

Validating the Results of an Establishment Survey of Occupational Requirements Using Direct Job Observations

Scott Fricker

Kristen Monaco

Kristin Smyth

Bureau of Labor Statistics

QDET2

November 10, 2016

Acknowledgements and Disclaimer

The work presented in this talk reflects the combined efforts of many individuals at BLS, including members of the survey management team, the procedures and training groups, and our professional field staff.

The views expressed here are those of the authors and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics.



Overview

- *Occupational Requirements Survey (ORS)*
 - ▶ Purpose
 - ▶ Design/Methods
 - ▶ Quality-Assurance Activities
 - ▶ Challenges
- Job Observation Methods
- 2015 ORS Job Observation Pilot Study (JOPT)
- Conclusions and Next Steps

Occupational Requirements Survey (ORS)

- Conducted by the U.S. Bureau of Labor Statistics (BLS) for the Social Security Administration (SSA)
 - ▶ *Purpose:* Support SSA disability adjudication process
 - SSA must determine whether claimant can perform her/any work
 - Existing sources of job-requirement data are inadequate
 - ▶ *2012 – 2015:* ORS development and testing
 - ▶ *Sept. 2015 – Sept. 2016:* 1st ORS production year
- ORS collects information on:
 - ▶ Physical and mental requirements of job
 - ▶ Vocational preparation (training/experience) and environmental conditions



ORS Design/Methods

- ORS is an establishment survey covering businesses in 50 states and the District of Columbia
 - ▶ 2-stage stratification:
 - Establishments w/in industry
 - Jobs w/in sampled establishments (proportional to employment)
 - ▶ # of selected occupations per establishment: 4 - 20
- Collection
 - ▶ Respondents – typically HR staff, hiring officials
 - ▶ Mode – primarily PV, but also phone & email
 - ▶ Items – 70+ data elements
 - Presence: Yes/No
 - Duration: Hours or Percentages

Select ORS Physical-Demand Elements

■ Postural

- ▶ Crawling
- ▶ Crouching
- ▶ Kneeling
- ▶ Stooping

■ Reaching/Manipulation

- ▶ Reaching overhead
- ▶ Reaching at/below shoulder
- ▶ Fine and gross manipulation
- ▶ Keyboarding

■ Pushing/Pulling

- ▶ Hands/Arms
- ▶ Feet/Legs

■ Climbing

- ▶ Ramps/Stairs
- ▶ Ladders/Ropes/Scaffolds

■ Communicate Verbally

ORS Quality-Assurance Activities

- Robust, iterative development process
 - ▶ Close interagency work to ensure that the measured constructs met SSA program needs
 - ▶ Small-scale cognitive testing (2012 – 2014)
 - ▶ Medium-scale regional tests (2013 – 2015)
 - ▶ Large-scale field test (2014 – 2015)
- Regular debriefings of respondents and field staff
- Interviewer training and mentoring programs
- Data diagnostics (edits/review, validation analyses)
- Soliciting external expert and stakeholder input

Challenges

- Establishment respondents may vary in their knowledge of occupational requirements
 - ▶ Some evidence from ORS testing; stakeholder comments
 - ▶ Other occupational studies involve directly interviewing incumbents or observing them performing their job
- No good benchmark dataset
- Need sufficient data to produce reliable estimates, examine patterns
 - ▶ Relatively small test sample sizes
 - ▶ ORS questions and procedures evolved during testing
 - ▶ Building library of expected relationships and variations takes time



Job Observation Methods

■ Goals:

- ▶ Capture within-person variability in activities
- ▶ Capture variability between people within an occupation
- ▶ Be unobtrusive, cost-effective, and efficient

■ Advantages

- ▶ Eliminates respondent error
- ▶ Natural setting provides richer context

■ Disadvantages

- ▶ Observer bias
- ▶ Time consuming/costly/burdensome



2015 ORS Job Observation Pilot Test

- Test occurred June – September 2015
 - ▶ Sample: subset of establishments that participated in the 2014 - 2015 field test/dress rehearsal
 - 540 pre-selected occupations (no substitutions; respondent selected which worker in the occupation to observe)
 - Criteria: occupations common in SSA disability claims; geography, industry; establishment size; sufficient sample in field test data

Sampled Occupations	
• Nursing assistants	• Cashiers
• Cooks	• Retail sales
• Waitress/Waiter	• Receptionists/clerks
• Dishwashers	• Team assemblers
• Janitors and cleaners	• Childcare workers
• Maids/Housekeeping	• Laborers/Movers

2015 ORS Job Observation Pilot Test, cont.

■ Test Procedures

- ▶ Two experienced ORS interviewers simultaneously observed the same employee performing their job
 - Attempted to observe “typical” work day and schedule
 - Neither interviewer was involved in collecting data from the sampled establishment during original field test
 - Interviewers did not review field test results or discuss their observations/codes with the other interviewer
- ▶ Observed employee in person for one hour
- ▶ Collected presence/duration information for Physical Demand elements* using semi-structured form
- ▶ 1-day observer training occurred 1 week prior to test
 - Study purpose, methods, use of observation form, etc.
 - Self and group study, plus video-based calibration exercises

Observation Test Results

- Contact Rate – 75% (405/540)
- Cooperation Rate – 60% (244/405)

Occupation (n)	Coop. Rate	Occupation (n)	Coop. Rate
• Nursing Assistants (9)	31%	• Laborers/Movers (21)	64%
• Childcare Workers (6)	37%	• Waitress/Waiter (19)	66%
• Dishwashers (13)	52%	• Cashiers (22)	67%
• Retail Sales (17)	57%	• Receptionist/Clerk (23)	68%
• Cooks, restaurant (16)	59%	• Maids/Housekeepers (20)	71%
• Cooks, institution/cafeteria (19)	61%	• Janitors/Cleaners (25)	74%

ORS Observation Test – Measures of Agreement

- Coded observations for each Physical Demand into four duration categories
 - ▶ Not present or seldom (LT 2%)
 - ▶ Occasionally (2% – 33%)
 - ▶ Frequently (34% - 66%)
 - ▶ Constantly (GT 66%)
- Inter-observer agreement
 - ▶ At least 0.90 agreement for 90% of elements
 - ▶ Lowest agreement was 0.77 – 0.79 for three elements
 - ▶ Most disagreements were 1-step differences

ORS Observation Test – Measures of Agreement, cont.

- Compared observation-based duration estimates with those derived from the field test interviews
 - ▶ Selected the max value from the two observations*
 - ▶ Common agreement measures (e.g., Cohen's Kappa) can be negatively impacted when distributions are not uniform
 - ▶ ORS physical elements tend to be highly skewed - for many of the jobs selected, the elements either are not present (e.g., crouching) or they occur frequently (e.g., gross manipulation)
 - ▶ Therefore, we present an adjusted kappa statistic (PABAK)

Observed vs. Interview Data Results

- Level of agreement generally was very good
 - ▶ Average adjusted kappa value: 0.68 (“substantial”)
 - ▶ 6 of 18 elements (stooping, reaching at/below shoulder, communicating verbally, fine and gross manipulation, and pushing/pulling with hands/arms) had low – moderate agreement (0.31 – 0.44)
- ORS is particularly interested in instances where the interview data may be underestimating durations – impacts SSA decisions
 - ▶ Sign test analysis revealed that in 5 of 6 low-agreement elements, observation resulted in higher duration estimates
- Logistic regressions indicate that agreement varies by job type and size of establishment



Lessons Learned & Next Steps

■ Job observation in ORS

- ▶ Provides promising source of convergent validity, and for targeting areas where ORS data may be sub-optimal
- ▶ Address stakeholder concerns
- ▶ Could supplement ORS data for certain jobs or elements
- ▶ Helped identify areas for improved interviewer training
- ▶ Improved ORS interviewers' understanding of how jobs are performed, and resulting confidence in ORS data quality
- ▶ Difficult to capture duration for some data elements (when speed of job is rapid or when multiple elements are present at same time (grasping, reaching, lifting))
- ▶ 1-hour observation may not be sufficient to reliably capture low-frequency actions

Lessons Learned & Next Steps

- Second Job Observation Test Planned for 2017
 - ▶ Mirrors the design of the 2015 test, but data compared to ORS production data
 - ▶ Expanded number and type of occupations selected
 - ▶ Targeting data elements that have higher nonresponse in the production collection interview
 - ▶ Single observer only
 - ▶ Inclusion of selected mental-demand elements (e.g., type of contacts, decision-making)
 - ▶ Explore additional paradata from observation

Questions or Comments about ORS or the ORS Job Observation Test(s)?



Contact Information

Scott Fricker

Senior Research Psychologist

Office of Survey Methods Research

www.bls.gov/osmr

202-691-7390

Fricker.scott@bls.gov

