Methodological Issues Related to ORS Data Collection

Michael J. Handel
January 28, 2014

EXECUTIVE SUMMARY

The Social Security Administration (SSA) has contracted with the Bureau of Labor Statistics (BLS) to produce occupational data for use as the main source of information about job demands in determining eligibility for Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) disability benefits. SSA uses five steps of sequential evaluation to determine whether claimants are eligible for benefits under these programs. The first three steps are decided primarily based on medical factors, but at steps four and five, eligibility depends on both a claimant's limitations based on his or her impairments, and the physical and mental demands of work that jobs require. Claimants found not disabled have the right to appeal in administrative law hearings and the Federal court system. The proposed Occupational Requirements Survey (ORS), currently in development, is intended to be an important component in a new Occupational Information System (OIS) that will be the primary source of information on job demands used in SSA's disability programs and considered in judicial appeals. The DOT's suitability for this purpose, given its growing age, has been a source of increasing concern, and SSA has explored alternatives since the late 1990s. ORS is an interagency program in which BLS is supplying technical expertise and institutional capacity to assist SSA by collecting important information about occupational requirements for inclusion in its OIS.

SSA's disability programs are large public benefits programs and an important part of the U.S. social safety net. The programs have been the subject of significant attention from political actors inside and outside the Federal government and from claimants' legal advocates. The long search for a DOT replacement has also brought different technical expert groups into the discussion. Given SSA's disability programs' importance to beneficiaries and the general public, the political attention it has received, and the professional interest the proposed data collection program has attracted, it is clear that BLS is entering somewhat new territory. The ORS' primary purpose is to assist directly in the administration of a large public program, rather than to provide standard statistical information that policy-makers use for general guidance as needed. The large responsibility and scrutiny associated with this data program argue for careful efforts to ensure the data's quality.

After completing several exploratory research phases, BLS requested a methodological report that would provide advice on achieving high levels of reliability, validity, and accuracy given the

ORS' parameters. Specific concerns included the relative merits of different sources of information and modes of data collection. Review of program documents for this report also raised questions regarding the high levels of precision desired for variables dealing with frequency and duration of certain task requirements. The report was also asked to address concerns raised by the public in a recent official comment period when earlier program documents were released for public reactions. The comments relating to BLS' role on ORS clustered around concerns that in-person observations of job performance by trained field analysts, which were a prominent part of the DOT, would play a much smaller role compared to interviews, and that the occupational classification system would be more aggregated than the DOT. In addition to addressing these concerns, this report also provides substantial background on general issues of reliability and validity as they relate to the ORS project.

The central recommendation is that the ORS conduct its own systematic tests and validation studies in order to ensure that the final survey design aims for the most reliable, valid, and accurate measures that can be expected given the various possible alternatives. While there are many existing studies of the reliability and validity of job analysis methods and a selection of the most relevant are reviewed, a systematic review of the literature would not be very productive and would leave many issues unresolved given the differences between the research design of most academic studies and the nature of the ORS program. The ORS is much larger than most other studies, which imposes certain constraints, and the measures are quite distinctive, reflecting SSA's specific needs. Given the significance of the proposed data, a dedicated effort tailored to the specific circumstances of the ORS program is recommended.

To accomplish this task the ORS needs to develop a coherent master plan for conducting methodological research. A strategy document should give a clear, compact summary and explanation of the ORS' specific data needs and the alternative data sources and collection procedures under consideration, including procedures, such as job observations, that are only feasible on a limited basis for the final phase of data collection but that can serve as a benchmark or gold standard for assessing validity more widely in smaller-scale methodological studies. The master plan should also identify occupations, measures, and specific occupational requirements with particular importance due to the frequency with which they present themselves in Social Security disability claims so that program resources, such as job observations, finer occupational detail, and enhanced measurement, can be targeted to particularly critical areas. This planning document is necessary both as a roadmap for the work that a validation phase must accomplish and because detailed study of reliability, validity, and accuracy requires prior specification of intended procedures and purposes, and required levels of precision. Where gold standards are difficult to find or utilize, alternative methods of data collection can be compared to assess levels of agreement, their ability to discriminate among occupations that are known to differ based on prior knowledge, and their levels of inter-rater reliability.

Systematic tests should be designed by an experienced industrial psychologist that seeks to identify the most significant sources of variance among raters using the same methods across jobs. Sources of variance to be considered can include

a. difficulty of the item (e.g., duration)
b. characteristics of the respondent
c. length of interview
d. characteristics of the job and establishment rated (e.g., skill level, task homogeneity, industry, urban/rural location, organizational size)
e. field economist (e.g., experience, training, judgment, personal style)
f. field office (e.g., variation in training, informal practices across offices)

As no set of tests can investigate all of these sources of variance, prior project reports should be mined and field economists and supervisors debriefed to identify those that are most likely to be the largest contributors. Clearly, rater-related variance is the traditional focus of reliability studies and should be a principal concern of ORS reliability studies. In addition, given the known difficulty in capturing the duration for certain data elements and the social distance separating many firm officials in large organizations from physically demanding front-line jobs, special attention should be given to these items and situations.

In short, the ORS needs to design and conduct a well-planned series of tests and validation exercises to evaluate the efficacy of different approaches to data collection and optimize the final design of the program. This involves

1. identifying all gold standards that might serve as validation criteria for accuracy
2. defining a reasonable range of other methods to assess convergent validity when gold standards are unavailable
3. identifying significant sources of measurement error for assessing reliability, including duration items and respondents relatively distant from the front-line job
4. considering methods for distinguishing error variance and true heterogeneity within critical occupations, and measuring their absolute and relative sizes
5. relating standard measures of validity and reliability to rates of classification disagreement across key boundaries (e.g., sedentary vs. non-sedentary) to assess the practical implications of ratings disagreements that are observed

Finally, the ORS should consider mining existing large-sample databases (e.g., O*NET, National Compensation Survey, DOT) for the insights they can provide regarding the properties of measures and procedures similar to the ORS, and the likely magnitude and practical significance of within-occupation variation in job ratings where the information is available to further address concerns about the occupational classification's level of detail.