automated procedures could, theoretically, allow for the tailoring of

[5]This committee was not originally charged with evaluating any aspect of the Employment Service automation program. Late in the life of the committee (January 1979) the staff conferred with representatives of the Employment Service and the Employment and Training Administration to discuss the committee's charge and the Employment Service's policy with regard to automation. As a result of these discussions, representatives of the Employment Service asked that some consideration be given to the conceptual bases of keyword matching and its potential utility to the Employment Service in comparison with other possible systems. We were not asked and could not, given the constraints of time and resources, conduct new field experiments to gauge the effects of the introduction of automation in different forms on Employment Service performance. Similarily, we could not design a completely new matching system. Rather, we agreed to raise conceptual questions and consider the nature and scope of appropriate and inappropriate technologies for classification.

matching strategies to the idiosyncrasies of local labor markets, or they could allow the matching strategies to adapt across time to changes in the functioning of the labor market. Unfortunately, the present keyword system does not exploit these possibilities.

To illustrate this point, let us consider how a job-worker matching system might use the information routinely collected by the Employment Service to improve placements and allow for changes across time or differences between labor markets in the types of job-worker matches that are appropriate.

At present, the full histories of workers enrolled at Employment Service offices are obtained by interviewers and coded onto the applicants' records. Indeed, annual national samples of several million of these worker histories are routinely collected by the Employment Service's Automated Reporting System (ESARS) and filed on computer tape in Washington, D.C. Our inquiries, however, indicate that these data are rarely, if ever, used.[6]

How might such data be used to improve job-worker matching? One way would be to build an adaptive matching system that used these data to learn what are "good" and "bad" matches. For example, over a period of time, thousands of secretaries obtain jobs through local Employment Service offices in any given state. There are also some entry-level executives who find jobs through Employment Service offices, and secretaries will sometimes be referred to entry-level executive jobs, albeit with lamentable infrequency. A computerized system that has been programmed to "learn" from its past experience in making referrals could over time recognize patterns of success or failure. The local Employment Service office has information on whether a person whom it referred was hired, and since unemployment insurance claims are also automated, information on how long a worker remained with an employer could also be obtained. Assuming that over time one found that ex-secretaries were hired as entry-level executives and did not appear as claimants under the unemployment insurance (UI) program, a matching system programmed to learn would begin to make more referrals of secretaries to entry-level executive positions. Alternatively, simply analyzing the work histories of Employment Service applicants (without resort to UI records) could provide similar information about the types of "matches" that are possible

[6]The worker history data, taken as nine-digit DOT codes, are keypunched through the first two digits only. This truncation, which could easily be avoided, precludes the use of these valuable data to evaluate the interoccupational mobility of Employment Service registrants. One suspects that if these data were actually used to study Employment Service operations and ways of better matching workers and jobs, the remainder of the DOT code would not have been discarded.

in a given labor market. In any case, with automated procedures the available job and referral histories could be used to improve job-worker matching.

Let us take another example. Consider a hypothetical labor market in which large numbers of engineers are laid off owing to cutbacks in the aerospace industry, while simultaneously, a computer firm opens a research center in the area. Assume that some aerospace engineers also have skills that are attractive to the computer firm, (e.g., in software development, telecommunications, and programming) and are hired for research and sales jobs. Assume also that some of these workers subsequently register with the Employment Service (perhaps seeking work in their former occupation). Their work histories are routinely recorded and coded. An adaptive computerized matching system would note that there were a number of computer-related occupations into which former aerospace engineers had transferred in that labor market. Thus even though most occupational classifications would not consider the occupations very similar, the labor market information routinely gathered by the local Employment Service might reveal that there was enough transferability of skill and knowledge so that some displaced aerospace engineers could find jobs in new occupations in the computer industry and thus were reasonable matches.

Such adaptive matching strategies use capabilities inherent in automation. Such matching strategies are dynamic; they can change over time. They can adapt to idiosyncrasies in local labor markets. They learn from experience. With such systems the labor involved in collecting information on each applicant's skills and qualifications and maintaining administrative records on the "hires" and "non-hires" that result from referrals becomes a resource for learning how to make better placements. This information teaches the system how to serve workers and employers better.

We are, however, describing a rather different system from the current keyword matching system, which has no facility for learning. In this respect it replicates current practices and overlooks the natural advantages of using machines as record-keepers. In a manual search one cannot continually revise matching strategies; the record-keeping and analysis would strain the resources of a staff many times larger than that of the Employment Service. However, machines, when they are properly programmed, are superb cataloguers, recorders, and tabulators of such information. Since all the necessary information for such an adaptive matching system is routinely collected and encoded by the Employment Service computers, no further staff time would be required—just a different computer program.

We suggest that the failure of the keyword system to use administrative records to evaluate and refine its matching strategies is its first and most important conceptual inadequacy.

The Definition of Similarity

In any system of matching, the crucial question is, How similar must two objects be to "match"? This is a basic question of classification. For some things, there appear to be natural categories: plant or animal, female or male, alive or dead. Distinctions such as these are not without anomalies (e.g., euglenae, hermaphrodites, and the brain dead), but they are useful classifications in many spheres of everyday life. In other areas, distinctions may involve questions of degree rather than categorization, such as the difference in rainfall between London and Paris or the difference in speed between a train and a plane. In some situations the similarity between two objects may be imperfect but sufficient for one to treat the objects as being equivalent.

Parallel questions arise in matching occupations. Some occupations are sufficiently similar that a worker in one occupation can be substituted for a worker in the other occupation in most circumstances. For example, for many jobs it is of little importance whether a statistician is an applied statistician or a mathematical statistician. In particular, if one asks only elementary questions about sampling theory or the measurement of association, either will do perfectly well. There are, of course, instances in which greater specificity will be required, for example, a university hiring someone to teach a course in applied statistics.

To judge the similarity between workers and jobs, one must make judgments about the transferability of skills between occupations. It must be decided how similar job requirements and worker attributes must be in order to match. The keyword approach to this problem is simple to state.[7] In the keyword system, jobs and workers are assigned to one of 36 major occupational units; a primary term from that unit describes the major aspect of the work. As noted above, these primary terms consist almost exclusively of occupational titles that have been grammatically transformed, e.g., anthropology work (anthropologist). The similarity between these primary terms is defined only in the special case in which one primary term is a subclass of another. For example, "sociology work" is subdivided into two subclasses: urban sociology work and rural sociology work (a division sociologists will, no doubt, find bizarre). Only in such

[7]In this discussion we leave aside matching criteria that are not occupational, such as geographic location (see footnote 3).

instances is there a procedure (called tree codes) for matching persons and jobs coded with different primary terms. Thus if the interviewer specifies a special search strategy called "explosion," "rural sociologist" will be considered for job openings coded simply as "sociologist."

Because the similarity of primary terms and hence occupations is largely undefined in the keyword system, there is no way to build linkages across work areas. Instead, "complementary" terms describing less central characteristics of occupations are used to accomplish such matching. Let us consider an urban sociologist. One might code experience in "research" (38063) and education plus experience in the study of migrants (52205) as complementary terms. How would this person be matched against a job file containing no openings coded with the primary term "sociology work"?

The similarity between the applicant and job orders is scored by counting the number of complementary keywords that the applicant and order have in common. Since there were no openings for sociologists, it is likely that there would be no order coded with the complementary term for the study of migrants. There would be, however, a number of jobs that would have research (38063) coded as a keyword. These might be in agricultural research, petroleum engineering research, chemical research, etc. Since the keyword system merely counts the incidence of matches on the complementary terms, our urban sociologists would be judged equally similar to jobs involving political science research and geography research (with only a single match—the complementary term research—both jobs will have identical similarity scores).

We suggest that the lack of similarity measures for the primary occupational terms is a fundamental deficiency of keywording. This deficiency affects all occupational areas. Indeed, one suspects that it will be most serious for occupations that do not require extensive credentialing or maintain professional guilds that restrict mobility from one occupation to another.

Adequacy of the Occupational Unit Division

The keyword system divides the occupational world into 36 occupational units. A list of the occupational titles included in each unit is provided in reference notes to the interviewers. More than 600 titles are listed in all. In order to assess the usefulness of the keyword units in matching workers and jobs, we attempted to assign the 441 detailed occupational titles of the

1970 Census to these 36 keyword units.[8] Although the translation of the Census codes into the keyword occupational units is subject to some ambiguity, it did provide a rough guide to the way in which the keyword system divides up the national labor force and to how adequately it groups together occupations between which considerable occupational mobility occurs.

How the keyword system groups together occupations is important because each of the 36 occupational units uses different sets of complementary terms. Matches across different occupational units are less likely than those within the same unit because there is a smaller number of complementary terms in common between units.

Our exploratory analyses indicated that the occupational unit structure of the keyword system did not capture a major portion of the regularity in the naturally occurring patterns of labor mobility. Cross-tabulations of workers' 1965 by 1970 occupations using the 36 keyword units showed less orderliness[9] than similar tabulations using other classifications such as the U.S. Census, the SOC, or the DOT. While the comparison of such statistics for classification having different numbers of categories can be problematic in some cases, in this instance we observed that a highly detailed classification (36 occupational categories) yielded lower Chi-square values than more abbreviated classifications (i.e., U.S. Census: 12 categories; DOT: 9 categories; SOC: 21 categories). In this case[10] it is reasonable to conclude

[8]We have performed this analysis in two ways: (1) by using the occupational titles listed at the beginning of each keyword unit to assign the census titles to the keyword units and (2) by having two Employment Service staff members (experienced in keywording) assign each census title to a primary keyword unit. Here we report the results from the latter procedure, which provided a more complete accounting of the 441 census occupations. The substantive results derived from the two methods are sufficiently similar that our conclusions would not be altered by using one or the other method.

[9]As measured by the likelihood ratio Chi-square statistic.

[10]The analysis was done using data for workers who (1) had reported occupations in 1965 and 1970 and (2) had changed occupations (as reflected in the detailed census codes). The DOT and SOC major group codes were assigned to a detailed census category using information from a double-coded CPS sample. Each of the 441 census occupations were assigned to the 1 of 9 DOT and 21 SOC major groups into which the majority of workers in this occupation were classified in the double-coded survey. The basic mobility matrix showing occupational transfers between 1965 and 1970 was then reordered by using the major groups of the new classifications. Likelihood ratio Chi-square statistics were computed for the simple independence model (i.e., random transfer between occupational categories) and for several models of quasi-independence (e.g., where the diagonal of the moblity matrix was purged) by using techniques developed by Leo Goodman. For the simple independence models the results were the following:

Continued overleaf

that the 36 keyword occupational units do a relatively poor job of mirroring the naturally occurring patterns of labor mobility. While the crudeness of our techniques requires us to be cautious in interpreting small differences, the results obtained in our exploratory analyses were quite substantial. Comparatively speaking, the keyword occupational units do not adequately group together occupations between which there is a substantial movement of workers.

The shortcomings of the occupational unit arrangement are another serious conceptual problem of the keyword system.

Diversity of Computer Hardware and Languages

A major practical problem of the present keywording system arises from the manner in which the Employment Service has implemented its automation plan. At present, major sections of the keyword system and related aspects of the Employment Service's Automated Report System (ESARS) are programmed in a variety of machine-dependent[11] languages running on computers supplied by four vendors: IBM, Honeywell, Univac, and Burroughs. The states participating in the automated matching system choose their own computer vendors. The decision to use machine-dependent languages (PL-I and assembly languages)[12] rather than machine-independent languages (e.g., COBOL, FORTRAN) was made, in part, because greater efficiency of computer operation could be achieved by tailoring the matching programs so as to take advantage of differences in the architecture of the various computers.

A major advantage of machine-dependent languages is their efficiency. Writing programs in the actual language of the computer (i.e., the machine or assembly language) ordinarily results in considerable efficiencies at execution. The programs take less time to operate, although they take

Classification	Number of Major Categories	df	Likelihood Ratio Chi-Square
Keyword	36	1,156	148,584
SOC	21	324	311,720
Census, 1970	12	121	283,899
DOT	9	64	189,596

This comparison indicates that despite the large number of degrees of freedom associated with the keyword-classified mobility matrix, it captures less of the orderliness of occupational transfers than classifications using a smaller number of categories.

[11]Machine-dependent and machine-independent refer to the ability of a computer program to run on machines other than that on which it was written, e.g., machines manufactured by other companies. A machine-dependent program cannot be run on a different machine.

[12]Development was initially done on IBM hardware, and substantial parts of the matching system were programmed in PL-I and IBM 360/370 assembly languages. These programming languages are not available on the machines produced by other manufacturers.

longer to write. Programs written in high-level, machine-independent languages must first be translated (i.e., compiled) into machine language before they can be run. The results of such general translations are not usually as efficient as a comparable program for the same task that was originally programmed in machine language.

A major advantage of higher-level languages is their portability. A program written in standard COBOL or FORTRAN can be translated by each manufacturer's hardware, and it will perform identically on IBM, CDC, Honeywell, Burroughs, or other computers. A matching system programmed in a higher-level language could be implemented, without reprogramming, in the 50 states regardless of where the states buy their computers.[13]

The fact that the Employment Service automation plan has used nonstandard hardware and programs written in machine-dependent languages causes delay in the implementation of program changes and uncertainty as to whether the states are operating in the same manner and producing comparable, correct statistics on local operations. Furthermore, the possibility of implementing major changes in the keyword matching system is impeded by the fact that any change would have to be programmed four times—once in IBM assembly language and once in the assembly languages of Univac, Burroughs, and Honeywell.[14]

The use of both nonstandard hardware and machine-dependent computer languages in the design of the automated matching system is a major problem of the current system. It has the important practical consequence of inhibiting attempts to improve the system through redesign.

CONCLUSION

In the foregoing pages we have presented specific criticisms of the Employment Service's automated keyword matching program. The apparent inadequacies and inefficiencies of this system are not trivial, and the thought and effort required to remedy them are likely to be substantial. The foregoing catalogue of criticisms was prepared, however, with limited resources in a relatively short period of time. It would be ill-advised, in our opinion, to attempt to redesign the current system on the basis of this assessment alone.

It would be a wiser investment of resources first to enlist the assistance

[13]Assuming that the hardware configurations are adequate (e.g., if one needs to read tapes, each machine must have tape drives, etc.).

[14]One Employment Service official knowledgeable in this area estimated that any major change to the architecture of the system would take a minimum of 2 years to program and implement (at current levels of staffing).

of a planning group, composed of individuals knowledgeable in the areas of artificial intelligence, the management of complex data bases (e.g., in computerized bibliographic searching), heuristic search techniques and their applications, and other areas of computer science, together with individuals knowledgeable in labor economics, occupational analysis, and the functioning of the Employment Service. This group should be charged with two missions. First, it should be asked to provide advice concerning short-term modifications that could be made to the current automated matching system in order to improve its performance. In this regard, the group could assess system modifications that might be expeditiously designed and incorporated in actual state Employment Service operations in 1–2 years' time. Such short-term enhancements to the keyword system might include (1) installation of supplementary matching strategies, (2) recoding of machine-dependent sections of programs to machine-independent languages in order to ensure standardization of Employment Service operations and statistics and to facilitate updating of the system, (3) improvement of keyword matching strategies, and (4) elimination of operational inefficiencies (e.g., use of ordinary English keywords rather than numerical codes).[15] These initial enhancements could be overseen by the planning committee and implemented by the Employment Service Division of Automated Matching, with assistance from outside contractors if it were appropriate.

The initial stage of work by such a group would attempt to provide useful enhancements to the current system. This, however, should not be the final goal of the group's work. In a second stage of its work, the group would undertake a fundamental reassessment of automated matching. Our limited review suggests that a fundamental reconsideration of the architecture and logic of the keyword matching would be appropriate.

The anticipated costs of the Employment Service's automation plan—currently estimated at $250 million—justify a careful and thoughtful review of the current system. If redesign is warranted and basic changes in the architecture and functioning of the system are contemplated, then sufficient time should be allowed for careful implementation of the new system and detailed evaluation of its impact. If properly planned, phased implementation of a new system could be carried out in such a way as to facilitate careful experimentation to monitor the effectiveness of the new system. This undertaking could also be overseen by an advisory committee expert in relevant areas of science and technology.

[15]It should be clear that these are merely examples of questions that might be considered. Clearly, the usefulness of such a committee would depend upon its ability to dismiss any or all of these suggestions and to consider an entirely different set of enhancements if that seemed appropriate.