



2022 No. 083

# O\*NET® Analyst Ratings of Occupational Abilities: Analysis Cycle 23 Results

## Final Report

**Prepared for:** National Center for O\*NET Development  
313 Chapanoke Road, Suite 130  
Raleigh, NC 27603

**Prepared under:** Subcontract Number (through RTI International):  
1-312-0207142-41224L

**Authors:** Timothy C. Burgoyne  
Matthew C. Reeder

**Date:** December 9, 2022



# O\*NET Analyst Ratings of Occupational Abilities: Analysis Cycle 23 Results

## Final Report

### Table of Contents

Introduction .....	1
Evaluation of Cycle 23 Analyst Ratings .....	2
Analysis 1: Cycle 23 Recommended Data Flags .....	2
Analysis 2: Cycle 23 Interrater Agreement.....	3
Analysis 3: Cycle 23 Interrater Reliability—Across Constructs within Occupations .....	8
Analysis 4: Cycle 23 Interrater Reliability—Across Occupations within Constructs .....	8
Summary.....	11
References.....	12
Appendix A Standard O*NET-IDI Codes and Titles for O*NET Analysis Cycle 23 Occupations .....	A-1
Appendix B Descriptive Statistics for O*NET Analysis Cycle 23 Importance and Level Ratings by Occupations .....	B-1
Appendix C Summary of Cycle 23 Interrater Agreement Indices .....	C-1
Appendix D Cycle 23 Interrater Reliability Coefficients.....	D-1

### List of Tables

Table 1. Number of Times Ability Level Flagged as Not Relevant.....	4
Table 2. Level Flags Due to Large $SE_M$ .....	6
Table 3. Interrater Reliabilities and Standard Errors of Measurement for Abilities Across Occupations in Cycles 14 through 23 .....	9



# O\*NET Analyst Ratings of Occupational Abilities: Analysis Cycle 23 Results

## Introduction

The Occupational Information Network (O\*NET) is a comprehensive system developed by the U.S. Department of Labor that provides information for over 900 occupations within the U.S. economy. This information is maintained in a comprehensive database. To keep the database current, the National Center for O\*NET Development is involved in a continual data collection process aimed at identifying and maintaining current information on the characteristics of workers and jobs. The information that populates the O\*NET database is collected from three primary sources: incumbents, occupational experts, and occupational analysts. Targeted job incumbents provide ratings on occupational tasks, generalized work activities (GWAs), knowledge, education and training, work styles, and work context (WC) areas. Importance and level information regarding the abilities and skills associated with these occupations is collected from occupational analysts. It should be noted that there are theoretical or philosophical reasons for preferring one rater group to the other for collecting different types of data. For example, incumbents are generally more familiar with the day-to-day duties of their job; therefore, they are the best source of information regarding tasks and GWAs. In contrast, it is likely that trained analysts understand the ability and skill constructs better than incumbents and therefore should provide the ability and skill data (Tsacoumis, 2007). Granted, it is imperative that the occupational analysts have detailed occupation information in order to rate the ability and skill constructs. It has also been suggested that some incumbents deliberately inflate their ratings to influence policy decisions regarding, for example, compensation and training (Morgeson et al., 2004). Given these considerations, occupational analysts as opposed to incumbents provide the ability and skill information in the O\*NET database.

This report focuses on results pertaining to the ability ratings only. Abilities are "... relatively enduring attributes of an individual's capability for performing a particular range of different tasks" (Fleishman et al., 1999, p. 175). Abilities are sometimes referred to as traits as they tend to remain stable over long periods. The 52 O\*NET abilities cover performance applicable to a broad range of jobs in the world's economy and are grouped into four categories within the O\*NET content model: cognitive, psychomotor, physical, and sensory-perceptual.

To facilitate the ability rating process, occupational analysts are provided relevant occupational information. Trained occupational analysts are responsible for rating the importance and level of the 52 abilities for each of the O\*NET occupations. More specifically, eight trained occupational analysts provided ratings for each occupation. For a description of the entire analyst data collection process, including the preparation and distribution of the occupational data, the steps associated with the ratings process, and the collection and management of the ability ratings, see *O\*NET Analyst Ratings of Occupational Abilities: Procedures Update* (Fleisher & Tsacoumis, 2012).

To ensure a controlled data collection and management process, occupational data are being collected in groups or "analysis cycles." This report describes the results from the data collection process for the 23<sup>rd</sup> analysis cycle of 80 occupations. Reports describing each of the previous cycles are available at <https://www.onetcenter.org/research.html?c=KSA>. Results for subsequent cycles will be reported in separate reports. For a description of the O\*NET Data Collection Publication Schedule see <http://www.onetcenter.org/dataPublication.html>. Appendix A includes a listing of the IDI codes and Occupational Titles addressed in Cycle 23.

## Evaluation of Cycle 23 Analyst Ratings

As mentioned above, occupational analysts provided ratings on the importance and level of the 52 abilities for each of the 80 occupations in Cycle 23. The mean, standard deviation, and standard error of the mean ( $SE_M$ ) of the importance and level ratings were computed. These results are shown in Appendix B.

We performed four sets of analyses to evaluate the ratings that occupational analysts provided. First, we focused on identifying the data that may be difficult to interpret based on limited agreement among raters or because there is an indication that the ability level rating is not relevant for a specific occupation. Thus, a set of recommended suppression criteria was established that flagged: (a) an ability level rating as not relevant to an occupation because of low importance ratings, (b) an ability with too little agreement in importance ratings across raters for a particular occupation, and (c) an ability with too little agreement in level ratings across raters for a particular occupation.

The remaining three sets of analyses focused on computing measures of interrater agreement and interrater reliability. Poor agreement as indicated by low reliability estimates may suggest that there is confusion about the constructs, potentially due to either the nature of the construct definition or rater training. Therefore, the second analysis involved estimating interrater agreement among the eight raters in each rating group. In the third analysis, we computed the interrater reliability of the raters to determine the extent to which raters agreed about the order of and relative distance between constructs on a particular scale (i.e., importance or level) within a particular occupation. This analysis provides information regarding the consistency across raters in terms of how they rate the required level or relative importance of the 52 ability constructs to performance in a particular occupation. Finally, in the fourth analysis, we computed another interrater reliability estimate to examine the consistency of ratings across occupations within constructs. This type of interrater reliability focused on the extent to which raters agree about the order of and relative distance between occupations on a particular scale for a particular construct. The following sections describe each of the four sets of analyses in greater detail.

### **Analysis 1: Cycle 23 Recommended Data Flags**

Three distinct criteria were established to flag the ability data. All three flags affect the presentation of publicly available data (e.g., [O\\*NET OnLine](#), [My Next Move](#), [O\\*NET Web Services](#)). First, the level rating of an ability was flagged as not relevant for a particular occupation if at least six of the eight occupational analysts rated its importance as one (1), the lowest possible rating. Thus, the level rating of an ability is considered “not relevant” when that construct is not important for performance in a particular occupation. For example, in the Cycle 23 data, the level ratings for Peripheral Vision were considered not relevant for several occupations, such as Securities, Commodities, and Financial Services Sales Agents (IDI: 01366.00.1) and Librarians and Media Collections Specialists (IDI: 01741.00.1), because Peripheral Vision was not considered important for performance in these occupations. In this cycle, there were 671 not relevant flags (see Table 1 for the number of not relevant flags across the past 10 cycles). To facilitate interpretation of these results, it should be noted that there are 4,160 sets of ratings (80 occupations x 52 abilities) in the current cycle. Given this, 16.13% (671/4,160) of the ability ratings were flagged as not relevant. The average percentage of ability ratings flagged as not relevant across the previous 22 cycles is 19.16% ( $SD = 5.08\%$ ); thus, the percentage of ratings flagged in the current cycle is below the average across previous cycles. Generally, the abilities flagged as not relevant for a large number of occupations in Cycle 23

were also flagged as not relevant for a large number of occupations in previous cycles (e.g., Dynamic Flexibility, Night Vision, Peripheral Vision). Given that these constructs capture fairly specific physical or sensory capabilities intuitively not required for many occupations, these results are not surprising.

The remaining two criteria for flagging an ability for a particular occupation involve the recommended suppression of any ability importance or level mean rating that had an  $SE_M$  greater than 0.51. These criteria were established to capture those ratings deemed to have insufficient agreement across raters. The value of 0.51 was selected because  $1.00/1.96 = 0.51$ . An  $SE_M$  greater than 0.51 means that the upper and lower bounds of the confidence interval are more than one scale point away from the observed mean. There were no instances in Cycle 23 where the mean importance rating was flagged for insufficient agreement. In fact, no importance ratings received flags for an  $SE_M$  greater than 0.51 since Cycle 3. The results of the suppression criteria for level for the past 10 cycles (Cycles 14-23) are presented in Table 2. There were 26 insufficient agreement flags for level ratings in Cycle 23, with the highest number of flags occurring for Wrist-Finger Speed and Reaction Time. The percentage of flags indicating insufficient agreement for level ratings in Cycle 23 was 0.63%, which is higher than was observed for previous cycles dating back to Cycle 11, which had 0.96%.

Dating back to Cycle 1, a decreasing trend exists across cycles with respect to the percentage of ability level ratings flagged for having a large  $SE_M$  (see Tables 1 and 2 in [Reeder & Tsacoumis, 2015](#) for results from Cycles 1-16 and subsequent annual reports for results from the following analysis cycles). Although the  $SE_M$  values have decreased over time, it is likely they have reached a lower asymptote in recent cycles as it is difficult to consistently obtain rates lower than 0.05-0.15% of the ratings. Exceptions in which there have been increases in flagged ratings across the cycles, such as the increase observed for Cycle 23, have been relatively rare. The increase in agreement observed in cycles over time could be attributable to the fact that most of the occupations rated have also been rated in a previous cycle, and slightly revised rating procedures were introduced to accommodate this large percentage of repeat occupations ([Fleisher & Tsacoumis, 2012](#)). In contrast, the decrease in agreement observed in Cycle 23 could be attributed to the fact that 32 of 80 occupations examined were “new” occupations arising from the recent taxonomy update ([Green & Allen, 2020](#); [Gregory et al., 2019](#)). It seems reasonable that agreement might be slightly lower because analysts did not have prior mean ratings for these occupations as a source of information to inform their current ratings. That said, these findings suggest there remains a high level of agreement among the occupational analysts. The detailed results of the recommended data flags and suppression criteria are depicted by the shaded cells in the results presented in Appendix B.

### **Analysis 2: Cycle 23 Interrater Agreement**

Interrater agreement was assessed to determine the level of absolute agreement among the occupational analysts in ratings within a construct for a particular occupation. Measures of interrater agreement index the extent to which the eight raters provided the same rating regarding the level of an ability (e.g., Written Comprehension) required to perform within a particular occupation. To examine agreement, we calculated the standard deviation ( $SD$ ) of ratings across occupational analysts for a given construct and scale for each occupation and the  $SE_M$  of these ratings. For both indices, lower values indicate greater agreement, and vice versa.

**Table 1. Number of Times Ability Level Flagged as Not Relevant**

Element Name	Cycle 14 (N = 106)	Cycle 15 (N = 126)	Cycle 16 (N = 102)	Cycle 17 (N = 116)	Cycle 18 (N = 110)	Cycle 19 (N = 90)	Cycle 20 (N = 100)	Cycle 21 (N = 100)	Cycle 22 (N = 100)	Cycle 23 (N = 80)
1 Oral Comprehension	0	0	0	0	0	0	0	0	0	0
2 Written Comprehension	0	0	0	0	0	0	0	0	0	0
3 Oral Expression	0	0	0	0	0	0	0	0	0	0
4 Written Expression	0	0	0	0	0	0	0	0	0	0
5 Fluency of Ideas	0	0	0	0	1	0	0	0	0	0
6 Originality	0	0	0	0	0	0	0	0	0	0
7 Problem Sensitivity	0	0	0	0	0	0	0	0	0	0
8 Deductive Reasoning	0	0	0	0	0	0	0	0	0	0
9 Inductive Reasoning	0	0	0	0	0	0	0	0	0	0
10 Information Ordering	0	0	0	0	0	0	0	0	0	0
11 Category Flexibility	0	0	0	0	0	0	0	0	0	0
12 Mathematical Reasoning	1	1	0	0	1	0	0	0	0	1
13 Number Facility	0	1	0	0	1	0	0	0	0	1
14 Memorization	0	0	0	0	0	0	0	0	0	0
15 Speed of Closure	0	0	0	0	0	0	0	0	1	0
16 Flexibility of Closure	0	0	0	0	0	0	0	0	0	0
17 Perceptual Speed	0	0	0	0	0	0	0	0	0	0
18 Spatial Orientation	60	62	32	48	51	50	63	52	50	33
19 Visualization	0	0	0	0	0	0	0	0	0	0
20 Selective Attention	0	0	0	0	0	0	0	0	0	0
21 Time Sharing	0	0	0	0	0	0	0	0	0	0
22 Arm-Hand Steadiness	11	19	10	12	10	14	28	17	16	7
23 Manual Dexterity	13	18	8	15	11	19	30	21	14	8
24 Finger Dexterity	0	0	5	0	0	1	3	0	1	2
25 Control Precision	7	17	10	12	9	19	33	16	13	11
26 Multilimb Coordination	18	23	15	26	25	27	45	27	18	22
27 Response Orientation	31	34	28	38	42	41	55	40	31	28
28 Rate Control	37	47	26	35	39	42	59	44	44	29
29 Reaction Time	31	33	23	33	38	38	54	35	32	30
30 Wrist-Finger Speed	35	50	4	11	17	19	26	30	14	2
31 Speed of Limb Movement	37	59	30	57	59	48	63	55	48	50
32 Static Strength	25	35	19	32	35	31	46	34	25	26
33 Explosive Strength	82	88	28	40	46	44	63	53	46	35



Table 1. (Continued)

Element Name	Cycle 14 (N = 106)	Cycle 15 (N = 126)	Cycle 16 (N = 102)	Cycle 17 (N = 116)	Cycle 18 (N = 110)	Cycle 19 (N = 90)	Cycle 20 (N = 100)	Cycle 21 (N = 100)	Cycle 22 (N = 100)	Cycle 23 (N = 80)
34 Dynamic Strength	32	50	15	28	29	34	43	36	34	21
35 Trunk Strength	24	24	0	1	0	0	1	1	0	0
36 Stamina	28	43	23	38	37	32	50	38	32	33
37 Extent Flexibility	28	49	17	32	31	31	49	36	28	23
38 Dynamic Flexibility	90	114	79	97	99	78	87	87	85	59
39 Gross Body Coordination	27	42	23	40	38	36	53	37	32	35
40 Gross Body Equilibrium	30	46	26	41	38	37	53	38	32	35
41 Near Vision	0	0	0	0	0	0	0	0	0	0
42 Far Vision	0	0	0	0	0	0	0	0	0	0
43 Visual Color Discrimination	0	0	0	0	0	0	1	0	0	0
44 Night Vision	66	81	42	63	65	60	75	69	64	43
45 Peripheral Vision	66	77	42	57	63	59	71	66	59	42
46 Depth Perception	12	10	6	6	11	8	16	12	9	13
47 Glare Sensitivity	62	73	39	54	61	56	69	65	58	43
48 Hearing Sensitivity	1	0	0	0	0	0	0	0	1	0
49 Auditory Attention	1	0	0	0	0	0	0	1	0	0
50 Sound Localization	66	76	43	55	62	57	70	65	61	39
51 Speech Recognition	0	0	0	0	0	0	0	0	0	0
52 Speech Clarity	0	0	0	0	0	0	0	0	0	0
Total Flags out of all possible ability ratings	16.71% (921/5512)	17.89% (1172/6552)	11.18% (593/5304)	14.44% (871/6032)	16.07% (919/5720)	18.82% (881/4680)	23.19% (1206/5200)	18.75% (975/5200)	16.31% (848/5200)	16.13% (671/4160)

**Table 2. Level Flags Due to Large  $SE_M$** 

Element Name	Cycle 14 (N = 106)	Cycle 15 (N = 126)	Cycle 16 (N = 102)	Cycle 17 (N = 116)	Cycle 18 (N = 110)	Cycle 19 (N = 90)	Cycle 20 (N = 100)	Cycle 21 (N = 100)	Cycle 22 (N = 100)	Cycle 23 (N = 80)
1 Oral Comprehension	0	0	0	0	0	0	0	0	0	0
2 Written Comprehension	0	0	0	0	0	0	0	0	0	0
3 Oral Expression	0	0	0	0	0	0	0	0	0	0
4 Written Expression	0	0	0	0	0	0	0	0	0	0
5 Fluency of Ideas	0	0	0	0	0	0	0	0	0	0
6 Originality	0	0	0	0	0	0	0	0	0	0
7 Problem Sensitivity	0	0	0	0	0	0	0	0	0	0
8 Deductive Reasoning	0	0	0	0	0	0	0	0	0	0
9 Inductive Reasoning	0	0	0	0	0	0	0	0	0	0
10 Information Ordering	0	0	0	0	0	0	0	0	0	0
11 Category Flexibility	0	0	0	0	0	0	0	0	0	0
12 Mathematical Reasoning	0	0	0	0	0	0	0	0	0	0
13 Number Facility	0	0	0	0	0	0	0	0	0	0
14 Memorization	0	0	0	0	0	0	0	0	0	0
15 Speed of Closure	0	0	0	0	0	0	0	0	0	0
16 Flexibility of Closure	0	0	0	0	0	0	0	0	0	0
17 Perceptual Speed	0	0	0	0	0	0	0	0	0	0
18 Spatial Orientation	1	0	0	1	0	0	0	0	0	1
19 Visualization	1	1	1	0	0	0	0	0	0	0
20 Selective Attention	0	0	0	0	0	0	0	0	0	0
21 Time Sharing	0	0	0	0	0	0	0	0	0	0
22 Arm-Hand Steadiness	0	0	0	0	0	0	0	1	0	0
23 Manual Dexterity	0	0	0	0	0	0	0	0	0	1
24 Finger Dexterity	0	0	1	0	1	0	0	0	0	0
25 Control Precision	1	0	2	0	0	0	0	0	0	0
26 Multilimb Coordination	0	1	0	0	0	0	1	0	0	2
27 Response Orientation	0	0	1	1	0	0	0	0	0	1
28 Rate Control	0	0	0	0	0	0	0	0	0	3
29 Reaction Time	0	0	0	0	0	0	0	0	0	4
30 Wrist-Finger Speed	1	1	1	0	0	0	0	0	0	6
31 Speed of Limb Movement	0	0	1	0	0	0	0	0	0	3
32 Static Strength	0	0	0	0	0	0	0	0	0	0
33 Explosive Strength	0	0	0	0	0	1	0	0	0	0

**Table 2. (Continued)**

Element Name	Cycle 14 (N = 106)	Cycle 15 (N = 126)	Cycle 16 (N = 102)	Cycle 17 (N = 116)	Cycle 18 (N = 110)	Cycle 19 (N = 90)	Cycle 20 (N = 100)	Cycle 21 (N = 100)	Cycle 22 (N = 100)	Cycle 23 (N = 80)
34 Dynamic Strength	0	0	0	0	0	0	0	0	0	0
35 Trunk Strength	0	0	0	0	0	0	0	0	0	0
36 Stamina	0	0	0	0	0	0	0	0	0	0
37 Extent Flexibility	0	0	0	0	0	0	0	0	0	1
38 Dynamic Flexibility	1	0	1	0	0	0	1	0	0	0
39 Gross Body Coordination	0	0	0	0	0	0	0	0	0	0
40 Gross Body Equilibrium	1	0	0	0	0	0	0	0	0	0
41 Near Vision	0	0	0	2	0	1	0	0	0	0
42 Far Vision	0	0	0	0	0	0	0	0	0	0
43 Visual Color Discrimination	0	0	0	0	0	0	0	0	0	0
44 Night Vision	0	0	0	0	0	0	0	0	0	0
45 Peripheral Vision	0	0	0	0	0	0	0	0	0	0
46 Depth Perception	0	0	0	0	0	0	0	0	0	0
47 Glare Sensitivity	0	0	0	0	0	0	0	0	0	0
48 Hearing Sensitivity	0	0	0	0	0	0	0	0	0	0
49 Auditory Attention	0	0	0	0	0	0	0	0	0	1
50 Sound Localization	0	1	0	0	0	0	0	0	0	3
51 Speech Recognition	0	0	0	0	0	0	0	0	0	0
52 Speech Clarity	0	0	0	0	0	0	0	0	0	0
Total Flags out of all possible ability ratings	0.11% (6/5512)	0.06% (4/6552)	0.15% (8/5304)	0.07% (4/6032)	0.02% (1/5720)	0.04% (2/4680)	0.04% (2/5200)	0.02% (1/5200)	0.00% (0/5200)	0.63% (26/4160)

A summary of these results is shown in Appendix C. The columns labeled “Mean of  $M_s$ ” show the mean of the occupational analyst mean importance and level ratings across the 52 abilities for each occupation<sup>1</sup>. The columns labeled “Median of  $SD_s$ ” show the median of the  $SD_s$  associated with each mean importance and level rating across the 52 abilities for each occupation. Finally, the columns labeled “Median of  $SE_{M_s}$ ” show the median of the  $SE_{M_s}$  associated with each mean importance and level rating across the 52 abilities for each occupation.

The importance ratings across all occupations had a median  $SD$  of 0.46 and a median  $SE_M$  of 0.16. The level ratings across occupations also had a median  $SD$  of 0.46 and a median  $SE_M$  of 0.16. These values are slightly higher than Cycle 22 (median  $SD$  = 0.35, median  $SE_M$  = 0.13), but still reflect strong agreement.

### **Analysis 3: Cycle 23 Interrater Reliability—Across Constructs within Occupations**

To examine the interrater reliability of the Cycle 23 ratings, we calculated intraclass correlations ( $ICC[C, k]$ ; McGraw & Wong, 1996) among the occupational analysts’ ratings to assess consistency across constructs within occupations. This statistic indicates the degree of the similarity in the rank ordering and relative distance between the abilities on a particular scale within an occupation. Our target level of interrater reliability is a median  $ICC(C, k)$  of 0.80 or greater. The value of 0.80 is judged to be a good rule-of-thumb that has been used in multiple contexts, including O\*NET (e.g., Clement et al., 2003; [McCloy et al., 1999](#); Rase & Tognetti-Stuff, 1983).

The results of these analyses are presented in Appendix D. The results revealed high levels of interrater reliability across the 80 Cycle 23 occupations. Specifically, the median  $ICC$  for importance ratings for the abilities across the occupations was 0.98 ( $M$  = 0.96,  $SD$  = 0.04). The median  $ICC$  for the level ratings was 0.98 ( $M$  = 0.96,  $SD$  = 0.05). The reliability for both the importance and level ratings exceeded the median target coefficient value of 0.80. All the reliability estimates were greater than 0.80, with the exception of the level reliability for Farmworkers and Laborers, Crop, Nursery, and Greenhouse ( $ICC$  = 0.77). Overall, the results support a very good level of reliability in the occupational analysts’ ratings.

### **Analysis 4: Cycle 23 Interrater Reliability—Across Occupations within Constructs**

Another way to evaluate the reliability of the occupational analysts’ ratings is to examine the consistency of the ratings across occupations within constructs. This type of reliability is the extent to which raters agree about the order of and relative distance among occupations on a particular scale for a particular construct. For example, is there consistency across raters in how they differentiate among occupations on the required level of the ability Oral Comprehension? To make this evaluation, McGraw and Wong’s (1996)  $ICC(C, k)$  is calculated for each construct on each scale (instead of for each occupation on each scale as described above). Consequently, each of the 52 ability importance scale ratings will have a reliability value. A median  $ICC(C, k)$  across the construct ratings for a particular domain on a particular scale of 0.80 or greater is the target interrater reliability for this coefficient (e.g., the median reliability across 52 ability level ratings should be at least 0.80). Again, the value of 0.80 has been judged to be a good rule-of-thumb.

---

<sup>1</sup> Although the mean is not a measure of agreement, it can affect the potential range of the  $SD$  and  $SE_M$ .

This reliability analysis was conducted for abilities across all occupations for the past 10 cycles<sup>2</sup> and results are presented in Table 3. The reliability analyses are based on 1,030 rating targets<sup>3</sup>. The values in the columns titled *ICC(C,1)* reflect the single-rater reliabilities, whereas the values in the columns titled *ICC(C,8)* reflect the reliability for eight raters. Overall, the median *ICC(C,8)* across the construct ratings for importance was 0.93 ( $M = 0.91$ ,  $SD = 0.07$ ) and for level was 0.95 ( $M = 0.94$ ,  $SD = 0.04$ ). This indicates that on the whole, the reliabilities achieved the target level. The majority of the abilities had high *ICC(C,8)* reliabilities for both importance and level. In fact, there were 36 abilities with reliabilities greater than 0.90 for the importance ratings and 44 abilities with reliabilities greater than or equal to 0.90 for the level ratings (e.g., Spatial Orientation).

**Table 3. Interrater Reliabilities and Standard Errors of Measurement for Abilities Across Occupations in Cycles 14 through 23**

Ability		Cycles 14 through 23 ( $N = 1,030$ )					
		Importance			Level		
		<i>ICC(C,1)</i>	<i>ICC(C,8)</i>	<i>SE</i>	<i>ICC(C,1)</i>	<i>ICC(C,8)</i>	<i>SE</i>
1	Oral Comprehension	0.54	0.90	0.12	0.70	0.95	0.14
2	Written Comprehension	0.67	0.94	0.13	0.79	0.97	0.14
3	Oral Expression	0.60	0.92	0.13	0.73	0.96	0.14
4	Written Expression	0.67	0.94	0.14	0.81	0.97	0.15
5	Fluency of Ideas	0.60	0.92	0.14	0.70	0.95	0.17
6	Originality	0.62	0.93	0.14	0.72	0.95	0.16
7	Problem Sensitivity	0.47	0.88	0.14	0.68	0.95	0.15
8	Deductive Reasoning	0.56	0.91	0.13	0.71	0.95	0.15
9	Inductive Reasoning	0.61	0.93	0.13	0.72	0.95	0.15
10	Information Ordering	0.35	0.81	0.15	0.57	0.92	0.14
11	Category Flexibility	0.37	0.83	0.14	0.57	0.91	0.15
12	Mathematical Reasoning	0.68	0.94	0.14	0.79	0.97	0.16
13	Number Facility	0.61	0.92	0.14	0.72	0.95	0.17
14	Memorization	0.37	0.83	0.15	0.51	0.89	0.16
15	Speed of Closure	0.37	0.82	0.15	0.52	0.90	0.17
16	Flexibility of Closure	0.40	0.84	0.15	0.55	0.91	0.16
17	Perceptual Speed	0.41	0.85	0.15	0.51	0.89	0.15
18	Spatial Orientation	0.72	0.95	0.13	0.75	0.96	0.18
19	Visualization	0.55	0.91	0.15	0.64	0.94	0.18
20	Selective Attention	0.21	0.68	0.13	0.34	0.81	0.15
21	Time Sharing	0.33	0.80	0.15	0.40	0.84	0.16
22	Arm-Hand Steadiness	0.84	0.98	0.14	0.86	0.98	0.17
23	Manual Dexterity	0.83	0.97	0.14	0.86	0.98	0.17
24	Finger Dexterity	0.66	0.94	0.15	0.69	0.95	0.20

<sup>2</sup> Starting in Cycle 22, interrater reliability analyses across occupations were limited to the past 10 cycles to reflect more recent trends. Previous reports (e.g., Reeder et al., 2020) include all cycles.

<sup>3</sup> A rating target refers to a unique instance of an occupation. An occupation can contribute more than one rating target if it has been rated more than once across data collection cycles.

**Table 3. (Continued)**

		Cycles 14 through 23 (N = 1,030)					
		Importance			Level		
Ability		ICC(C,1)	ICC(C,8)	SE	ICC(C,1)	ICC(C,8)	SE
25	Control Precision	0.82	0.97	0.14	0.85	0.98	0.18
26	Multilimb Coordination	0.82	0.97	0.14	0.87	0.98	0.17
27	Response Orientation	0.76	0.96	0.13	0.80	0.97	0.19
28	Rate Control	0.80	0.97	0.13	0.83	0.97	0.17
29	Reaction Time	0.82	0.97	0.13	0.86	0.98	0.18
30	Wrist-Finger Speed	0.53	0.90	0.15	0.60	0.92	0.22
31	Speed of Limb Movement	0.65	0.94	0.13	0.70	0.95	0.19
32	Static Strength	0.84	0.98	0.13	0.89	0.98	0.16
33	Explosive Strength	0.53	0.90	0.14	0.53	0.90	0.21
34	Dynamic Strength	0.71	0.95	0.14	0.79	0.97	0.18
35	Trunk Strength	0.68	0.94	0.16	0.71	0.95	0.21
36	Stamina	0.80	0.97	0.12	0.85	0.98	0.16
37	Extent Flexibility	0.82	0.97	0.13	0.88	0.98	0.17
38	Dynamic Flexibility	0.42	0.85	0.10	0.42	0.85	0.14
39	Gross Body Coordination	0.80	0.97	0.11	0.85	0.98	0.15
40	Gross Body Equilibrium	0.78	0.97	0.11	0.80	0.97	0.16
41	Near Vision	0.24	0.72	0.15	0.41	0.85	0.15
42	Far Vision	0.38	0.83	0.15	0.46	0.87	0.17
43	Visual Color Discrimination	0.55	0.91	0.15	0.63	0.93	0.19
44	Night Vision	0.69	0.95	0.10	0.70	0.95	0.16
45	Peripheral Vision	0.76	0.96	0.10	0.77	0.96	0.15
46	Depth Perception	0.69	0.95	0.14	0.75	0.96	0.19
47	Glare Sensitivity	0.76	0.96	0.10	0.80	0.97	0.16
48	Hearing Sensitivity	0.54	0.90	0.15	0.59	0.92	0.20
49	Auditory Attention	0.51	0.89	0.15	0.59	0.92	0.19
50	Sound Localization	0.71	0.95	0.11	0.73	0.96	0.16
51	Speech Recognition	0.40	0.84	0.14	0.55	0.91	0.15
52	Speech Clarity	0.51	0.89	0.14	0.65	0.94	0.16

Note. These ICCs indicate how consistently raters rated (rank ordered) occupations on a given ability. SE = Standard error of measurement = Observed score standard deviation times the square root of one minus ICC(C,8).

The lowest importance ICC(C,8) reliabilities were for Selective Attention and Near Vision (0.68 and 0.72, respectively). These abilities were among those that had lower importance reliabilities (around 0.70 or less) in Cycles 21 and 22 as well. These were the only two constructs that had importance ICC(C,8) values less than 0.80 in Cycles 22 and 23. The construct with the lowest level rating reliability was Selective Attention (0.81). No constructs had level reliabilities less than 0.80, replicating a finding from Cycle 22.

Some variation in calculated values is likely to occur by chance. As previously described, the goal was for the  $ICC(C,8)$  reliabilities to have a median value of 0.80 or greater across constructs, which was achieved for both importance and level (0.93 and 0.95, respectively). These results suggest that there was a very good level of agreement among the raters with respect to the order and relative distance among occupations on specific constructs for importance and level.

## Summary

The main findings of the analysis of Cycle 23 analyst ratings were as follows:

- About 84% of the ability ratings were considered important for performance in a given occupation. Constructs that were flagged as not relevant for performance were very similar to those flagged in previous cycles and are not unexpected given the specificity of those abilities.
- No importance ratings were flagged based on a  $SE_M$  greater than 0.51.
- Although still low in an absolute sense, a higher percentage of level ratings (0.63%) was flagged for having an  $SE_M$  greater than 0.51 compared to recent cycles. This is likely due to many of the occupations examined this cycle having not been rated previously due to updates stemming from the transition to the 2019 O\*NET-SOC taxonomy.
- There was strong interrater agreement this cycle as evidenced by the overall low medians of  $SE_{MS}$ .
- All but one of the within-occupation  $ICC$  reliabilities were above the target value of 0.80. These high levels of interrater reliability indicate that the occupational analysts rank ordered the abilities within each occupation similarly on both importance and level.
- Nearly all across-occupation  $ICC$  reliabilities were above the target value of 0.80. These high levels of interrater reliability indicate that analysts rank-ordered occupations within each ability similarly on both importance and level.

Given these results, it appears that the analysts are calibrated with one another and understand the abilities and associated definitions. Agreement was high and there is clear evidence regarding the high quality of the data. Nevertheless, project staff will continue to review the constructs and data collection process with returning analysts prior to each new cycle and as needed, throughout a cycle.

## References

- Clement, L., Chauvot, J., Philipp, R., & Ambrose, R. (2003). A method for developing rubrics for research purposes. In N. A. Pateman, B. J. Dougherty, & J. T. Zilliox (Eds.), *Proceedings of the 2003 joint meeting of PME and PMENA* (Vol. 2, pp. 221–227). Honolulu: CRDG, College of Education, University of Hawaii.
- Fleisher, M. S., & Tsacoumis, S. (2012). *O\*NET® analyst occupational abilities ratings: Procedures update* (FR-11-66). Alexandria, VA: Human Resources Research Organization. <https://www.onetcenter.org/reports/AnalystProcUpdate.html>.
- Fleishman, E. A., Costanza, D. P., & Marshall-Mies, J. (1999). Abilities. In N.G. Peterson, M.D. Mumford, W. C. Borman, P. R. Jeanneret, & E. A. Fleishman (Eds.), *An occupational information system for the 21st century: The development of O\*NET* (p.175-195). Washington D.C.: American Psychological Association. <https://doi.org/10.1037/10313-000>
- Green, J. P., & Allen, M. T. (2020). *O\*NET-SOC 2019 Taxonomy Development (2020 No. 045)*. Alexandria, VA: Human Resources Research Organization. <https://www.onetcenter.org/reports/TaxonomyDev2019.html>.
- Gregory, C., Lewis, P., Frugoli, P. & Nallin, A. (2019). *Updating the O\*NET-SOC Taxonomy: Incorporating the 2018 SOC Structure*. National Center for O\*NET Development. <https://www.onetcenter.org/reports/Taxonomy2019.html>.
- McCloy, R., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999, July). *Determining the occupational reinforce patterns for O\*NET occupational units (Volume I: Report)*. Raleigh, NC: National Center for O\*NET Development. <https://www.onetcenter.org/reports/ORP.html>.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30-46. <https://doi.org/10.1037/1082-989X.1.1.30>
- Morgeson, F. P., Delaney-Klinger, K., Mayfield, M. S., Ferrara, P., & Campion, M. A. (2004). Self-presentation processes in job analysis: A field experiment investigating inflation in abilities, tasks, and competencies. *Journal of Applied Psychology*, 89, 674-686. <https://doi.org/10.1037/0021-9010.89.4.674>
- Rase, C. W. & Tognetti-Stuff, R. K. (1983). Reliability of the auditing process at the University of Montana's Physical Therapy Department. *Physical Therapy*, 64(7), 1088-1090. <https://doi.org/10.1093/ptj/64.7.1088>
- Reeder, M. C., Burgoyne, T. C., & Allen, M. T. (2020). *O\*NET® analyst ratings of occupational abilities: Analysis cycle 21 results (2020 No. 077)*. Human Resources Research Organization.
- Reeder, M. C., & Tsacoumis, S. (2015). *O\*NET® analyst occupational ratings: Linkage revisit (2015 No. 022)*. Alexandria, VA: Human Resources Research Organization. <https://www.onetcenter.org/reports/LinkageRevisit.html>.



Tsacoumis, S. (2007). *The feasibility of using O\*NET to study skill changes*. Presented at the Workshop on Research Evidence Related to Future Skill Demands organized by the National Academies Center for Education, Washington, D.C.