This article reviews concepts and measures of skill in the social sciences. Conceptual positions differ in the ultimate origin of bases for skill, the locus and nature of skill valuation mechanisms, the extent of social constructionist influences, and the dimensionality of skill. An emerging consensus posits two organizing dimensions: substantive complexity and autonomy control. The major research designs that include the study of skill are highly varied and complementary in their knowledge yield. The major measurement strategies have shifted in recent years from nonmeasures and indirect measures of skill to direct measures of two types: expert systems and self-report measures. The review considers in detail the major expert system, the Dictionary of Occupational Titles, *in terms of population coverage, aggregation bias, reliability, validity, and relationship to self-report measures.*

Skill

MEANINGS, METHODS, AND MEASURES

KENNETH I. SPENNER
Duke University

Over the past 20 years, a renaissance of research has used the concept of skill in a variety of contexts. These include studies of technological change, the labor process, overeducation-underemployment, job redesign, and comparable worth. This article reviews the state of the art in the measurement of job skills. What are the dominant approaches? How are they grounded in theory and concept? What do we know about the validity and reliability of different approaches? How much progress have we made in measurement in the past two decades?

To summarize my arguments: The major approaches include nonmeasures, indirect measures, and direct measures, all of which are used in aggregate and case studies. Measurement sources include external, expert judgments and self-reports of people in jobs. The linkage of measurement strategy to the conceptual-theoretical approach has become more problematic, but in a way that is "healthy" for cumulative science. We have learned some things about issues of validity and reliability in the measurement of job

WORK AND OCCUPATIONS, Vol. 17 No. 4, November 1990 399-421

skills, but we still have a great deal to learn. Key issues remain unresolved, such as dimensionality and agreement on standardized measures. In short, we have made progress over the past 20 years—relatively more in awareness of measurement limitations and sensitivity to pitfalls and relatively less in the development and diffusion of best practice measurement strategies and procedures.

CONCEPTUAL GROUNDINGS

Several distinctions help define the major conceptual approaches to skill. Most writers concur on the necessity of distinguishing the skills that people bring to jobs (i.e., talents, abilities, capacities, and so on) from the skills that jobs require (i.e., task demands, role requirements, and what positions demand, permit, and encourage). The former are often viewed as individual possessions, the latter as characteristic features of social roles. Further, both as individual possessions and as job demands, skills involve issues of (a) social valuation (What shall be rewarded?); (b) social definition, construction, and historical rootedness (What shall custom, language, technical requirements of tasks, and the contests and power of interested parties define as a skilled performance?); and (c) supply, demand, and the governance structures that rule transactions involving skills in jobs and people (What transactional logic, that is, control and efficiency, and system, that is, market, subcontracting, and internal hierarchy, matches people with jobs or translates specific technologies into a form of work organization and configuration of jobs with given skill demands?).

These dimensions and issues approximately define the space of different theoretical and conceptual approaches to skill in people or jobs. In this issue, Attewell and others provide comprehensive reviews of the different theoretical approaches.[1] I briefly define the positions here because design decisions and measurement strategies often follow from theoretical stance.

First, theoretical positions differ in the ultimate origin of bases for skill. Marxist perspectives, and more generally proponents of what Form (1987, p. 30) labeled the idealist tradition, locate "skill" in the extent to which work provides balanced fulfillment of human nature. Skilled work balances the physical and the mental, the conception and execution of task. Braverman (1974) provided the best example in his use of the classic craft-artisan model of the 19th century. More recent examples occur in many case studies in the deskilling literature (see the collections by Wood, 1982; Zimbalist, 1979). On the other hand, neo-Weberian and neoclassical economic approaches

situate the ultimate origins of skill in the nature of technology and the associated intrinsic nature of tasks (Bell, 1973; Kerr, Dunlop, Harbison, & Myers, 1964).

Second, theoretical positions differ in the locus and nature of mechanisms of skill valuation. Neoclassical demand-side perspectives, such as the theory of the firm (Henderson & Quandt, 1971), and supply-side perspectives, such as human capital theory (Becker, 1975), use market mechanisms and logics of efficiency and return on capital as the places and rules of valuation. From a measurement point of view, these perspectives tend to stress the "objective" nature of skills and to use market reflections as indicators of skill. On the other hand, neo-Marxist and most labor process perspectives point to logics of control, capital accumulation and exploitation, and valuational mechanisms, such as contests over work content and closed hiring, promotional, and evaluational systems. These views are more likely to see skill as subjective, constructed, and requiring measurement of that which is valued along with that which is not valued or devalued (Littler, 1982).

Third, related to the first and second points, theoretical positions vary in the extent to which skill is socially constructed and defined. The extreme constructionist position views all or most of what is "skilled" as subjective and constructed; accordingly, the measurement of skill attends to what interested parties want (i.e., managers, unions, professions, and individuals; see Burawoy, 1979; More, 1982; Stone, 1974), or what bureaucracies do (i.e., DiPrete's [1988] status redefinition hypothesis). A corollary question is posed: Who controls the social construction process? Neoclassical views would ignore the issues or say "no one"; some labor process positions would point to those in positions of power (Braverman, 1974). More recent labor process research has suggested that workers and managers engage in various forms of skill construction and definition (Wood, 1982, 1989). Also, Form (1980) suggested that social constructionist influences include the ways in which investigators err in romanticizing the skills demanded by jobs in the past century.

This third theoretical contrast is important for the measurement of skill because it defines the nature and amount of measurement error in empirical indicators, relative to an ideal of a "true" underlying objective level of skill. Neoclassical perspectives would suggest that the observed level is close to the unobserved, true level; the error is small and random. Social construction perspectives would posit a large difference between observed skill labels and measures and true underlying "skill," and would postulate nonrandom errors of measurement (i.e., over- or underestimation of true skill conditional on the winners of the skill construction contests). Alternately, if the constructionist

influences of contestants cancel out, there occurs the measurement paradox of true underlying skill equaling observed, measured skill levels.

Most observers would agree that skill labels and measures reflect at least some degree of ideological and constructionist influence, thus posing the critical question: How much and in what direction? I know of no study that carefully measures the degree and direction of bias, although some areas of research contain studies that show how alternate measurement strategies provide different indications of skill. For example, studies of job evaluation systems show how different "expert" raters, different factors or components of jobs, and different analytical procedures produce different judgments of job skill (for review, see England & Dunn, 1988). In another example, Myles (1988) showed how expert ratings of a job compare with individuals' ratings of their own job on the same skill measure (specific vocational preparation). For the Canadian labor force, expert ratings appeared to modestly overstate skill levels as compared with measures taken from individuals; both indicators lead to similar conclusions concerning compositional upgrading of the job structure. Hence there is some bias—but in which direction depends on the source that we treat as the true standard.

Finally, the various theoretical positions differ in the postulated dimensionality of skill. Even today, many analysts treat the issue as unimportant, the meaning of skill being obvious (but unstated) and the construct being unidimensional, for example, Flynn's (1988) review of nearly 200 case studies in economics. On the other hand, most recent sociological studies of skill have acknowledged the multidimensional nature of skill and measured it accordingly, although this trend is stronger in the United States compared with Europe and elsewhere (cf. Spenner, 1988a; Wood, 1989).

As hypothesis and pragmatic approach, consider a two-dimensional organizing scheme that I proposed elsewhere (Spenner, 1983, 1985). The thrust of empirical research suggests at least two dimensions of job skills: substantive complexity and autonomy-control. Each dimension has analytic and substantive subdimensions. Substantive complexity refers to the level, scope, and integration of mental, manipulative, and interpersonal tasks in a job. The subdimensions of mental, interpersonal, and manipulative refer to the classic functional foci of "data, people, and things" as dimensions of interface between a person and a task. The level, scope, and integration of these substantive dimensions capture important empirical variations in the chemistry of the dimensions as they are found in jobs. Most studies consider *levels* of one or more of the functional foci, taken separately or in combination.

Theoretical and empirical research also have suggested that autonomy-control is an important organizing dimension of jobs and skills. Autonomy-

control refers to the discretion or leeway available in a job to control the content, manner, and speed with which tasks are done. We must carefully distinguish autonomy-control from formal authority. Authority refers to between-role differences in legitimated power; autonomy-control refers to the discretion and leeway within a role. The two are positively correlated. Further, closeness of supervision may tightly constrain the autonomy-control of a job but is not synonymous with it; there exist other sources of constraint and freedom on jobs, such as the technology configuration or the integration and dependence of tasks in one job on those of another job.

Elsewhere, I offered detailed citation and consideration of the theoretical arguments and empirical evidence concerning these dimensions (see Spenner, 1983, 1985, 1988a, 1988b). However, several recent arguments and evidence warrant mention. First, empirical studies now show that the correlation between substantive complexity (measured as levels of the functional foci) and autonomy-control is in the range of $r = .5^0$ to $.7^0$ for fairly to highly heterogenous samples of workers or jobs. The level varies, conditional on the specific measures, scaling, sample and level of aggregation in occupations and whether the measures are based on expert systems, such as the *Dictionary of Occupational Titles* (DOT; U.S. Department of Labor, 1977), or are taken from people in jobs (Kalleberg & Leicht, 1986; Spenner, 1980; Vallas, 1988).

Second, some have suggested additional dimensions or subdimensions of skill or, alternately, different ways of organizing these two dimensions with other dimensions (e.g., see Miller, Treiman, Cain, & Roos, 1980). For example, some major systems in the private research sector, such as the Position Analysis Questionnaire (McCormick, Mecham, & Jeanneret, 1977) measure nearly 200 job features. But often these systems are more compilations or listings of every conceivable way in which jobs may vary; the linkage to a conceptualization of skill and to the number of distinct analytic dimensions remains problematic. Further, many of the arguments about additional dimensions of skill derive from large-scale factor analyses of vectors of job characteristics. Most of the additional dimensions refer to physical abilities, forms of dexterity, coordination, endurance, and tolerance that have biological or psychobiological content. To a large extent, the validity of these arguments is an empirical matter. My major points are that (a) dimensionalizations of skill should be conceptually and theoretically grounded —fully inductive orthogonal factor analyses are not sufficient; and (b) multidimensional conceptualizations of skill should *at least* include dimensions for substantive complexity and autonomy-control—more detailed dimen-

sionalization may be theoretically appropriate or show insightful nuances of empirical effects.

Finally, some have suggested that autonomy-control is not really a dimension of skill but is an additional nonskill, hierarchical dimension of jobs (Form, 1987). This argument often rests on a priori and unstated notions of "skill," or arguments that educational and training institutions neither provide nor recognize human capacities related to autonomy-control (i.e., independence, leadership, and problem solving), and hence we should not treat it as a dimension of skill. (I find more compelling an argument that schools and training institutions should recognize and address these components of human capability as they affect job performances, thus the problem is not with social science concepts but in training institutions.) For the most part, I believe the problem is one of labeling. Whether or not researchers, and more generally, the public, think of and evaluate autonomy-control as a dimension of skill, there is clear evidence that it is analytically separable from substantive complexity, that employers and employees evaluate and value it (Coser, 1975), that it affects self-judgments, personality, and health (Karasek, 1979; Kohn et al., 1983), and that it has distinct patterns of relationship to other social processes. For example, in the deskilling literature, there is far more consistent evidence of deskilling with respect to the effect of technological change on levels of autonomy-control than on substantive complexity. Autonomy-control deserves attention for a variety of reasons; whether or not we label it a "skill" matters less as long as we consider it in our studies.

DESIGN STRATEGIES

The various literatures that use concepts of skill as human capabilities or as job demands range from industrial and experimental psychology, education and sociology, industrial relations and management, to institutional micro- and macroeconomics. Hence research designs stem from a broad range of purposes and substantive questions. These include issues such as how technological change alters the quality and quantity of work, ergonomic and productivity considerations in job design and redesign, the measurement of comparable worth vis-à-vis comparable pay (including a growing number of private sector consultants in court litigation), optimal workplace organization and the labor process, human resource training models and planning, overeducation and underemployment, projections of future labor force configurations and needs, task structure and human motivation, and task orga-

nization in relation to stress, health, psychological disorder, and personality, to name only some areas of investigation.

The research designs used to study skill encompass similar variety and can be classified in a number of ways. The designs include experimental and nonexperimental designs. In the former, skill is sometimes manipulated as an independent variable among people or jobs (tasks) or is studied in relation to other manipulated independent variables. For example, one tradition investigates the causes and consequences of skill mismatch in novel games where task complexity, task autonomy, and subjects' perceptions of the relevant capabilities are all manipulated. O'Brien (1986) provided a comprehensive review of many of the experimental designs in fields of organizational behavior and industrial psychology. Variations on the experimental theme include laboratory settings with college students to the workplace with real employees. The obvious advantages of the experimental designs in this area are the classic ones: control over sources of variation and the ability to make causal inferences in the most rigorous sense; the major limitation involves external validity, or the extent to which the experimental results can be generalized to everyday situations. In general, little of the experimental research has made its way into sociological literatures, but as will be shown, some of it does inform key methodological issues in the study of skill, for example, in judging the extent to which people are accurate reporters of actual levels of task complexity and autonomy, or studies of how manipulated changes in task structure affect perceptions.

Research designs for studies of skill also differ in whether they are *primary*, gathering new data and making direct measurements of one sort or another, versus *secondary*, analyzing extant data, with their implicit or explicit measures of skill. For example, many of the *aggregate studies* of skill change and studies of comparable worth use secondary designs, aggregate referring to sample sizes greater than 1 and usually involving heterogeneity in people, occupations, industries, firms and so on. On the other hand, many *case studies* of a single or small set of workers, occupations, firms, and so on gather primary data on skill through surveys, interviews, observation, participant observation, and judgments and measurements from archival materials. The advantages of secondary and aggregate designs hinge on issues of population coverage and generalizability; the disadvantages center on a limited ability of the design to tap process, understand context, or average out key types of variability (Flynn, 1988). On the other hand, case study designs in the study of skill offer general advantage in their ability to see context, process, and dynamics of change; the overriding limitations

involve issues of generalizability, sample selections, and sufficient information to perform psychometric analyses or provide statistical controls.

Research designs differ in the number timing, and availability of repeated measures of skill, ranging from cross-sectional over units to longitudinal, panel, time-series, and repeated cross-sectional designs within and across units. The major issues in comparing research designs pertain to the quality of causal inferences (e.g., causes and consequences of skill transformations) and the quality of inferences about temporal changes. For example, regarding the latter, age, period, and cohort effects often become problematic because of the relevance of temporal shifts in skill (classic upgrading and deskilling argument) in context of well-known lifetime increase in skills both over the jobs in people's careers and as assessed in terms of developmental variations in human capabilities (Spenner, 1988b). In addition, in some areas, there are reasons to expect cohort effects relevant to skill transformations in, for example, the life cycles of products and technologies, legislation and court cases affecting job evaluations systems, or the rapid diffusion of a job redesign system or training program throughout an industry.

What an optimal design is clearly depends on the substantive questions, the state of knowledge in a subarea, issues of cost and, often, what is available. Across areas of research on skill, the dominant methodological wisdom suggests that no single type of research design is sufficient to understand a problem. For example, in studies of technology and skill transformation and in studies of comparable worth, both aggregate and case studies and primary and secondary designs significantly informed the current state of knowledge. In either of these areas, reliance on a single type of research design would lead to seriously distorted conclusions concerning the skill phenomenon in question.

In summary, research designs to study skill have characteristic strengths and weaknesses, some of which are accentuated or damped by the nature of the research question and the state of knowledge in an area. The major evaluational criteria for designs to study skill include validity issues pertaining to causal inferences, the validity of temporal and subgroup contrasts, external validity, and the quality of samplings of time and space. In some areas of research, the samplings of time and space are relatively strong, for example, measurements of education, training, and wage levels and practices across firms and over time. But in most areas, the samplings of time and space tend to be weak, particularly for time periods prior to World War II and certainly before the turn of the century because of both major upgrades in the census and other major data sources and the advent of systematic sampling

and data collection methodologies. Also, in some areas of research, the ability of a research design to distinguish age, period, and cohort effects is important.

MEASUREMENT ISSUES

An earlier classification of skill measurement strategies into nonmeasures, indirect measures, and direct measures continues to describe current practice (Spenner, 1983).

The nonmeasurement strategy simply asserts or assumes the empirical level of skill for a comparison. No empirical operations accompany the assertion. Frequently, occupational groupings, such as professional, managerial, blue-collar, or white-collar, are assumed to accurately capture skill levels. This measurement strategy remained popular into the 1980s, particularly in parts of economics and in some labor process research. For example, in Flynn's (1988) otherwise excellent book, which reviewed nearly 200 case studies of technology and skill in economics and presented a theoretical model of the skill-training life cycle, I am unable to locate either a specific definition of "skill" or concern with direct measurement of skill in the treatment of the case study literature. Wood's (1982) collection of insightful essays on the labor process—one of the major works on the topic, with a subtitle of "Skill, Deskilling and the Labour Process"—only ambiguously offered a specific definition of what the authors meant by skill. Further, I am unsure as to whether any of the investigators in their case studies used empirical operations or explicit protocols to measure skill—even in archival research—ones that could be replicated by other researchers. The nonmeasurement strategy involves substantial problems of validity. The referent dimension or dimensions are not clear in the cross section, to say nothing of how an unknown referent dimension(s) may have changed over time or whether the causes and consequences of subdimensions are isomorphic.

The indirect measurement strategy typically uses wage rates or education levels of people or occupational groups as an indication of overall skill level. This strategy, too, contains validity problems because the identity between indirect indicator and true skill level requires strong assumptions about (a) the dimensionality of skill; (b) other factors that generate variation in the indirect indicator but not necessarily in the true skill level; and (c) interactions between (a) and (b), that is, if skill is multidimensional, then do the factors in (b) operate constantly over the dimensions in (a)? (Field, 1980). The indirect measurement strategy remains popular in studies of skill in education and economics. In sociology, indirect measures of skill seem to

be on the decline, and when they are used, it is with a greater sensitivity to their limitations (Wallace & Kalleberg, 1982).

The most reliable and valid approach to the measurement of skill is direct measurement, "direct" meaning empirical operations and/or explicit protocols for the designation of skill levels. Applications include both quantitative survey-type research and qualitative, archival, and historical research. Sources of direct measures divide into two categories: expert ratings and self-report ratings. Expert includes outside observers, job analysts, and researchers, and more generally, imputations about skill from any source outside of the person or job that is the subject of study. Self-report includes persons in jobs and key informants reporting on the skill levels of people or jobs, if the informant is inside and part of the system under study.

The number of different direct measures is massive. I consider only major and selected applications, in particular, measures based on the *Dictionary of Occupational Titles (DOT)*. It is the most widely used expert system and serves as the major basis for construction of self-report indicators of skill in many studies. Further, in social science research, the *DOT* is by far the standard among expert systems. Other expert systems exist, for example, the Position Analysis Questionnaire, but none of them have extensive use in social science studies of skill. Most of them involve proprietary measures and validation and data bases based on samples of convenience. Hence I do not consider them here. Such systems may be an important social science resource in the future, but they have not had a major impact thus far.

EXTERNAL AND EXPERT SYSTEMS: THE *DICTIONARY OF OCCUPATIONAL TITLES*

Although designed as an aid to employment counseling, the *Dictionary of Occupational Titles* (DOT; U.S. Department of Labor, 1965, 1977) contains the most widely and frequently used ratings that serve as the basis for direct measures of skill. Miller et al. (1980), Spenner (1980, 1983, 1988a), and Cain and Treiman (1981) provided detailed review and critique of the *DOT* as a source of occupational data. The fourth edition of the *DOT* (1977) contains measures of over 40 variables for over 12,000 jobs, and the third edition (1965) contains nearly similar ratings for over 13,000 distinct jobs. The methodologies used in generating the ratings changed a bit between the two editions.

Table 1 lists the variables rated by employment analysts affiliated with the U.S. Department of Labor. The indicators include levels of involvement with data, people, and things, general educational development (GED: mathematical, language, and reasoning development required for an average perfor-

TABLE 1: Variables Measured in

Concept	*Variable*	*Scoring*
Worker functions	Complexity of function in relation to data	0-6
	Complexity of function in relation to people	0-8
	Complexity of function in relation to things	0-7
Training times	General educational development	1-6
	Specific vocational preparation	1-9
Aptitudes	Intelligence	1-4
	Verbal aptitude	1-5
	Numerical aptitude	1-5
	Spatial perception	1-5
	Form perception	1-5
	Clerical perception	1-5
	Motor coordination	1-5
	Finger dexterity	1-5
	Manual dexterity	1-5
	Eye-hand-foot coordination	1-5
	Color discrimination	1-5
Temperaments	Direction, control, and planning	0/1
	Feelings, ideas, or facts	0/1
	Influencing people	0/1
	Sensory or judgmental criteria	0/1
	Measurable or verifiable criteria	0/1
	Dealing with people	0/1
	Repetitive or continuous processes	0/1
	Performing under stress	0/1
	Set limits, tolerances, or standards	0/1
	Variety and change	0/1
Interests	Communication of data versus activities with things	*
	Scientific and technical activities versus business contact	*
	Abstract and creative versus routine concrete activities	*
	Activities involving processes, machines, or techniques versus social welfare	*
	Activities resulting in tangible, productive satisfaction versus prestige and esteem	*
Physical demands	Lifting, carrying, pulling, pushing	1-5
	Climbing, balancing	0/1
	Stooping, kneeling, crouching, crawling	0/1
	Reaching, handling, fingering, feeling	0/1
	Talking, hearing	0/1
	Seeing	0/1

(continued)

TABLE 1 Continued

Concept	*Variable*	*Scoring*
Working conditions	Outside working conditions	1-3
	Extreme cold	0/1
	Extreme heat	0/1
	Wet, humid	0/1
	Noise, vibration	0/1
	Hazardous conditions	0/1
	Fumes, odors, dust, gases, poor ventilation	0/1

SOURCE: Adapted from information in *Handbook for analyzing jobs* (U.S. Department of Labor, 1972).
* Scoring includes preference for one or the other of the pair or for neither.

mance), specific vocational preparation (SVP: total training time for an average performance at the job), 11 aptitudes, and a variety of temperaments, interests, and working conditions. The latter ratings are not personality temperaments and interests as such but are more like typical role demands of a job.

The main advantages of the *DOT* include its comprehensiveness, national scope, and relative ease of use in secondary data analysis. For example, numerous studies in different areas that investigate skill merge one or more *DOT* variables to individual record on the basis of detailed census or *DOT* occupational codes for respondents' jobs.[2] For purposes of secondary data analysis, the *DOT* is probably the best comprehensive system. The items most frequently used as measures of skill include the data, people, and things measures, GED, SVP, and selected items of the aptitude, temperament, interest, and working conditions variables (for review and examples, see Spenner, 1983). However, the *DOT* contains significant limitations as a source of occupational data, some of which are general to studies of skill while others are specific to certain applications. The most comprehensive study of the *DOT*'s methodological properties is by Miller et al. (1980). I summarize several of the major points and refer the reader to the original volume for detail. Anyone contemplating use of the *DOT* as a source of occupational data is advised to study Miller et al.'s original report carefully.

Population Coverage

The *DOT* is not a probability sample of jobs in the economy. Rather, it is an attempt to include all jobs based on work by analysts at field centers, who

analyze and visit convenience samples of firms in different industries. Some job ratings are based on hundreds of schedules; others are based on a few or one. In general, manufacturing jobs are oversampled having a much larger percentage of base titles compared with the corresponding fraction of the labor force. Conversely, professional, managerial, clerical, and service jobs are underdetailed, having fewer job categories than one would expect, based on share of the labor force.

Aggregation Bias

Any use of the *DOT* variables at the level of occupations involves some aggregation over person-instances of the job, for example, across firms, jobs within an occupation, and so on. Thus the *DOT* variables include measurement error to the extent that they do not capture these contexts and are relevant to the problem under study. At the level of individuals in jobs relative to the 12,000 *DOT* occupations, the slippage can only be indirectly assessed. A number of studies show the correlations between the same variable measured by self-report of job incumbents with the parallel indicator from the *DOT*, merged on the basis of occupation codes, are low to modest for indicators most closely related to substantive complexity and autonomy control, ranging from $r = .2^0$ to $.7^0$ (for examples, see Karasek, Schwartz, & Pieper, 1982; Kohn et al., 1983; Spenner 1980). But one cannot treat the zero-order correlations as simple validity coefficients because the measures are not parallel or equivalent (Lord & Novick, 1968). Indeed, individual-level, self-report measures may contain more measurement error relative to the occupation-level measures. Further, formal validity estimates would vary from unacceptable to quite reasonable, depending on the scaling, use of multiple indicators, and the psychometric-structural model employed. The slippage involved in using one occupation classification system versus another (e.g, *DOT* and 1970 census categories) appears to be minimal in terms of lost variance, if the number of categories is 200 to 300 or greater (Spenner, 1980; cf. Jones, 1980).

Reliability

Estimates of (interrater) reliability for *DOT* variables are in the range of .63 to .70 for commonly used variables and subscales, generally acceptable but not high values. However, there are several clear exceptions. For example, the ratings for levels of involvement with things and those for required strength are well below acceptable social science standards. On the other

hand, reliabilities for measures of data and people, GED subscales, and the SVP scale are in the range indicated earlier. Generally, the reliabilities are higher for manufacturing jobs compared with service jobs.

Validity

The issue of validity is complicated, ongoing, and well beyond the scope of this article in terms of formal review. Further, issues of validity depend highly on the context of the study, the nature of the problem under study, the way in which *DOT* indicators are taken to reflect constructs and so on. However, there are several recent studies and comparisons that lend some confidence to use of the *DOT* variables, but there are three specific ways in which there are potentially serious validity problems with the *DOT*.

Karasek and colleagues (Karasek, 1979; Karasek et al., 1982) conducted an extensive validation of *DOT* variables relative to individual, self-report measures for two constructs: job demands (similar to substantive complexity) and job decision latitude (similar to autonomy-control). The substantive study investigated job characteristics, strain, and health. Individual-level measures came from the Quality of Employment Surveys, national probability samples of the labor force taken in the 1970s. The *DOT* variables included indicators for data, people, and things, GED, SVP, and temperament variables related to direction-control, variety, change, and special instructions.

The central conclusions were quite encouraging. First, the *DOT*'s occupation-based estimates quite closely replicated the covariance patterns in the individual-based measures, but with some loss of statistical power. Second, the implied measurement reliabilities were acceptable in measurement and self-employment helped improve the quality of the estimates. Most important, the substantive conclusions that were reached using the two approaches were about the same. A number of other comparisons of occupation-level and individual-level indicators reached similar conclusions (Kohn, 1969; Myles, 1988; Spenner, 1980; Temme, 1975).

Parcel and Benefo (1987) estimated an extensive set of covariance structure models, which compared 42 indicators in the third and fourth editions of the *DOT*. Their models included constructs for unpleasant work, physical activities, physical dexterity, and complexity. Each construct had multiple indicators from both editions of the *DOT*, and the models permitted correlated measurement errors within and across editions. Several conclusions are important. First, the highest validities and lowest error estimates (random and nonrandom) involved the overall complexity construct. Second, for selected indicators, the models showed nonrandom correlated errors across editions,

confirming the suspicions of Miller et al. (1980) that a majority of titles in the fourth edition had their rating "confirmed" relative to titles in the third edition. This appears, at least in part, to have been a direct borrowing or relisting of the ratings verus independent confirmation. The estimates showed a very high degree of stability in construct covariances between editions (cross-time correlation coefficients above .9), even *after* adjustments for measurement errors.

A final type of comparison that has become available in recent years and informs that validity of skill measures taken from the *DOT* asks whether *DOT*-based conclusions are broadly replicated in specific problem areas by studies that use alternate methodologies. In general, the answer is yes. For substantive problems as diverse as job stress and coronary heart disease (Karasek, 1979), aggregate studies of skill transformations (Spenner, 1988b), job content, and pay inequity (England, Chassie, & McCormack, 1982; England & Dunn, 1988; McLaughlin, 1978), the same general conclusions are supported with alternate methodologies.

However, the *DOT* does have serious limitations. First, the *DOT* may contain gender biases, particularly in the title descriptions and perhaps the ratings and more so in the third edition for stereotypical female occupations (Miller et al., 1980). Studies that do gender comparisons are on more solid footing with the fourth edition, and all gender comparisons based on the *DOT* require cautious interpretation. This issues bears most directly on comparable pay-salary discrimination research.

Second, the fourth edition of the *DOT* contains many titles that were merely confirmed by analysts, using third edition ratings as a reference point. As a result, any uses of the *DOT* for temporal comparisons may be subject to substantial underestimates of change, although the direction and extent of the bias is unclear (see Cain & Treiman, 1981; Spenner, 1983).

Third, some of the *DOT* indicators are particularly suspect. The strength indicator needs overhaul (Miller et al., 1980). The GED indicator produces suspiciously high estimates of occupational upgrading—higher than any other indicator or method. There is some indication of a serious validity problem with GED types of indicator (overall "intelligence" or "reasoning" abilities required for occupational performance), as raters may as much be scoring social desirability and overall goodness of an occupation as any specific mental requirements (see Duncan, Featherman, & Duncan, 1972; Siegel, 1971; Spenner, 1983). Unfortunately, a number of major studies that reach upgrading conclusions—for example, the recent *WorkForce 2000* report (Johnston, 1987), which gained substantial national press attention—

rely heavily or exclusively on the GED subscales, and the conclusions likely contain serious error.

While the *DOT* contains limitations, it is the most frequently used and most accessible source of measures of skill. The verdict on quality is cautiously encouraging but with the significant exceptions noted earlier. The major report by the National Academy of Sciences (Miller et al., 1980), now over a decade old, called for major new government initiatives in expanding and updating the *DOT* and in providing a whole range of data and studies, cross-sectional and ongoing, which would provide a dramatic increase in the methodology of skill measurement and data. Federal budget allocations and priorities during the 1980s lost ground in this area rather than making progress.

SELF-REPORT MEASURES

The self-report measurement strategy takes direct measures of skill from people in jobs through questionnaire or interview procedures. The distinguishing measurement feature is that the respondent as subject and actor interprets and reports on his or her experiences. Distinguish here a more neutral reporting on one's own job from evaluations that are explicitly normative, such as questions about what people wish, desire, value, see as important and so on. The latter types of measures are clearly and highly problematic as measures of skill because they expressly include sources of variation other than, or in addition to, what people experience. This is not to prejudge the issue of whether people are accurate observers and reporters of the skill demands of their jobs.

There are many different examples and approaches to self-report measures, but several central tendencies are apparent.

First, many of the detailed questions have conceptual basis in the indicators from the *Dictionary of Occupational Titles* (for examples, see Kohn, 1969; Quinn, Mangione, & Seashore, 1975). That is, question protocols contain content highly comparable to the stimulus questions and scalings found in the *DOT*.

Second, many of the specific questions used as indicators of skill capture dimensions or subdimensions of substantive complexity and autonomy control (Albin, Hormozi, Mourgos, & Weinberg, 1984; Karasek, 1979). The specific indicators often center on core wordings of complexity, variety, challenge, creativity, learning new things, autonomy, repetition, discretion or working under one's own influence, and task scope. More detailed applications use measures that capture functional foci or axes of interaction

between worker and job, such as physical and intellectual demands (Hull, Friedman, & Rodgers, 1982) or conceptual-cognitive demands (Vallas, 1988). For example, Kalleberg and Leicht (1986) used items for training, overall "skill," learning new things, and variety as indicators of complexity; they used items for respondents determining speed, amount of freedom, and closeness of supervision as indicators for autonomy control. These scalings produced acceptable reliabilities and were supported in oblique factorizations, with the two dimensions correlated about .45.

Third, a number of studies simply measure skill with the key word "skill," leaving the definition to the respondent (Kalleberg & Leicht, 1986; Karasek, et al., 1982). These single items usually produce high-item-to-item correlations for scalings of substantive complexity type constructs and correlate well with other indicators of skill. Obviously, they are subject to greater error if used as a single indicator, both measurement error and confounding of subdimensions of skill. For heterogenous samples of adults, this evidence suggests that what respondents see and mean when they use the word "skill" is most closely related to substantive complexity of tasks but is also correlated with other subdimensions of skill.

Fourth, some studies tailor the specific measure of skill to the work task or occupation being studied. For example, Kelley (1988; also see Zicklin, 1987) measured the skill transformation in blue-collar machining occupations due to programmable automation (PA) by the extent to which workers do reprogramming on PA machinery and the extent to which workers are required to follow specific written work orders from management for PA machining. This approach possibly has higher validity within occupations or the domain of a case study, but it becomes unworkable and problematic (i.e., changing dimensions from occupation to occupation and comparable scaling across occupations) with more heterogeneity of workers and jobs.

Self-report measures engender a number of methodological issues. At a most general level, instrumentation effects might operate, for example, if earlier questions raise the salience of the skill level of one's work, creating biases in the direction of social desirability. I know of no empirical studies of the issue for skill measures. The safest strategies in instrument design include randomly scrambling scale items to eliminate order effects and strategic distancing and buffering with other items inside an instrumentor interview. Another general limitation involves the use of single indicators versus multiple-item scales, the latter in simple additive form or with more complicated multivariate scaling models. As is well established for many areas of research, single indicators are particularly subject to measurement

error, and the advantage rests with multiple indicators with direct specifications of reliability and validity.

Are people accurate reporters and observers of their jobs? Or are job reports contaminated with biases, distortion, and selective perception? Overall, empirical studies support respondents as being relatively accurate reporters (see Hackman & Lawler, 1971; Oldham & Hackman, 1981; Quinn, 1977; for review, see O'Brien, 1986; Spenner, 1988c; but cf. Lopata, Norr, Barnewolt, & Miller, 1985). There is some experimental evidence that several features of personality, ability, and social cues distort job perceptions in laboratory and nonlaboratory settings (Humphrey, 1985; Thomas & Griffin, 1983), but a major tradition of experimental studies shows that manipulations of objective job attributes generate corresponding changes in job perceptions (O'Brien, 1986). For central job attributes, including most of the basic components of skill, the correlations between objective and perceived measures are high in most studies.

A related, more complicated type of argument suggests an interactive effect of work values, job involvement, or other features of personality with the accuracy of perceptions or as determinants of peoples' perception of job characteristics (for reviews, consult O'Brien, 1986; Spenner, 1988c). Although the jury can never weigh in on such an open-ended set of possibilities, the scientific search for empirical support of this interaction hypothesis has been disappointing (Kohn et al., 1983; Spenner, 1988c). If there are shreds of evidence, they are at the margins of job stimuli and work conditions (Miller, 1988). That is, over the broad range of people and job situations, such interactive effects do not hold; at the margins, for example, the very extremes of high complexity and low autonomy, workers may compensate in their perceptions and attributions to create and "see" a more comfortable state of affairs, damping the reported levels relative to true levels. Accentuation effects may happen at other extremes of job stimuli. I know of no comprehensive tests of these ideas with reference to skill measurement. A special version of this hypothesis may operate for workers new to a job, who simply may not have the knowledge or have had sufficient time in the role to arrive at good reports.

In summary, self-report measures offer characteristic strengths and weaknesses for skill measurement, similar to other types of self-report measurement. There is no systematic evidence that people seriously distort reporting of their job characteristics; to the contrary, most of the evidence suggests that people, by and large, are fairly accurate perceivers and reporters of their immediate job situation (but not necessarily beyond—for example, characteristics of the organizational structure in which they work). There are special

problems with self-report measures, including wording, order and instrumentation effects, scaling, and approaches to measurement error. There is no consensus on a specific set of detailed questions to measure skill, in part because of differences in the conceptualization of skill. On the other hand, there is broad de facto consensus on the general pool of types of items to use, and these have begun to coalesce around indicators for substantive complexity and autonomy-control. Exact, word-for-word and procedural replication across studies is rare, while strong conceptual and measurement similarities are increasingly common in sociology.

CONCLUSIONS

Twenty years ago, the dominant methodological approach to the measurement of job skill was nonmeasurement or a unidimensional concept—we assumed that we knew what it was—or indirect measurement, usually with education or wages. Only an occasional study used direct measures of multiple dimensions of job skills. Braverman (1974) and others strongly critiqued the usual approach and exposed the limitations. On balance, we do much better now. Sociology, in particular, has made progress on these fronts: in direct measurement, in considering multiple dimensions, in treating ideological and social constructionist influences on skill as problematic, and in paying more attention to the ways in which theory, concept, and design dictate measurement and shape conclusions.

Nonetheless, the methodological agenda remains substantial. The issue of the dimensionality of job skills is taking shape. One strong contending hypothesis involves the correlated dual dimensions of substantive complexity and autonomy-control. But there are other possibilities, for example, in the empirical factorizations found in a number of studies (Miller et al., 1980; Parcel & Benefo, 1987) or in other theoretical dimensionalizations of jobs (Baron & Bielby, 1982; Karasek, 1979; Kohn et al., 1983). The 1990s should offer empirical basis for choosing an optimal dimensionalization for specific problem areas or perhaps across problem areas.

For the foreseeable future, measurement of job skills will benefit from dual reliance on expert systems and self-report measures. Each approach has characteristic strengths and weaknesses and needs for the future. The *DOT*-based system needs upgrade, expansion, and a full-scale embedded research program. These initiatives are conceptualized (Miller et al., 1980) but await funding, particularly from governmental sources. Other, newer methodologies for the expert-based direct measurement of skill requirements exist, but

these are in experimental or developmental stages.[3] The best prospect for the generation of new measures and methodological research rests in self-report measures. Consensus does not yet exist on replicated measures, but the 1980s showed the emergence of many commonalities in the pool of approaches and items. Here, too, the 1990s should produce considerable progress in reaching measurement consensus and in producing substantial new methodological research that permits the separation of fact from artifact.

NOTES

1. For other comprehensive reviews, consult Baldry (1988), England and Dunn (1988), Flynn (1988), Form (1987), O'Brien (1986), Rumberger (1983), Spenner (1988a), Wood (1982, 1989), and other articles in this issue.

2. Such matrices of job scores are available through the National Technical Information Service and the InterUniversity Consortium for Social and Political Research for both the third and fourth editions of the *DOT* and for 1960, 1970, and 1980 census occupation-industry categories (see Miller et al., 1980).

3. A number of economists and sociologists, including Paul Adler, Peter Cappelli, David Hachen, Mary Ellen Kelly, and Rui Teixeira, are actively working on new measures and methodologies for job skills.

REFERENCES

Albin, P. S., Hormozi, F. Z., Mourgos, S. L., & Weinberg, A. (1984). An information system approach to the analysis of job design. In S. Chang (Ed.), *Management and office information systems* (pp. 385-400). New York: Plenum.

Baldry, C. (1988). *Computers, jobs and skills: The industrial relations of technological change.* New York: Plenum.

Baron, J. N., & Bielby, W. T. (1982). Workers and machines: Dimensions and determinants of technical relations in the workplace. *American Sociological Review, 47,* 175-188.

Becker, G. (1975). *Human capital* (2nd ed.). New York: Columbia University Press.

Bell, D. (1973). *The coming of post-industrial society.* New York: Basic Books.

Braverman, H. (1974). *Labor and monopoly capital: The degradation of work in the twentieth century.* New York: Monthly Review Press.

Burawoy, M. (1979). *Manufacturing consent.* Chicago: University of Chicago Press.

Cain, P., & Treiman, D. J. (1981). The *Dictionary of Occupational Titles* as a source of occupational data. American Sociological Review, *46,* 253-278.

Coser, R. L. (1975). The complexity of roles as a seedbed of individual autonomy. In L. A. Coser (Ed.), *The idea of social structure: Papers in honor of Robert K. Merton* (pp. 236-263) New York: Harcourt.

DiPrete, T. A. (1988). The upgrading and downgrading of occupations: Status redefinition vs. deskilling as alternative theories of change. *Social Forces, 66,* 725-746.

Duncan, O.D., Featherman, D.L., & Duncan B. (1972). *Socioeconomic background and achievement*. New York: Seminar Press.

England, P., Chassie, M., & McCormack, L. (1982). Skill demands and earnings in female and male occupations. *Sociology and Social Research*, 66, 147-168.

England, P., & Dunn, D. (1988). Evaluating work and comparable worth. *Annual Review of Sociology, 14*, 227-248.

Field, A. J. (1980). Industrialization and skill intensity: The case of Massachusetts. *Journal of Human Resources, 15*, 149-175.

Form, W. (1980). Resolving ideological issues in the division of labor. In H. M. Blalock (Ed.), *Sociological theory and research: A critical appraisal* (pp. 140-155). New York: Free Press.

Form, W. (1987). On the degradation of skills. *Annual Review of Sociology, 13*, 29-47.

Flynn, P. M. (1988). *Facilitating technological change: The human resource challenge*. Cambridge, MA: Ballinger.

Hackman, F. R., & Lawler, E. E., III. (1971). Employee reactions to job characteristics. *Journal of Applied Psychology Monographs, 55*, 259-286.

Henderson, J. M., & Quandt, R. E. (1971). *Microeconomic theory* (2nd ed.). New York: McGraw-Hill.

Hull, F. M., Friedman, N. S., & Rodgers, T. F. (1982). The effect of technology on alienation from work. *Sociology of Work and Occupations*, 9, 31-57.

Humphrey, R. (1985). How work roles influence perception: Structural-cognitive processes and organizational behaviour. *American Sociological Review, 50*, 242-252.

Johnston, W. B. (1987). *Workforce 2000: Work and workers for the twenty-first century*. Indianapolis, IN: Hudson Institute.

Jones, F. F. (1980). Skill as a dimension of occupational classification. *Canadian Review of Sociology and Anthropology, 17*, 176-183.

Kalleberg, A. L., & Leicht, K. T. (1986). Jobs and skills: A multivariate structural approach. *Social Science Research, 15*, 269-296.

Karasek, R. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly, 24*, 285-308.

Karasek, R., Schwartz, J., & Pieper, C. (1982). *A job characteristics scoring system for occupational analysis*. Unpublished manuscript, Center for the Social Sciences, Columbia University, New York.

Kelley, M. R. (1988). *Beyond the deskilling debate: A statistical model for explaining the design of blue-collar jobs under programmable automation*. Unpublished manuscript, John F. Kennedy School of Government, Harvard University, Boston.

Kerr, C., Dunlop, J. T., Harbison, C., & Myers, C. A. (1964). *Industrialism and industrial man*. New York: Oxford University Press.

Kohn, M. L. (1969). *Class and conformity: A study in values*. Homewood, IL: Dorsey.

Kohn, M. L., & Schooler, C., with the collaboration of Miller, J., Miller, K. A., Schoenbach, C., & Schoenberg, R. (1983). *Work and personality: An inquiry into the impact of social stratification*. Norwood, NJ: Ablex.

Littler, C. R. (1982). *The development of the labour process in capitalist societies: A comparative study of the transformation of work organization in Britain, Japan and the U.S.A.* London: Heinemann.

Lopata, H. Z., Norr, K. F., Barnewolt, D., & Miller, C. A. (1985). Job complexity as perceived by workers and experts. *Work and Occupations, 12*, 395-415.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

McCormick, E. J., Mecham, R. C., & Jeanneret, P. R. (1977). *Technical manual for the Position Analysis Questionnaire*. Layafette, IN: Purdue University Press.

McLaughlin, S. (1978). Occupational sex assessment and the identification of males and female earnings inequality. *American Sociological Review, 43*, 901-921.

Miller, A., Treiman, D. J., Cain, P. S., & Roos, P. A. (1980). *Jobs, and occupations: A critical review of the* Dictionary of Occupational Titles. Washington, DC: National Academy Press.

Miller, J. (1988). Jobs and work. In N. J. Smelser (Ed.), *Handbook of Sociology* (pp. 327-360). Newbury Park, CA: Sage.

More, C. (1982). Skill and the survival of apprentice. In S. Wood (Ed.), *The degradation of work? Skill, deskilling and the labour process* (pp. 109-121). London: Hutchison.

Myles, J. (1988). The expanding middle: Some Canadian evidence on the deskilling debate. *Canadian Review of Sociology and Anthropology, 25*, 335-364.

O'Brien, G. E. (1986). *Psychology of work and unemployment*. New York: Wiley.

Oldham, G. R., & Hackman, J. R. (1981). Relationships between organizational structure and employee reactions: Comparing alternate frameworks. *Administrative Science Quarterly, 26*, 66-83.

Parcel, T. L., & Benefo, K. (1987). Temporal change in occupational differentiation. *Work and Occupations, 14*, 514-532.

Quinn, R. P. (1977). *Effectiveness in work roles: Employee responses to work environments*. Ann Arbor, MI: Institute for Social Research.

Quinn, R. P., Mangione, R. W., & Seashore, S. E. (1975). *1972-73 Quality of Employment Survey*. Ann Arbor, MI: Institute for Social Research.

Rumberger, R. W. (1983). *A conceptual framework for analyzing work skills* (Project Report No. 83-A8) Stanford, CA: Institute for Research on Educational Finance and Governance.

Siegel, P. M. (1971). *Prestige in the American occupational structure*. Unpublished doctoral dissertation, University of Chicago.

Spenner, K. I. (1980). Occupational characteristics and classification systems: New uses of the *Dictionary of Occupational Titles* in social research. *Sociological Methods and Research, 9*, 239-264.

Spenner, K. I. (1983). Deciphering Prometheus: Temporal change in the skill level of work. *American Sociological Review, 48*, 824-837.

Spenner, K. I. (1985). The upgrading and downgrading of occupations: Issues, evidence, and implications for education. *Review of Educational Research, 55*, 125-154.

Spenner, K. I. (1988a). Technological change, skill requirements and education: The case for uncertainty. In R. Cyert & D. Mowery (Eds.), *The impact of technological change on employment and economic growth* (pp. 131-184). Cambridge, MA: Ballinger.

Spenner, K. I. (1988b). Occupations, work settings and the course of adult development: Tracing the implications of select historical changes. In P. Baltes, D. Featherman, & R. Lerner (Eds.), *Life-span development and behavior* (Vol. 8, pp. 244-285). Hillsdale, NJ: Lawrence Erlbaum.

Spenner, K. I. (1988c). Social stratification, work and personality. *Annual Review of Sociology, 14*, 69-97.

Stone, K. (1974). The origins of job structures in the steel industry. *Review of Radical Political Economics, 6*, 61-97.

Temme, L. (1975). *Occupation: Meanings and measures*. Washington DC: Bureau of Social Science Research.

Thomas, J., & Griffin, R. (1983). The social information processing model of task design: A review of the literature. *Academy of Management Review, 8*, 672-682.

U.S. Department of Labor. (1965). *Dictionary of Occupational Titles* (3rd ed.). Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor. (1972). *Handbook for analyzing jobs.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor. (1977). *Dictionary of Occupational Titles* (4th ed.). Washington, DC: U.S. Government Printing Office.

Vallas, S. (1988). New technology, job content and worker alienation: A test of two rival perspectives. *Work and Occupations, 15*, 148-178.

Wallace, M., & Kalleberg, A. L. (1982). Industrial transformation and the decline of craft: The decomposition of skill in the printing industry, 1931-1978. *American Sociological Review, 47*, 307-324.

Wood, S. (1982). *The degradation of work? Skill, deskilling and the labour process.* London: Hutchinson.

Wood, S. (1989). *The transformation of work?* London: Unwin Hyman.

Zicklin, G. (1987). Numerical control machining and the issue of deskilling: An empirical view. *Work and Occupations, 14*, 452-466.

Zimbalist, A. (1979). *Case studies of the labor process.* New York: Monthly Review Press.