Thirty years ago, research on the *Dictionary of Occupational Titles* (DoT) transformed the study of job skills and related job characteristics (Spenner 1979; Cain and Treiman 1981). Criticism of the quality of DoT measures and an impasse over how to measure job skill requirements soon followed. Kenneth Spenner observed, ‘Our conceptualization and measurement of skill are poor. Unidimensional, undefined concepts, nonmeasures, and indirect measures of skill have not served us well’ (1983, 825). More recent reflections express some of the same concerns (Borghans, Green, and Mayhew 2001).

Recent debates on work, education, and inequality brought greater attention to job skill requirements than ever before, but there has been little sustained effort to develop better measures and there remains, in fact, little hard, representative data on what people actually do at work. Researchers have only a general sense of required skill levels, even less information on different skill dimensions and rates of change, and no well-defined intellectual framework.

This chapter presents a conceptual map of skill domains and a strategy for measuring them called ‘explicit scaling’. The goal is to demonstrate the feasibility and validity of a relatively comprehensive yet tractable set of measures. The approach is validated using
the first wave of the survey of skills, technology, and management practices (STAMP), a nationally representative, two-wave panel study.

This chapter reviews the skill debates, introduces explicit scaling, develops a conceptual framework for understanding job skill requirements, and presents evidence on the quality of the STAMP measures derived from them.

THEORIES OF THE CHANGING NATURE OF JOBS

Numerous debates hinge on claims regarding job skill levels and trends without articulating a clear vision for directly measuring them.

The long-running debate between deskilling and post-industrial theories is well known (Braverman 1974; Bell 1976; Attewell 1987; Form 1987), but methodological problems hindered progress.

The frequent reliance on case studies limited generalisability. There was no way to know if cases selected for study were representative and little consistency in measuring skill requirements across cases, so no convincing profile emerged of the overall job structure or its evolution. Skill changes due to shifts in occupation shares were beyond the scope of the method altogether (Spenner 1983).
In response, researchers used standardised measures derived from the DoT, often matched to representative labour force surveys (Spener 1979; Howell and Wolff 1991). Enthusiasm soon faded as DoT ratings were based on a convenience sample and not consistently updated since the 1960s. Consequently, this debate stalled by the late 1980s amid calls for better data (Cain and Treiman 1981; Attewell 1990; Spener 1990; Vallas 1990; US Department of Labor 1993, 20).

Quite separately, Bluestone and Harrison discovered the rise in earnings inequality in the US in the early 1980s. They argued that firms returned to profitability after the economic crises of the late 1970s by squeezing workers: eliminating institutional protections (unions, minimum wages), decreasing employment security (outsourcing, casualisation), and job downgrading (pay/benefit cuts, sweating labour) (Bluestone and Harrison 1982; Harrison and Bluestone 1988; Harrison 1994; see also Graham 1993; Green 2006). Many of these ideas, based on declining worker bargaining power, also remain debated because of measurement problems.

By contrast, Piore and Sabel (1984) argued that the same crises were forcing firms to abandon strategies based on low wages in favour of competing on quality, continuous innovation, and customisation. Flexible specialisation implied job enrichment, worker autonomy, employee involvement (EI) practices, and Japanese-style quality control techniques. Automation and computers reinforced both skill upgrading and decentralised authority in this view (Zuboff 1988).
Mutually reinforcing relationships between skill, technology, and EI practices contrast with the previous perspective, for which computers increase surveillance and narrow autonomy, with EI either a method for intensifying effort or merely a token gesture (Graham 1993; Vallas 2003).

Representative data addressing these debates is also scarce and key constructs, like self-directed teams, subject to vague or varying definitions. There is no real consensus regarding the prevalence of EI practices, much less the relationships between skills, technology, EI, and the other elements of the flexible specialisation and critical paradigms.

In response to Bluestone and Harrison’s structural account of inequality growth, mainstream labour economists focused on the spread of computers, arguing new technology raised the relative demand for skill and increased returns to education and other human capital (Katz and Murphy 1992; Autor, Katz, and Krueger 1998). This theory of skill-biased technological change is dominant among labour economists and favoured by some sociologists (e.g. Fernandez 2001), education researchers, policy analysts, and popular writers.

The exact causal argument relating computers to skills and wages remains unsettled. The wage premium observed for using a computer at work suggested the complexity of computer hardware and software required significant specific training investment (Krueger 1993; Dickerson and Green 2004). Others believe computers demand more
general cognitive skills because they increase the information content of work and decentralise problem solving (Autor, Levy, and Murnane 2002; Spitz-Oener 2006), even incorporating the flexible specialisation position on EI practices (Bresnahan, Brynjolfsson, and Hitt 2002).

One notable criticism argues that skill upgrading is the secular trend but did not accelerate when inequality grew, pointing to the need for long time series data (Mishel and Bernstein 1998).

There is a large research literature on skill-biased technological change (for reviews see Handel 2003a, 2004), but its original foundation was very rough proxies for job skill requirements, such as workers’ own educational attainment, broad occupation category, or DoT scores. Even today there is no consensus on a conceptual framework and operational measures of workplace cognitive skill requirements. Consistent time series data on skills necessary to test the acceleration hypothesis is even scarcer. Measures of technology are limited mostly to indicator variables for computer use or the value of an industry’s computer capital investment, rather than the complexity of computer skills used by workers.¹

Somewhat independently, official reports in several countries expressed doubts about the quality of national education systems, perceiving or fearing a growing mismatch between

¹ For exceptions using British data see Dickerson and Green (2004) and Borghans and ter Weel (2004).
the skills graduates possess and those required by the new economy.\footnote{See US Department of Labor, Secretary’s Commission on Achieving Necessary Skills 1991; HM Treasury 2006; see also Keep and Mayhew 1996; Payne 1999; Krahn and Lowe 1998; McIntosh and Steedman 2000; Haahr et al. 2004.} Accelerating demand for cognitive and teamwork skills is taken for granted, reflecting popular versions of the theories discussed above. Reports conclude with a long list of academic goals considered essential, but lack credible evidence on the share of jobs requiring different levels or kinds of academic skills.\footnote{For a review, including contrary evidence, see Handel (2003b). Education research includes Murnane and Levy (1996) and Rosenbaum and Binder (1997).}

Poverty researchers also saw the inner-city poor as suffering from a mismatch between their levels of education and the cognitive demands of growing industries and occupations (Wilson 1996; Holzer 1996; Moss and Tilly 2001).

Finally, diverse research streams focused on interpersonal job requirements as a new dimension of interest. The growing share of white-collar and service jobs shifts demand away from manual and toward interpersonal skills (Bell 1976; Reich 1991), as does the perceived growing importance of teams within occupations. This may be a problem for low-income minority workers with non-standard cultural capital, communication repertoires, and personal styles (Wilson 1996; Moss and Tilly 2001). Research on gender inequality focuses on emotional and caring labour.\footnote{See Hochschild 1983; Leidner 1993; Wharton 1999; Steinberg and Figart 1999; Glomb, Kammeyer-Mueller, and Rotundo 2004; England 2005.} Other studies show that jobs in fashionable boutiques, restaurants and bars, consulting, and investment banking require self-presentation work and ‘aesthetic labor’ (Payne, this volume).
By contrast, service jobs in fast food and call centres can be highly routinised, with very short-cycle, scripted, and ritualised interactions (Leidner 1993). Both the deskilling and emotional/caring labour perspectives agree that interactional demands can be a new source of job stress distinct from traditional forms of overwork. Nevertheless, as with cognitive skills, there has been limited effort to measure interpersonal job requirements.\(^5\)

**Conclusion**

The various debates over skills and the changing nature of work suggest that a new approach is needed to address the following substantive questions:

- How many and which jobs require what levels of various skills, computer competencies,\(^6\) and participation in EI practices? In brief, what is the skill profile of the evolving job structure?
- What are the functional and causal relationships between skill requirements, computer use, and EI?
- What are the effects of skills, computers, and EI on wages and other outcomes that define desirable or undesirable jobs (e.g. work intensity, promotion opportunities, layoffs, job satisfaction)?
- What are the trends in skill requirements, technology, and EI practices, their interrelationships, and their effects on the other outcomes mentioned above?

\(^5\) Exceptions include Steinberg and Figart (1999); Brotheridge and Lee (2003); Glomb et al. (2004); Hampson and Junor (2010).

\(^6\) For ease of exposition, ‘computer use’ sometimes covers the broader category ‘computer and other technology use’.
AN EXPLICIT SCALING APPROACH

This paper proposes a method called explicit scaling to address the impasse over the measurement of job skill requirements and related job characteristics that limits progress in the intellectual debates described above. Consider a question from the Quality of Employment Survey asking employees how much they agreed: ‘My job requires a high level of skill’ (Quinn and Staines 1979). Such measures are relatively common (Karasek 1979; Fields 2002, 72ff.).

Among their problems, respondents must decide for themselves what ‘skill’ means and judge their job’s overall level. The question and response options, ranging from ‘strongly agree’ to ‘strongly disagree’, contain no objective guidelines. The indefinite referents mean researchers can never be sure what the numerical scores mean in any concrete sense.

The question does not prevent responses based on relative standards rather than an absolute yardstick. Respondents are likely to rate their own job partly in comparison to jobs close to their own rather than using the entire range of jobs in the economy as their frame of reference due to its unfamiliarity. Indeed, job analysis has long wrestled with the problem of obtaining self-ratings based on an objective or common standard to ensure that measures mean the same thing across people and jobs.

---

7 ‘The European Union’s quinquennial European Survey of Working Conditions asks: ‘Does your main job involve complex tasks’ (yes/no). The Household, Income and Labour Dynamics survey, an annual Australian panel, asks level of agreement with the statement: ‘My job is complex and difficult.’
The DoT’s use of expert job analysts and workplace observations avoided many of the problems associated with self-reporting. Nevertheless, many DoT measures, such as a job’s relationship to data, people, and things, also did not correspond to obvious, unambiguous, or concrete concepts, and the different levels of some scales are not even clearly ordinal (Table 1, top panel) (US Department of Labor 1991, 3-1). Despite the value labels, the distinctions between different levels of data, people, and things appear rather arbitrary, to limit their interpretability. Further, though the use of job analyst site visits produced consistent scores across jobs, its costliness precluded replication so there is no time series for these scores.

***** TABLE 1 ABOUT HERE ***** [for file containing all tables, see Handel_Tables_LSaayman-edit.docx]

The official replacement for the DoT, the Occupational Information Network (O*NET) database, uses mostly survey self-reports from representative samples of employees instead of trained raters and non-probability sampling used for the DoT (Peterson et al. 1999, 2001; US Department of Labor 2005). O*NET is an impressive effort to develop comprehensive measures of occupational characteristics. However, most O*NET survey items are complex or multi-barrelled, abstract, and vague, producing interpretive difficulties (see Table 1, bottom panel). Many O*NET items have moderately strong predictive validity, but the use of rating scales with indefinite referents means one can never be quite sure what O*NET scores actually mean in terms of specific real-world tasks (Handel 2013a). If two occupations require different levels of estimating skills, with
one rated 3 and the other 4, one cannot really explain how they differ on this dimension beyond the difference in scores themselves because the scores have no inherent or external meaning. This is also true for standardised factor analytic scores constructed from rating scales and other unit-free measures (e.g. Miller et al. 1980, 176ff.; Spenner 1990, 403). Any effort to understand how much more skill might be involved in current jobs compared to the past will be limited insofar as it is not clear what is being quantified or if the quantification is not a real count or lacks cardinality.

Job analyses using arbitrary scales are also a problem for studies of mismatch between jobs and workers because it is even more difficult to find measures of workers’ estimating skills or factor scores on the same scale to compare directly to job requirements in order to determine match quality.

Explicit scaling seeks to avoid these problems with measures that are objective, concrete, correspond directly to the target of interest, and have absolute meaning. Survey questions refer to specific facts, events, and behaviours, rather than attitudes, evaluations, and holistic judgements, so they have greater external validity. Questions are general enough to encompass diverse work situations, but sufficiently concrete that they have stable meanings across respondents. Response options discriminate a wide range of levels to avoid floor and ceiling effects, and use natural units when possible. Rating scales, vague quantifiers, and factor scores, which have arbitrary metrics and lack specific or objective referents, are a last resort.
Explicit scales are intrinsically desirable because they have definite and easily interpretable meanings compared to measures that are more abstract or vague. Less room for subjective interpretation and self-enhancing biases hopefully also means less measurement error.

On the question of mismatch, for example, a job’s educational requirements can be compared easily with a person’s own level of educational attainment because both are measured using a common, natural unit. When measures of job and worker characteristics are in non-equivalent units, any analysis of match quality is more or less ad hoc.

Realistically, producing measures that are behaviourally concrete and meaningful in absolute or objective terms is often difficult. Questions that are very precise, referring to occupation-specific skills, for example, may achieve clarity and concreteness at the expense of relevance for most other jobs. Some constructs, such as work intensity, resist explicit scaling because they are complex and differ qualitatively across jobs. Concrete measures of work intensity for pilots, teachers, and assembly-line workers are as diverse as continuous flying hours, class size, and parts worked per hour. Even if a survey had space for a very long inventory of occupation-specific measures, there is no obvious way they could be mapped to a common scale for full-sample analyses due to their incommensurability. This applies to many constructs, such as job autonomy, task variety, and most kinds of occupation-specific skill and training (e.g. operating printing machinery, programming in Perl). When behavioural specificity is incompatible with generality, rating scales may be unavoidable.
DATA AND METHODS

The STAMP survey represents an explicit scaling approach to comprehensive measurement of job skill requirements and related job characteristics. STAMP is a nationally representative, two-wave panel study of employed adults first conducted between October 2004 and January 2006, using standard random-digit-dial telephone survey procedures (n=2,304). The survey contains approximately 166 unique items on job characteristics, summarised in Table 2. This represents the conceptual framework for measuring skills discussed in the next section.

***** TABLE 2 ABOUT HERE *****

A central goal of explicit scaling is to maximise the clarity and interpretability of items for both respondents and researchers, that is, strong face validity. The corpus of measures should also cover the different facets of constructs reasonably well and avoid floor or ceiling effects in measuring the levels of constructs, that is, content validity (Anastasi 1982). The survey’s success in these areas can be judged informally from the descriptions in the next section.

Criterion validity is assessed formally by comparing means across occupation and education groups and by correlating STAMP skill scores with standard measures of occupational standing and reward expected to be associated with the underlying
constructs of interest (occupational prestige, socioeconomic index [SEI] scores, mean occupational wages) (Anastasi 1982, 140ff.).

Construct validity is assessed using several methods of internal consistency (Cronbach’s α, principal components analysis, confirmatory factor analysis) (Anastasi 1982, 144ff.; Peterson et al. 1999). Because most STAMP items are dichotomous, nonlinear principal components analysis is used rather than standard PCA (Meulman et al. 2004) and categorical confirmatory factor analysis (CFA) rather than models for continuous measures.8

Because many STAMP items were designed to represent a hierarchy of complexity, Mokken scale analysis, a probabilistic form of Guttman scaling, is used to assess construct validity. Classical methods that assume items are parallel are not strictly appropriate for items representing different levels of difficulty (Sijtsma and Molenaar 2002, 55). Mokken scales use Loevinger’s coefficient of homogeneity at the item (H_i) and overall scale (H) levels to test whether departures from strictly hierarchical response patterns are sufficiently small that persons and items are consistently orderable. H_i is a measure of item discrimination, somewhat analogous to item-rest correlations in classical item analysis, and H is a measure of overall scalability.9 These are alternatives to classical measures of a scale’s unidimensionality.

---

8 Nonlinear PCA procedures were performed using the CatPCA command in SPSS. Categorical CFA analyses were conducted with Mplus.

9 H_i and H are one minus the ratio of observed to maximum possible Guttman errors, where the latter is the joint distribution of item frequencies based on the item marginals and assuming independence between items. Mokken recommended that all H_i exceed 0.30 and classified H values as indicating weak scalability (0.30 ≤ H < 0.40), medium scalability (0.40 ≤ H < 0.50), and strongly scalability (H ≥ 0.50). He also
Convergent and divergent construct validity is evaluated by examining whether skill measures correlate more strongly with job educational requirements than personal educational attainment, and by comparing correlations between STAMP scales and parallel and non-parallel measures from the DoT and O*NET.

Finally, measures corresponding more directly to some observable, real-world condition have external or ecological validity because they are more meaningful outside the context of a particular survey instrument than less direct or concrete measures. Verisimilitude, which distinguishes explicit scaling from prior approaches, is a key consideration in judging the quality of the STAMP measures.

**VALIDITY AND RELIABILITY OF STAMP MEASURES**

**Skills**

Although the concept of job skill requirements has proven quite debatable, it is defined here as technical task demands that are defined by employers as necessary for effective job performance. STAMP divides the skill domain into cognitive, interpersonal, and physical skills following the DoT’s data-people-things scheme, which has been validated extensively (Kohn and Schooler 1983; Peterson et al. 1999; Autor, Levy, and Murnane 2003; Autor and Handel 2009).

---

developed a test of the null hypotheses that $H_0=0$ and $H=0$ (Sijtsma and Molenaar 2002, 51ff.; van Schuur 2003, 149). All calculations were performed using the -msp- command in Stata.
Most research focuses on cognitive job skill requirements but faces significant measurement challenges, as discussed above. Covering this domain requires measuring transversal skills that apply to a wide range of jobs, both particular academic skills (e.g. math, reading, writing) and more general thinking and reasoning demands (e.g. problem solving). Required level of education captures aspects missed by the others and is an important overall measure.

More difficult to measure in a comparable fashion across jobs are the diverse occupation- and job-specific skills that are critical for most jobs (e.g. plumbing, computer programming). The only obvious solutions are required prior experience in related jobs and learning time for current job, which can be measured consistently across jobs and correspond closely to human capital concepts.

**Math**

Mathematics is relatively well structured into levels of complexity that can be represented by a series of items, from simple counting to the use of calculus and other higher math (Table 3). Respondents report whether they perform each task as a regular part of their jobs (1=yes). The questions are objective and behaviourally specific yet remain applicable to all occupations and have stable meanings across respondents. Unlike subjective rating scales, they can be related easily to specific aspects of math curricula
and to the personal skill levels of workers and students, which are advantages in investigating skills mismatch.

The top panel of Table 3 shows measures of construct validity. Cronbach’s α is reasonably large (0.81) and most correlations between individual items with a subscale excluding the item are above 0.50. The first principal component accounts for 63% of the total variance, well above the 30–40% cut-off commonly recommended for a strong dominant factor, and the item loadings are large.

***** TABLE 3 ABOUT HERE *****

When the items are fit to a Mokken scale, the values of Loevinger’s H for the full scale (0.83) and each of the items (>0.70) are all significantly different from zero and show very strong scalability. Indeed, about 91% of all respondents answered the math items in a strictly cumulative fashion, consistent with the Guttman model. This strongly suggests that the items represent a single hierarchy of skill.

The categorical confirmatory factor analysis is more consistent with a two-factor solution distinguishing basic and complex math (columns 4 and 5). The table shows fully standardised loadings and uses the root mean standard error of approximation (RMSEA) as the overall fit index. The fit is acceptable but the hierarchical quality of the items creates interdependencies among the items that makes CFA problematic, so these and similar results should be taken as suggestive.
When the sample is divided into five broad occupation groups, the proportion reporting more complex math use generally follows a pattern consistent with expectation (see Handel 2013b).10

When groups are defined by job education requirements (job education) and personal education attainment (own education), the math scales correlate significantly with both criteria but more strongly with job education (≥0.40) than own education (≥0.30), indicating both convergent and divergent validity (Table 3).

**Reading and writing**

Unfortunately, reading is not as easy to measure on a single difficulty continuum and/or to equate to specific curricular concepts or years of schooling. Unlike mathematics, there are no standard categories available to researchers or survey respondents that capture text complexity in a satisfying way. Educational psychology often uses some combination of average word length and average sentence length. Despite dissatisfactions, no alternative has achieved similar currency (US Department of Education, National Center for Education Statistics 2001). However, writing can be ranked clearly above reading because it requires skills needed to create as well as interpret text at a given level of complexity (e.g. logically organise and develop ideas) (ACT 2002, 19ff.).

---

10 The occupational groups are upper white-collar (managers, professionals, technical workers), lower white-collar (clerical, sales), upper blue-collar (craft, repair workers), lower blue-collar (operators, labourers), and service workers (health, food, personal, and protective service workers).
STAMP opted to use a hierarchy of items based on text length at the low end and text complexity in the middle and upper ranges to reflect the varying approaches within education research (Table 3, middle panel). It was hoped the items for the middle and upper levels would be well ordered, relatively unambiguous for respondents, and cover the range of text typically used on the job in terms of both qualitative variety and level of difficulty. Reading matter not meant to be read as continuous, running text, such as manuals and bills or invoices, were also part of this series but did not scale with the other items. Unfortunately, unexpected issues in question wording also reversed the expected relative frequencies of respondents saying they read work-related books compared to presumably easier reading material (e.g. articles in trade magazines or newspapers).

Nevertheless, the reading scale has reasonable construct validity. Cronbach’s $\alpha$ (0.80), variance explained in a nonlinear PCA (0.75), and Loevinger’s H (0.68), and most measures of item discrimination are high (Table 3, middle panel). Large majorities answered the reading items in a strictly cumulative fashion without Guttman errors, whether the scale is constructed assuming books are less difficult than both kinds of articles (71%) or more difficult (65%). Dependencies among the items required a trimmed CFA model, which also had good fit (RMSEA=0.04).

The reading items also have high criterion validity. The proportions of positive responses among the five broad occupational groups differed as expected (Handel 2013b). The
various scales of reading complexity correlate even more highly with job education (~0.60) and personal education (~0.50) than the math scales.\textsuperscript{11}

The writing items parallel those for reading except that writing professional articles and writing books were collapsed into a single question because few positive responses were anticipated for either task, which was the case.

While the first two columns suggest this scale does not perform as well as the reading scale, the Mokken scale results suggest it performs better (Table 3, bottom panel). Reflecting this fact, a very high proportion of respondents answered these items in strictly cumulative fashion (95%). The writing items discriminate among complexity levels better than the reading series; the proportion of positive responses drops much more sharply after the second level. Perhaps as a result, various CFA models did not fit well and are not reported.

The correlations between writing tasks and job and own education are similar to those for reading, indicating strong criterion validity (Table 3, bottom panel), and writing is more strongly associated with occupational group than reading (Handel 2013b).

\textit{Problem-solving}

\textsuperscript{11} The Mokken scale was constructed assuming that books represented a higher level of difficulty than news or journal articles. Cases with Guttman errors were corrected by arbitrarily assigning values equal to the highest level of self-reported reading task used on the job, with the further restriction that in order for cases with Guttman errors to be coded as reading books the respondents also had to report reading at least one kind of article for their job. The criterion correlations for this scale are somewhat lower than those for the other reading scales. However, if the sample is restricted to cases without Guttman errors the correlations with job education (0.72) and personal education (0.59) would be about 0.10 higher than those using the more conventional scales.
Discussions of the skills crisis often mention the need for problem-solving skills, but typically leave the concept undefined. Problem solving usually seems to refer to the application of general reasoning ability and common sense to novel or non-routine situations, but sometimes seems stretched to cover non-cognitive dimensions, like willingness to take initiative in novel situations, that is, conscientiousness. The concept of problem solving used here avoids motivational aspects, such as proactive work orientations.

STAMP operationalises problem solving as dealing with new or difficult situations that require thinking for a time about what to do next. Respondents reported how often they faced easy problems, defined as those that could be solved right away or after getting a little help from others, and hard problems, defined as requiring more time and a lot of work to solve. The response options for these items were vague quantifiers (never, rarely, sometimes, often), but respondents who dealt with hard problems also reported the number they face in an average week. As a group, the three items measure both frequency and complexity of problem solving.

The two items dealing with hard problems have high reliability (α=0.85) and account for a large proportion of the variance in a nonlinear PCA (0.85). An alternative scale using the two items with vague quantifiers performs a bit better by some measures and worse by others. Correlations with job education (~0.45) and personal education (~0.36) are

---

12 The number of hard problems was logged after adding a very small number to zero values; both variables were standardised before the scale was constructed.
also strong. In this case, the most behaviourally specific item – number of hard problems per week – may have been difficult for respondents to answer. Nevertheless, the latter is a useful check on the comparability of the vague quantifiers across respondents.

*Education, experience, and training*

Three summary measures eliminate most of any gaps in coverage left by the previous cognitive skill measures.

- Level of education needed by the average person to perform the respondent’s job (*job education*)
- Years of prior experience in related jobs needed by someone with that level of education (*related job experience*)
- Length of time needed to learn the job by someone with the required education and experience (*training time*)

All three use objective scales, operationalise important concepts in human capital and other theories, and have intuitive meanings for researchers, policy analysts, practitioners, and laypersons.

By design, job education and personal education are measured on the same scale so they can be compared directly. Mismatch is easily defined, avoiding the problems with the DoT’s general educational development (GED) variable, which required judgement or
strong estimation assumptions to impute an occupation’s required education level to assess match quality with individuals’ own education (Berg 1971; Halaby 1994).

A significant problem remaining is accounting for specific skills with measures that are general, parsimonious, and explicit despite the great number and qualitative diversity of occupation- and job-specific skills. Specific skills may appear incommensurate but the labour market rewards them mostly in a common currency and they are too large to be omitted, which would effectively render them the skill equivalent of dark matter.

STAMP uses required experience and training time to capture all non-academic, job-specific skills on an absolute scale. This appears to be the only way to measure the variety of job-specific skill demands on a common basis (Cully et al. 1999, 63).

Data on job education and training times in two waves of the Panel Study of Income Dynamics (1976, 1978) yield test–retest correlation for job stayers of 0.83 (n=1,356) and 0.60 (n=1,446), respectively. The corresponding correlations for workers who changed employers and three-digit occupation and industry were 0.51 (n=228) and 0.22 (n=257), respectively (Handel 2000, 187f.). This indicates a reasonably high level of consistency for job stayers and a much lower level for job changers, as expected.

---

13 Job stayers are respondents with at least two years’ job tenure in 1978 and whose reported three-digit occupation and industry were identical across the 1976 and 1978 waves.
Results in Table 3 and elsewhere also show that almost all specific skill measures are strongly associated with job education and exceed correlations with personal education, demonstrating criterion, convergent, and divergent validity.

**Interpersonal skills**

Job-related interpersonal skills are much less well theorised and well measured than cognitive skill requirements. Even at the most basic level, this domain is weakly conceptualised. This makes it difficult to assess content and construct validity, which require some level of consensus on the elements properly included in the domain.

The literature on interpersonal skills includes communication skills, courtesy and friendliness, service orientation, caring, empathy, counselling, selling skills, persuasion and negotiation, and, less commonly, assertiveness, aggressiveness, and even hostility, at least in adversarial dealings with organisational outsiders (e.g. police, bill collectors, lawyers, businessmen) (Hochschild 1983; US Department of Labor 1991; Hampson and Junor 2010). If dealing with co-workers as well as outsiders were considered, the list would also include leadership, cooperation, teamwork skills, and mentoring skills.

These elements seem qualitatively diverse, rather than different levels of a single trait. Many could be considered ancillary job characteristics, which, while often useful, are exercised at the discretion of the employee, rather than job or employer requirements.
Often it is not easy to separate interpersonal skills from more purely attitudinal and motivational aspects of work orientations (Moss and Tilly 2001).

On a practical level, survey questions produce very high rates of agreement and low variance if they do not distinguish relations with co-workers from dealings with organisational outsiders, such as customers and clients. Pre-tests show many people respond reflexively that working always requires a positive attitude, willingness to cooperate with others, and so on. Indeed, most managers feel pressure to engage in intensive impression management (Kanter 1977; Morrill 1995; Riesman, Glazer, and Denney 1950 on other-directedness).

To reduce yea-saying biases, STAMP asked a set of relatively specific questions that could apply to relations with organisational insiders or outsiders, as well as several general questions about extended interactions with outsiders only.

The first group included whether jobs require giving people information, counselling, dealing with tense or hostile people, teaching or training, interviewing, and giving formal presentations lasting at least 15 minutes. The questions dealing with outsiders asked if the jobs require contact with customers or the public, frequency of contacts lasting more than 15 minutes, and self-rated importance of such contact for their jobs. While these items attempt to be relatively concrete, substantial room for individual interpretation and yea-saying biases undoubtedly remain.
Caring labour and selling skills were omitted because pre-tests suggested they were relevant only for jobs that could be identified easily from the occupational title. The more generic negotiating, persuading, and influencing skills had the reverse problem. Rather than applying too narrowly, they seemed too open to broad agreement.

Despite these complexities, a single scale showed reasonable levels of overall consistency ($\alpha=0.72$, variance explained=0.68, RMSEA=0.04), and all individual items loaded on the latent construct (Table 4).

As expected, a number of items have relatively high levels of endorsement. However, the more specific items on working with the public yield a stronger gradient across occupations in the expected manner.\(^{14}\) Correlations with education are comparable to those for cognitive skills, though it is not clear that education is appropriate for establishing the criterion validity of interpersonal skills (Table 4).

**** TABLE 4 ABOUT HERE ****

*Physical demands*

Physical job requirements are bodily activities usually involving materials, tools, and equipment, corresponding to the final category of the DoT’s classification of job tasks as involving data, people, and things.

\(^{14}\) For example, on an 11-point scale for importance of working with customers, clients, and the public, the mean is 8.8 for upper white-collar workers, 8.3 for lower white-collar workers, 6.9 for service workers, 5.0 for upper blue-collar workers, and 4.2 for lower blue-collar workers.
Simple physical tasks include gross physical exertion (e.g. carrying heavy loads), elementary movements (e.g. sorting mail), use of simple tools or equipment, and machine tending. These are the kinds of tasks assumed to be common in deskilling theory and vanishing rapidly by post-industrial theory.

More complex physical tasks require more training, experience, and background knowledge regarding the properties of physical materials, mechanical processes, and natural laws. Deskilling theory predicts strong declines, but other theories give them less prominence. Unfortunately, specific craft skills are difficult to capture parsimoniously, so this domain is sparser than desirable.

Items on standing and lifting (simple physical tasks) follow principles of explicit scaling using objective yardsticks (bottom panel, Table 4). The question on hand-eye coordination and arm steadiness is the only measure of more skilled physical demands beyond whether required education is vocational. A final item uses a subjective rating scale for a more global report of physical demands to capture otherwise unmeasured aspects of this domain.

Scales using the four items have strong construct validity ($\alpha=0.79$, variance explained=0.80, RMSEA=0.00). The scales correlate negatively with job and personal education (-0.34), as expected. Blue-collar and service workers are much more likely to
report their jobs involve physical work. Skilled blue-collar workers are the most likely to say their work requires good eye–hand coordination or a steady hand (Handel 2013b).

**Technology**

Theories of skill change invariably implicate technology. Yet despite several attempts to develop standardised measures using both nominal classifications and ordinal scales, none has achieved widespread acceptance. Despite its centrality to research on skill and work roles, technology and its various dimensions remain weakly conceptualised.

The concept ‘computer literacy’, for example, has surprisingly little precise meaning despite its currency among both professionals and the general public. Test-makers have failed to develop reliable measures of computer task complexity or computer literacy (Statistics Canada 2005, 23; ACT 2002).

STAMP used 50 questions on computers, automation, and non-computer technology to capture the prevalence of common and important workplace technologies and the levels of skill they require.

**Information technology**

Four approaches were used to measure computer task complexity.
• A count of 18 specific software applications and another variable for frequency of computer use

• Five items for higher-level computer tasks (e.g. scientific/engineering calculations) and one for highly routine activity (data entry), to address both post-industrial and deskilling concerns

• Length of computer learning times, a natural metric for skill in human capital theory. Anticipating respondent difficulty recalling time required for learning general computer skills, STAMP asked respondents using job-specific software or learning new programmes in the previous three years how long they needed to learn the most complex such programme

• A measure of overall computer task complexity using a subjective rating scale varying from ‘very basic’ (0) to ‘very complex’ (10)

Finally, to measure possible computer skill deficits, STAMP asked respondents if they had all the computer skills they needed for their current job and if lack of computer skills had affected their chances of employment, promotion, or pay raise.

A scale combining (1) number of programmes and (4) self-rated complexity of computer tasks has reasonable construct validity (Cronbach’s α=0.71, PCA variance explained=0.77) (Table 5). The scale is more highly correlated with job educational requirements (0.43) than own education (0.31), indicating criterion validity. These calculations are also conservative because they use only the sub-sample of computer
users. Using the full sample of users and non-users increases $\alpha$ (0.89) and the correlations with job education (0.56) and own education (0.45) (not shown).

There are clear differences across occupational groups. Upper white-collar workers use six applications on average, while service workers use fewer than 1.5. Much larger proportions of white-collar workers use special software (~60%) compared to blue-collar and service workers (~25%). Upper white-collar workers were the most likely to perform higher-level tasks, such as using spreadsheet formulas, and to have learned new software in the previous three years (Handel 2013b).

By contrast, the two items on computer skill deficits did not scale with the other items or with one another ($\alpha$=0.30). Their correlations with both job and own education are less than 0.10 (not shown).

***** TABLE 5 ABOUT HERE *****

Non-computer technology

Technology associated with blue-collar jobs such as heavy machinery and industrial equipment has received even less attention than computers in employee surveys.

Respondents who use heavy machinery other than vehicles were asked 16 questions that addressed sociologists’ concerns with deskilling (e.g. machine tending, assembly line
work), traditional craft skills (e.g. machine set-up, maintenance, repair), and newer, high-technology skills (e.g. programmable automation).

Those using equipment introduced in the past three years also reported the time needed to learn the most complex such equipment, providing numerical estimates of both the skill requirements of new technology and the rate of technological change in blue-collar jobs.

All respondents rated the level of mechanical knowledge needed for their jobs (0–10 scale) and whether they required ‘a good knowledge of electronics, such as understanding transistors or circuits’ (1=yes).

Finally, to derive estimates of technological displacement, respondents who lost jobs in the previous three years were asked if this was because a machine or computer had replaced them.

Consistent with expectation, the groups most likely to use heavy machinery and industrial equipment are upper blue-collar (65%) and lower blue-collar (46%) workers, while the other three occupational groups have very low incidence rates (~10%). The ratings for required level of mechanical knowledge exhibit a similar pattern (Handel 2013b).

Management practices
Management practices are the third main leg of recent skill debates. This includes employee involvement practices, other aspects of autonomy and control in the workplace, and various aspects of job downgrading.

**Employee involvement**

Employee involvement is widely argued to be a significant driver of skill upgrading. Its major elements are relatively well defined: job rotation/cross-training, formal quality programmes, self-directed teams, and supportive training and compensation (e.g. pay for skill, gain-sharing, performance-based pay). STAMP covers all of these dimensions but this section focuses on teams.

To move away from very general questions regarding teams and quality programmes, which may elicit false positive responses (Cully et al. 1999, 42ff.), the questions on team membership include frequency of team meetings and 10 questions on specific areas of team authority and responsibilities corresponding to the concept of a self-directed team used in current research (Appelbaum et al. 2000).

The 10 team authority items have adequate construct validity using only the sub-sample of team members for conservative tests ($\alpha=0.69$, variance explained=0.67, RMSEA=0.09) (Table 5), but do not form a consistent hierarchy according to a Mokken analysis (not shown).
Somewhat unexpectedly, there is no strong association between the EI measures and occupational group (Handel 2013b). This may reflect relatively uniform diffusion or problems with the questions; current knowledge of EI practices is still too limited to draw strong conclusions. Despite trying to achieve precision, the EI items probably contain significant noise, consistent with other research (Gerhart et al. 2000).

*Autonomy, closeness of supervision, and authority*

Autonomy—control, authority, and closeness of supervision are related to but distinct from EI and skills.

Autonomy refers to discretion and the ability to work independently, as opposed to working under external control. Psychologists distinguish between employees who can set goals and basic rules, and those who can decide how to meet goals set by others within an existing structure. The former is control over a work situation, or strategic autonomy, and the latter control within a work situation, or operational autonomy (de Jonge 1995, 25f.). Braverman’s criticism (1974, 35ff.) of EI practices as a weak substitute for craft autonomy relies implicitly on a similar distinction.

Most measures focus on operational autonomy, such as control over hours of work and break times, the sequence and pacing of job tasks, closeness of supervision,
restrictiveness of rules, and task routinisation. In contrast, strategic autonomy includes authority to make decisions, direct subordinates, and allocate resources.\(^{15}\)

There is no objective standard for many of these concepts, and jobs are so diverse that measures of control over work methods are unlikely to be both concrete and widely applicable. Overly general items may be susceptible to self-enhancing biases (Wright 1997; Cully et al. 1999; Handel 2000; Gallie et al. 2004).

STAMP contains four items covering both operational and strategic autonomy: (1) freedom from prescriptive rules and supervisor’s instructions, (2) task repetitiveness, (3) closeness of supervision, and (4) policy-making authority.

Construct validity is mixed. Cronbach’s \(\alpha\) is quite low (0.49), albeit similar to some other autonomy scales (Kalleberg and Lincoln 1988, S136), but the variance explained by the first component of a nonlinear PCA is respectable (0.64) and the item loadings are high.

Criterion validity is also mixed. Upper white-collar workers are much more likely to participate in policy-making decisions (44%) than lower white-collar workers, upper blue-collar workers (~20%), and blue-collar and service workers (~15%). Occupational differences for the other three autonomy items are much narrower, but the additive scale using all four items discriminates among occupations reasonably well (Handel 2013b).

\(^{15}\) Examples are Kohn and Schooler 1983, 22ff; Wright 1985, 316; Kalleberg and Lincoln 1988; de Jonge 1995, 65; Cully et al. 1999, 141; Fields 2002.
The scale correlates relatively well with job education (0.43), own education (0.34), and management status (0.39), but not as well with other criteria, such as problem solving (0.25) and job satisfaction (0.29). The scale’s correlation with the employee involvement scale (0.15) is even lower, but may reflect the weaknesses of the EI scale. The evidence suggests it is more reasonable to consider these variables an index rather than a scale. Indeed, the definition of autonomy suggests a variety of concepts rather than a unidimensional trait.

*Job downgrading*

Though skill upgrading dominates the discourse on job quality, one cannot ignore the opposing position that explains inequality trends on the basis of worsening job quality (Bluestone and Harrison 1988). If skills are claimed to be more or less important than other dimensions of work, the latter require measurement.

A few standard surveys include questions on non-standard employment arrangements and recent job loss. STAMP also includes factual questions on downsizing, transfer of work outside the establishment (outsourcing), and pay and benefit cuts. Internal labour markets and work intensity are trickier to operationalise as explicit scales but survey questions on personal promotion prospects, organisational promotion policies, and workload, pace, and stress can help address questions regarding the importance of job downgrading versus upskilling.
Further validity measures

All scales were also converted to occupational means to compare them with measures from the DoT and O*NET, occupational prestige and SEI (Nakao and Treas 1994; Hauser and Warren 1997), and occupational wages (Current Population Survey).

Correlations between STAMP column variables and corresponding DoT and O*NET variables in Tables 6 and 7 measure the STAMP constructs’ convergent validity (in bold), while correlations with non-parallel DoT and O*NET variables measure their divergent validity. Correlations between STAMP scales and outcomes like occupational prestige, SEI, and wages measure criterion validity. The final columns of the tables present correlations between DoT and O*NET measures and occupational wages, which can be compared to the STAMP wage correlations in the bottom row of each panel.

STAMP measures have generally strong convergent validity. More than three-quarters of the correlations with parallel DoT and O*NET measures are between 0.65 and 0.92 (mean=0.72; median=0.75). Most are also larger than correlations with non-parallel measures in the same column, demonstrating divergent validity.

Most STAMP measures also have high predictive criterion validity. Cognitive scales correlate strongly with occupational prestige, SEI, and wages (Table 6, top panel). Correlations for autonomy, decision making, repetitiveness, and closeness of supervision
are more mixed but generally high (Table 6, bottom panel), as are correlations for interpersonal skills and physical job demands (Table 7).

**CONCLUSION**

Understanding job skill requirements and possible drivers such as new technology and management practices require a clear conceptual framework and direct measures. The various debates that hinge on assumptions regarding skill levels and trends will always remain constrained by speculation in the absence of surveys that ask employees about a wide range of specific skills in concrete detail.

Explicit scaling addresses this problem with objective, behaviourally concrete questions and response options with natural units, rather than rating scales and vague quantifiers. Items and scales using this approach have high validity and reliability and are easily interpreted. By measuring job requirements in ways that can be compared to worker traits, they also facilitate assessment of the congruence or mismatch between people and jobs.
References


