

U.S. DEPARTMENT OF LABOR
BUREAU OF LABOR STATISTICS

Occupational Requirements Survey Third Wave Testing Report

September 2022



The Occupational Requirements Survey (ORS) is an ongoing voluntary establishment survey conducted by the Bureau of Labor Statistics (BLS) under an agreement with the Social Security Administration (SSA). The survey provides information regarding physical demands; environmental conditions; education, training, and experience; as well as cognitive and mental requirements for occupations in the U.S. economy. BLS and SSA continue to work together to evaluate measurement objectives from published results and plan to incorporate changes with the third wave in fiscal year 2023 to produce data relevant to SSA's disability programs.¹

In fiscal year 2021, SSA proposed several changes to ORS measurements which required additional testing for the third wave. Exhibit 1 displays the eight occupational requirements tested; three are new and five are modifications to existing requirements.

Exhibit 1. Proposed third wave changes tested

Occupational requirement	Description	New or modified requirement
Cognitive and mental requirements		
Public work area	Measure whether work occurs in setting where the worker is exposed to the general public.	New
Verbal interactions: Internal	Measure maximum frequency of verbal interactions <u>within</u> a worker's organization and update the frequency scale.	Modified
Verbal interactions: External	Measure maximum frequency of verbal interactions with the general public and update the frequency scale.	Modified
Problem solving complexity	Measure problem-solving complexity as a replacement for existing problem-solving measurement and change the measurement scale.	Modified
Adaptability: Work schedule variability	Measure the incidence of whether the work schedule varies (different days or times or different number of hours per week).	New
Adaptability: Task changes	Measure the frequency workers change critical tasks.	New
Frequency of work being checked	Update the frequency scale.	Modified
Physical demands		
Fine manipulation (including keyboarding)	Include keyboarding in the measurement for fine manipulation.	Modified

Testing overview and design

During fiscal years 2021-2022, BLS tested these proposed changes in three consecutive phases: focus groups, a structured cognitive test, and a field test. The overall goal of testing was to determine the feasibility of producing new or modified occupational requirements before these measurements are introduced for wide-scale collection in the third wave. The first phase of testing, focus groups, worked with ORS economists to identify concepts or terms that are difficult to explain to respondents and where additional resources or definitions would be helpful. The second phase was a structured cognitive test to determine whether respondents were able to understand the concepts and definitions explained by field economists in order to collect data in an accurate and consistent manner. The final stage of testing, a field test, was a production simulation using a modified interview to inform final recommendations for third wave measurements. As part of the second and third phases of testing, field economists re-contacted participating establishments from the second wave of ORS collection.

BLS used this testing to identify critical issues with the proposed changes and potential clarifications of measurement objectives, as well as to refine future survey collection procedures and best practices. An overview of the methods used for these three phases are summarized below. The [Results](#) section describes the findings for each new or modified ORS measurement across the three testing phases. Each phase of testing was used to help design and inform subsequent work and recommendations.

Phase 1: Focus groups

Focus groups are a useful first step in a testing process because they allow researchers to collect rich, qualitative data on a specific topic. The group setting allows participants to hear each other's perspectives and build on one another's feedback, which is valuable because ORS collection involves complex measurements that are not easily conveyed in a single question. This experience is valuable for understanding where field economists notice difficulty in explaining a concept or where additional clarification would be beneficial to data collection.

The focus groups were moderated and documented by the Office of Survey Methods Research (OSMR). Focus group participants included field economists with varying levels of experience and procedures, training, and review staff. Prior to meeting, participants were provided with materials describing the proposed changes. These materials were based on the information provided by SSA in development of the measurements and similar in format to procedures provided in the ORS collection manual.²

During each focus group, a free-flowing discussion was encouraged. However, from time to time, the moderator also called on focus group members to provide input to make sure that the perspectives of all focus group members were heard. Each new or modified ORS measure was reviewed and discussed with focus group participants, including a description of each proposed measurement question with response options. The information from the focus groups was used to identify potential issues and concerns with the proposed changes, including changes that could be difficult to collect or that could have unintended consequences on the coding of related requirements.

Phase 2: Structured cognitive test

Cognitive testing techniques involve one-on-one individual interviews with respondents to gauge their reactions to the measurement questions, response options, and concepts. Using structured follow-up probes, interviews can uncover whether the question measures the intended construct and what sources of measurement error may be invoked by the question, including issues with comprehension or mapping answers onto the response categories. Since the proposed measurements reflected complex concepts and occupational behaviors, the test did not focus solely on understanding the measurement questions exactly as worded. The protocol also included structured follow-up probes about the respondent's understanding of each ORS measurement.

Experienced field economists collected data for the structured cognitive test. Before the test began, all field economists participated in a training session, which focused on the goals of the test, test logistics and materials, concepts and questions, collection and write up, and debriefing plans. Additionally, field economists practiced administering the questions and response options through mock interviews.

Field economists made attempts to contact the prior ORS respondent at each establishment. If the prior respondent was no longer available to participate, field economists attempted to interview a new respondent from the sampled establishment. The interviews were conducted remotely, using the telephone (94 percent) or Microsoft Teams (6 percent), and lasted between 15 to 30 minutes. The interviews consisted of a series of questions including both the new and revised measurements and structured respondent probes following each question. Field economists followed a script to ask each ORS measurement question and follow-up probes. Respondents were either emailed a showcard with

response options in advance of the interview, or field economists displayed the card during the interview if conducted via Microsoft Teams.

After each interview, field economists documented how well respondents understood each question and the accuracy of the responses, as well as any problems respondents had answering the questions (e.g., comprehension issues, response selection issues). After data collection was completed, a debriefing session was held to assess the questions and get additional field economist feedback on how the new or modified ORS measurements performed. The debriefing questions focused on getting field economists’ general impressions of the overall test and how easy or difficult it was for respondents to understand the concepts.

The sample for the structured cognitive test was designed to support completing up to 100 interviews covering jobs in ten detailed occupations.³ The ten specific detailed occupations were targeted because they were relevant for the questions collected and were representative of a wide range of occupations and industries in the national economy. Establishments were selected from those that provided information for jobs in the targeted occupations as part of ORS production samples from 2018 and 2019. Field economists collected information for a job in one occupation per interview. To reach this number of respondents, a starting sample of approximately 500 previous participating establishments was selected to ensure field economists would have enough sample to reach the target goal of 100 interviews. A total of 90 interviews were completed.

Table 1. Number of interviews completed in the structured cognitive test by occupation code and title

Occupation code	Occupation title	Number of interviews
11-1021	General and Operations Managers	11
13-2011	Accountants and Auditors	9
25-2021	Elementary School Teachers, Except Special Education	9
29-1141	Registered Nurses	7
35-3023	Fast Food and Counter Workers	9
37-2011	Janitors and Cleaners, Except Maids and Housekeeping Cleaners	9
41-2031	Retail Salespersons	7
43-4171	Receptionists and Information Clerks	10
49-9071	Maintenance and Repair Workers, General	8
53-7062	Laborers and Freight, Stock, Material Movers, Hand	11

Phase 3: Field test

A field test, also referred to as a dress rehearsal, was conducted as a final evaluation of the proposed measurements using typical ORS interview methods and procedures. A field test is a valuable final phase of survey testing that provides a last opportunity to identify any potential problems before new or modified survey questions are introduced into production.

Experienced field economists collected data for the field test. Before the test began, all field economists participated in a training session, which focused on the goals of the test, test logistics and materials, concepts and questions, collection and write up, and debriefing plans.

Like the structured cognitive test, field economists made attempts to contact the prior respondent at selected establishments regarding one to two designated jobs collected in the prior production sample. If the prior respondent was no longer available to participate, the field economists attempted to interview a new respondent from the selected establishment. Field economists indicated that most of the respondents they interviewed in the field test were the prior

ORS respondent. The field test interviews were conducted remotely, using the telephone (96 percent) or Microsoft Teams (4 percent), and lasted between 15 to 30 minutes. The field test consisted of a modified ORS interview focusing on final changes to the new measurements and included a subset of other existing measurements to provide context and improve question comprehension, including questions about gross manipulation, pushing/pulling with hands/arms, and at/below the shoulder level reaching. Because the field test aimed to assess how the new and revised ORS questions worked in a more traditional collection environment, field economists could either ask the questions as written or use conversational interview methods as they would in a production ORS interview. Field economists provided the response options for each ORS question verbally and/or using the showcard and were instructed to answer respondent questions and/or provide clarification based on the procedures.

After each interview, field economists documented how often they had to change or clarify each measurement, how well respondents understood each measurement, the accuracy and confidence of the responses, what information they had to clarify for respondents, and anything the respondent found confusing. After data collection was complete, a field economist debriefing session was also conducted to get field economists’ feedback and how easy or difficult it was for respondents to understand the concepts.

The sample for the field test was designed to collect up to 130 interviews covering jobs in thirteen detailed occupations. Establishments were selected from previous production samples with occupations that were relevant to the measurements being tested; ten of the occupations were the same ones used in the structured cognitive test and three new occupations were added as they were most relevant for the questions collected. Establishments were selected from those that provided information for jobs in the targeted occupations as part of two ORS production samples collected August 2019 to July 2021. Field economists collected information for jobs in up to two occupations per interview. A total of 96 interviews were completed and 138 jobs were collected; 54 were obtained from respondents with one occupation in the sample and 84 jobs were obtained from 42 respondents with two occupations in the sample.

Table 2. Number of total interviews completed in the field test by occupation code and occupation title

Occupation code	Occupation title	Number of interviews
11-1021	General and Operations Managers	13
13-2011	Accountants and Auditors	11
25-2021	Elementary School Teachers, Except Special Education	10
29-1141	Registered Nurses	10
31-1131	Nursing Assistants	10
33-3051	Police and Sheriff’s Patrol Officers	10
35-3023	Fast Food and Counter Workers	10
37-2011	Janitors and Cleaners, Except Maids and Housekeeping Cleaners	12
41-2031	Retail Salespersons	10
43-4171	Receptionists and Information Clerks	11
47-2061	Construction Laborers	10
49-9071	Maintenance and Repair Workers, General	11
53-7062	Laborers and Freight, Stock, Material Movers, Hand	10

Results

The results section will cover the main findings across each of the three testing phases for each proposed ORS measurement. Exhibit 2 shows the new or modified ORS measurements that were administered in each testing phase. Two of the new proposed measurements were dropped prior to the final field test.

Exhibit 2. New or modified ORS requirements administered in each test phase

Occupational Requirement	Focus groups	Structured cognitive test	Field test
Public Work Area	Yes	Yes	Yes
Verbal Interactions: Internal and External	Yes	Yes	Yes
Problem Solving Complexity	Yes	Yes	No
Adaptability: Work Schedule Variability	Yes	Yes	Yes
Adaptability: Task Changes	Yes	Yes	No
Frequency of Work Being Checked	Yes	Yes	Yes
Fine Manipulation (including Keyboarding)	Yes	Yes	Yes

Public work area

The public work area measurement was designed to collect whether workers are required to perform critical job tasks in a setting or environment where contact with the general public is likely to occur, and whether there is a likelihood that a member of the general public could approach or communicate in person with workers while doing so. Members of the general public could include any individual outside of the worker’s organization (e.g., customers and clients, routine visitors, vendors, couriers, delivery personnel, etc.). The measurement specified that settings in which barriers designed to prevent a member of the general public from approaching or communicating with workers should be excluded.

Barriers were defined as anything intended to restrict or prevent access to a worker or area where the worker performs critical tasks, including walls, doors, partitions, and designated areas (e.g., areas marked “private,” “do not enter,” or “restricted access”).

Findings

Feedback from the focus groups indicated that the concept of barriers and the types of jobs that would require them were unclear. Thus, additional guidance and examples of physical barriers intended to prevent contact, as well as examples of jobs that would require barriers, were incorporated into the documentation.

In the structured cognitive test, field economists indicated that most respondents understood the question well and provided accurate answers. However, minor issues occurred for some jobs, such as fast food and counter workers and teachers. These issues were related to what physical areas to include (drive-thru versus inside of store, inside or outside of classrooms) and what to include for people who do not work for the company (e.g., contractors, temporary workers, parents, or students). Respondents were also somewhat split on whether their answer included only in-person contacts, or other ways of contact (e.g., talking on the telephone). Modifications for the field test included adding the term “physically approach or communicate” into the question to highlight that it is collecting only in-person contacts. Additionally, “read if needed” clarification information was added to the public work area question to provide guidance regarding settings and examples of people who do not work for the same company and are therefore considered to be part of the general public.

The question performed well in the field test and field economists were confident that respondents understood the question and provided accurate responses. Field economists did report some issues in jobs where workers could potentially be approached by the general public, but would not typically interact with them (e.g., a highway worker). The concept of a barrier was also interpreted inconsistently across respondents and more broadly than intended.

Verbal interactions: Internal and external

The existing verbal interactions measurement was modified to distinguish between two types: internal and external. The first measurement, internal verbal interactions, was designed to measure the frequency of interactions with individuals within the worker’s organization. The second measurement, external verbal interactions, was designed to

collect the frequency of interactions with those outside of the worker's organization. The modification retained the definition and intent of the existing verbal interaction requirement, which captures how often workers must begin verbally interacting with others while performing critical job tasks and included in-person contact, interactions by telephone or videoconferencing technologies, or any other real-time interaction. Additionally, the existing frequency scale was modified to include five options: every few minutes, less than every few minutes but at least once per hour, less than once per hour but at least once per day, less than once per day, and never.

Key differences between the two measurements were:

- *Internal verbal interactions* are limited to any individual within the workers' organization including co-workers and supervisors, regardless of whether the individual is familiar or unfamiliar to the worker.
- *External verbal interactions* include any individual outside of the workers' organization, such as customers and clients, regardless of whether the individual is familiar or unfamiliar to the worker. Examples of external contacts include routine visitors, vendors, couriers, delivery personnel, or frequent clients of the business which are all considered members of the general public.

Findings

Feedback from the focus groups indicated that it was unclear whether the questions on verbal interactions were designed to measure the *maximum* or *typical* frequency of verbal interactions, whether group interactions would count as a single interaction or a series of multiple interactions, and who would be considered internal versus external contacts. Additional clarification on these issues was incorporated into the documentation, including guidance to capture the maximum frequency of verbal interactions during a typical workday and the frequency of new interactions within group meetings, as well as examples of persons who are considered internal and external to an organization.

In the structured cognitive test, field economists indicated that most respondents understood both questions well and provided accurate answers. However, some additional time was required to educate respondents on the difference between internal and external interactions. Some respondents expressed confusion over the wordiness of the questions and response options and what types of verbal interactions were considered work-related. Finally, some respondents indicated they only included in-person interactions in their responses. For the field test, the wording used for the frequency scale options was shortened and optional prompts were added to clarify the types of verbal interactions to include or exclude.

During the field test, field economists indicated that they were confident that respondents understood the questions on internal and external verbal interactions and provided accurate responses. The shortened frequency response options and prompts clarifying the types of verbal interactions to include improved respondent comprehension of these questions in the field test.

Problem solving complexity

The existing problem solving measurement was designed to capture the frequency workers are faced with a problem which requires them to weigh alternatives when thinking what to do next and make a decision. Complexity in the existing measurement was defined as problems taking more than five minutes to determine a solution and only captured moderate to complex problem solving.

The modified problem solving complexity measurement retained the existing problem solving definition, however, the measurement scale was changed to assess the maximum level of complexity of problems workers are required to handle when performing their critical job tasks across three levels: low, moderate, or high. It was also assumed some level of problem solving exists for all jobs.

Each level of problem solving complexity encompassed multiple dimensions including the level of complexity of the problem itself, the information required to understand the problem, and the uniqueness and possible outcomes of solutions.⁴ If the complexity for any one dimension was higher than other dimensions for critical tasks, the highest level required should be captured.

Findings

Feedback from the focus groups indicated the multidimensional nature of problem solving made the measurement both difficult to understand conceptually and subsequently explain to respondents. Focus group participants also noted that distinguishing between three levels of problem-solving, especially understanding the distinction of moderate compared to low or high problem solving complexity would be challenging and lead to inconsistent classification. Thus, a respondent showcard with a table outlining key words for each dimension and level was created to be used during structured cognitive testing. Additional guidance was also provided for the structured cognitive test clarifying that the maximum level of any one dimension should be used to classify workers when a job falls into different levels across dimensions.

In the structured cognitive test, field economists indicated this question was one of the most problematic, and a significant number of respondents did not understand the question or provide consistent responses. Analysis of the responses across detailed occupations suggested a potential bias to select an average level of complexity for many jobs, and thus “Moderate” was the most frequent response option chosen. Respondents struggled to select the maximum level and tended to default to the typical level of problem solving complexity. Further, respondents noted the distinction between the moderate and complex levels was nuanced and seemed to overlap. Although some respondents indicated the showcard was helpful, many noted that the showcard options needed to be further simplified and required considering too much information together and therefore, they were not able to consider all dimensions at once to arrive at a single answer.

As a result of focus group feedback and structured cognitive testing, a reduction of response levels from three to two and simplified response option definitions were recommended. However, because this measurement could not meet the needs of SSA without three levels, this question was dropped and not included in the field test.

Adaptability: Work schedule variability

The intent of work schedule variability was to measure whether a job’s work schedule changes from week to week. A work schedule was defined as the number of hours, time of day, and days worked by the employee, within the work week set by the employer. Jobs with a set, unchanging schedule one week to the next were not considered to have work schedule variability. Jobs would not need weekly schedule changes to be considered to have work schedule variability.

Findings

Feedback from the focus groups indicated that this measurement would be relatively straightforward to collect. Focus group participants did note that it was unclear how a worker’s choice to have flexibility in scheduling affects work schedule variability. Additional guidance was provided that clarified schedule changes under the worker’s control should be excluded when determining work schedule variability.

In the structured cognitive test, field economists indicated that the majority of respondents understood this question well and provided accurate answers. However, testing did reveal that respondents did not consistently:

- Exclude changes under the worker’s control
- Include times when workers may be asked to report early or may be dismissed early
- Include seasonal changes

- Consider a work schedule based on the time period intended (weekly).

Additional “read if needed” prompts were added to the work schedule variability measurement to clarify these situations.

The measurement performed well in the field test with field economists indicating that almost all respondents understood the question well and provided accurate responses. Field economists did note that additional information was needed on how to collect salaried jobs who generally have flexibility over their start and stop times but may need to work until the job gets done during busy periods.

Adaptability: Task changes

Adaptability – task changes was designed to measure the maximum frequency of critical task changes occurring during the course of a typical workday. The measurement defined a critical task change as a change requiring the worker to stop performing one critical task in order to focus on a different critical task. The measurement specified that changes between sub-tasks, defined as a sequence of tasks required to perform the critical task, be excluded. Changes between critical tasks could include performing the same critical task more than once per day at different times as long a different intervening critical task occurred. The frequency scale included five options: every few minutes, less than every few minutes but at least once per hour, less than once per hour but at least once per day, less than once per day, and never.

Findings

Feedback from the focus groups indicated capturing the frequency of critical task changes could be difficult since differentiating between critical tasks and subtasks is complex and likely not able to be done consistently, especially among different detailed occupations. Focus group participants also indicated that typical ORS respondents (i.e., human resources personnel) may not have the in-depth knowledge to be able to accurately identify the frequency in which critical task changes occur.

In the structured cognitive test, field economists indicated this question was one of the most problematic tested, yielding issues with reporting consistency across jobs. The issues were due to inconsistent understanding of tasks at the detailed level needed to identify changes between critical tasks. Several respondents expressed difficulty arriving at an answer. This question had the highest proportion of non-substantive responses (e.g., ‘don’t know’) among all questions tested. Respondents reported difficulty defining what it meant to switch from one task to another task. There was no consistent understanding between respondents and switching was often not based on the type of changes intended to be measured:

- Switching locations but doing the same type of work was considered switching.
- Respondents often included ‘sub-tasks’ when thinking about switching.
- Interacting with different people but doing the same type of work was considered switching.

Additionally, field economists often had difficulty due to respondents’ confusion. While field economists can explain the concept of critical tasks to respondents, they ultimately rely on respondents to identify the critical tasks of jobs at the establishment. This question required additional consideration among and between critical tasks that currently cannot be consistently measured. Using lawn care as an example, respondents are able to report whether fine manipulation is required, or interaction with other people is required in the performance of lawn care. However, respondents are inconsistent in whether the components of lawn care would be separate critical tasks, and further, what would be a task vs. sub-task.

Findings from the structured cognitive test indicated that this measurement cannot be consistently or accurately measured. Issues with this measurement were due to difficulties with the source construct (critical tasks) and cannot be

addressed with question wording changes. As a result, the decision was made to drop the measurement from further testing, and it was not included in the field test.

Frequency of work being checked

The frequency of work being checked was designed to measure how often during a typical workday that a supervisor or lead worker routinely reviews and assesses the performance of workers in an occupation. The measure used the new frequency scale to be consistent with other cognitive and mental requirements measuring frequency. The frequency scale included five options: every few minutes, less than every few minutes but at least once per hour, less than once per hour but at least once per day, less than once per day, and never.

Findings

Few issues were identified with this measurement across all testing phases. The frequency response descriptions were shortened as a result of feedback from the structured cognitive test that the questions and response options were too wordy.

Fine manipulation including keyboarding

Fine manipulation was designed to measure workers' duration of touching, picking, pinching, or otherwise working primarily with fingers rather than with the whole hand or arm, as in gross manipulation. This modified measurement included the time spent keyboarding as fine manipulation rather than measuring keyboarding separately. The measurement also to captured whether one or both hands are required; however, both hands would not be assumed when keyboarding as it may not require the use of both hands (i.e., unilateral or bilateral) simultaneously.

Findings

Focus group participants indicated that clarification would be needed when determining unilateral and bilateral requirements. While some jobs could be performed unilaterally, it might be inefficient to do so. Another issue brought up in the focus groups was how to capture the use of a mouse while keyboarding. Additional guidance established that the use of a traditional computer mouse would be measured as both fine manipulation and gross manipulation.

In the structured cognitive test, field economists indicated that most respondents understood the question well and provided accurate answers. However, testing did reveal difficulties with respondents understanding they should include only "active" fine manipulation, including keyboarding, and that this may have led to overestimating duration. In some cases, respondents indicated that a job needed two hands, but thought that job requirements could be met using only one hand. The intent of unilateral and bilateral measurements is to determine whether a worker could reasonably perform the critical tasks with one hand. For this reason, minor changes were made to the wording of this question.

The fine manipulation measurement performed as expected in the field test, with occupations where keyboarding was likely required more of the time (e.g., accountants, and receptionists) having higher durations. The field test also assessed new wording asking about unilateral and bilateral requirements. However, field economists noted that the new wording primed respondents to think more about what they would be willing to allow instead of what is required by the critical tasks. Respondents included factors such as accommodations, performing tasks in a manner that may cause safety and health concerns or unreasonable losses of productivity, and how the job is typically performed versus how it could be performed.

Summary

This report outlines the testing activities for the third wave of ORS collection. Testing included three phases – focus groups, structured cognitive interviews, and a field test. The findings were taken together to help identify critical issues

with the proposed changes and potential clarifications of measurement objectives, as well as to refine future survey collection procedures and best practices. Most proposed changes performed well during testing with field economists and respondents noting they understood the new or modified measurements well with only minor modifications needed. However, because of critical issues identified during testing, two proposed cognitive measurements measuring the frequency of task changes and problem solving complexity were dropped. BLS and SSA continue to work together to evaluate these results and incorporate changes to the third wave to produce data relevant to SSA's disability programs.

¹ More information on ORS concepts, design, and history can be found in the Handbook of Methods at www.bls.gov/opub/hom/ors/.

² The collection manuals and forms for all reference years are available on the [information for survey respondents page](#).

³ Occupations in the ORS are classified using the [Standard Occupational Classification](#) (SOC) system. Detailed occupations are the lowest level of aggregation in the SOC system and are indicated by a six-digit SOC code that does not end in a zero. ORS distinguishes jobs from occupations; a job is considered a position where one or more workers are employed at a specific establishment whereas an occupation refers to a generalized job or family of jobs common to many industries and areas doing similar work throughout the national economy.

⁴ Defining characteristics of high problem solving complexity included not easily understood problems, with numerous options to consider because essential information is missing or needs to be sought out, requiring time to consider and review resulting in unique solutions that are unlikely to apply to other types of problems. Defining characteristics of moderate problem solving complexity included unclear problems with several options to consider because secondary information is missing or needs to be sought out, requiring time to consider and review resulting in non-routine or rule-based solutions. Defining characteristics of low problem solving complexity included straightforward, easy to understand problems with limited options to consider and all needed information available resulting in routine solutions that can be applied to multiple common problems.