FR 07-02

HumRRO

Human Resources Research Organization

O*NET Analyst Occupational Abilities Ratings: Analysis Cycle 7 Results

Shannon Willison
Suzanne Tsacoumis
Carrie N. Byrum

Prepared for: National Center for O*NET Development
700 Wade Avenue, P.O. Box 27625
Raleigh, NC 27605

Prepared under: Contract Number: 2108-009

June 2008

66 Canal Center Plaza, Suite 400 • Alexandria, VA 22314 • 703.549.3611 • fax. 703.549.9025
www.humrro.org

O*NET ANALYST OCCUPATIONAL ABILITIES RATINGS: ANALYSIS CYCLE 7 RESULTS

Table of Contents

List of Tables

O*NET ANALYST OCCUPATIONAL ABILITIES RATINGS: ANALYSIS CYCLE 7 RESULTS

Introduction

The Occupational Information Network (O*NET) is a comprehensive system developed by the U.S. Department of Labor that provides information about nearly 1,000 occupations within the U.S. economy. The National Center for O*NET Development is in the process of collecting occupational data for 812 occupations. The data collection effort includes job incumbent ratings on occupational tasks, skills, generalized work activities (GWA), knowledge, education and training, work styles, and work context areas. Importance and level information regarding the abilities associated with these occupations is being collected from 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's likely that trained analysts understand the ability constructs better than incumbents and therefore should provide the ability data (Tsacoumis, 2007). Abilities are "… relatively enduring attributes of an individual's capability for performing a particular range of different tasks" (Fleishman, Costanza, & Marshall-Mies, 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. These abilities are grouped into four categories: cognitive, psychomotor, physical, and sensory-perceptual constructs.

To facilitate the ability rating process, analysts are provided relevant occupational information. Trained analysts are responsible for rating the importance and level of the 52 abilities for each of the O*NET occupations. More specifically, eight trained 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 Occupational Abilities Ratings: Procedures* (Donsbach, Tsacoumis, Sager, & Updegraff, 2003).

To ensure a controlled data collection and management process, occupational data is being collected in groups or "analysis cycles." This report describes the results from the data collection process for the seventh analysis cycle, which included 101 occupations. Reports describing each of the previous cycles are available at http://www.onetcenter.org/resData.html#waves. Results for subsequent cycles will be reported in separate reports. For a description of the O*NET Data Collection Publication Schedule see www.onetcenter.org. The O*NET-SOC Codes and Titles included in O*NET Analysis Cycle 7 are presented in Appendix A.

Evaluation of Cycle 7 Analyst Ratings

As mentioned above, analysts provided ratings on importance and level of the 52 abilities for each of the 101 occupations in Cycle 7. The mean, standard deviation, and SE_M of the importance and level ratings were computed. These results are presented in Appendix B.

Four sets of analyses were performed to evaluate the ratings that 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 criteria was established which 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 or reliability estimates may be an indication that there is confusion about the ability constructs, potentially due to either the nature of the definition or rater training. Specifically, the second analysis involved computing the interrater agreement among the eight raters in each rating group. Next, the interrater reliability of the raters was computed 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. That is, 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, another interrater reliability of the raters was computed to examine the consistency of ratings across occupations within constructs. In other words, 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.

Cycle 7 Recommended Data Flags

Three distinct criteria were established to flag the ability data. All three flags affect the presentation of data within the publicly available O*NET Online (online.onetcenter.org). First, the level rating of an ability was flagged as not relevant for a particular occupation if two or fewer of the eight analysts rated its importance as two or greater. Thus, the level rating of an ability is considered not relevant when that ability is not important for the performance of the particular occupation. For example, in the Cycle 7 data, the level ratings for the Number Facility ability were considered not relevant for Crossing Guards (33-9091.00) as well as Laundry and Dry Cleaning Workers (51-6011.00) because Number Facility was not considered important for the performance of these two occupations. In this cycle, there were 605 not relevant flags (see Table 1). To facilitate interpretation of these results, it should be noted that there are 5,252 sets of ratings (101 occupations x 52 abilities) in the current cycle. Given this, 11.52% (605/5,252) of the ability ratings were flagged as not relevant.

As can be seen in Table 1, the most common abilities identified as not relevant remain consistent with the Cycles 1-6 results. The abilities with the most flags in Cycle 7 include Dynamic Flexibility, Explosive Strength, Peripheral Vision, Night Vision, and Sound

Localization; each of these abilities has received large numbers of flags in earlier cycles. Given that these constructs capture fairly specific physical capabilities intuitively not required for many occupations, these results are not surprising.

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

Element Name	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7
Oral Comprehension	0	0	0	0	0	0	0
Written Comprehension	0	0	0	0	0	0	0
Oral Expression	0	0	0	0	0	0	0
Written Expression	0	0	0	0	0	0	0
Fluency of Ideas	0	2	0	0	0	0	0
Originality	0	7	2	0	0	1	0
Problem Sensitivity	0	0	0	0	0	0	0
Deductive Reasoning	0	0	0	0	0	0	0
Inductive Reasoning	0	0	0	0	0	0	0
Information Ordering	0	0	0	0	0	0	0
Category Flexibility	0	0	0	0	0	0	0
Mathematical Reasoning	0	6	4	1	3	4	4
Number Facility	3	5	0	1	1	3	2
Memorization	0	1	0	0	0	5	3
Speed of Closure	0	2	3	0	0	0	1
Flexibility of Closure	0	2	0	0	0	0	0
Perceptual Speed	0	1	1	0	0	0	0
Spatial Orientation	36	48	66	81	54	48	35
Visualization	0	6	3	0	2	0	0
Selective Attention	0	0	0	0	0	0	0
Time Sharing	0	0	0	0	0	0	0
Arm-Hand Steadiness	9	14	11	49	15	14	7
Manual Dexterity	9	19	9	54	16	16	7
Finger Dexterity	0	6	3	0	1	0	0
Control Precision	6	19	13	48	16	12	7
Multilimb Coordination	13	31	23	50	25	15	10
Response Orientation	30	72	50	66	39	28	16
Rate Control	35	88	57	73	43	29	18
Reaction Time	27	65	40	66	39	23	13
Wrist-Finger Speed	26	50	54	76	50	32	21
Speed of Limb Movement	28	57	49	65	47	34	20
Static Strength	21	38	33	56	36	23	15
Explosive Strength	44	104	90	93	85	93	93
Dynamic Strength	28	61	46	65	42	28	15
Trunk Strength	8	16	23	29	30	21	6
Stamina	21	42	38	58	38	25	14
Extent Flexibility	22	47	36	64	40	24	15
Dynamic Flexibility	52	104	102	98	90	99	99
Gross Body Coordination	21	46	36	58	38	25	14
Gross Body Equilibrium	27	67	53	61	44	26	14

Continued on next page

Element Name	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7
Near Vision	0	0	0	0	0	0	0
Far Vision	0	4	3	0	0	0	0
Visual Color Discrimination	2	18	7	2	1	0	0
Night Vision	44	99	83	83	58	53	40
Peripheral Vision	44	85	79	82	55	54	41
Depth Perception	11	21	24	35	12	13	6
Glare Sensitivity	41	93	68	84	48	45	30
Hearing Sensitivity	2	39	32	16	3	0	0
Auditory Attention	2	10	4	1	2	0	0
Sound Localization	44	95	83	84	54	52	39
Speech Recognition	0	0	0	0	0	0	0
Speech Clarity	0	0	0	0	0	0	0
Total Flags out of all possible ratings	23.36% (656/2808)	22.74% (1,490/6,552)	21.67% (1,228/5,668)	30.75% (1,599/5,200)	21.70% (1,027/4,732)	16.25% (845/5,200)	11.52% (605/5,252)

The remaining two criteria for suppression involve identifying any importance or level mean rating that had a standard error of the mean (SE_M) greater than .51. These criteria were established to capture those ratings deemed to have insufficient agreement across raters. The value of .51 was selected because 1.0/1.96 = .51. An SE_M greater than .51 means that the upper and lower bounds of the confidence interval are more than one scale point away from the observed mean. The results of these two suppression criteria for Cycles 1-7 are presented in Table 2. For ease of presentation, the data from Cycles 1 and 2 are averaged and presented in one column. As can be noted, there were no instances in Cycle 7 where the mean importance rating was flagged for insufficient agreement. There were 119 insufficient agreement flags for level ratings, 110 of these flagged constructs also had ability level ratings flagged as not relevant (92.44% of 119). It should be noted that the number of flags indicating insufficient agreement with level ratings in Cycle 7 more than doubled as compared to Cycle 6. Cycle 6 had 49 flags for level (.94%), whereas in Cycle 7 there were 119 level flags (2.27%). However, the number of level flags for Cycle 7 is consistent with results from Cycles 4 and 5.

In Cycle 7, the abilities that were flagged the most for the level criteria included: Reaction Time (n = 13), Wrist-Finger Speed (n = 15), Extent Flexibility (n = 10), and Auditory Attention (n = 10). In most cases, the abilities with the most flags in Cycle 7 also received many flags in previous cycles. However, there are a few points of note. Visualization, identified in Cycle 6 as one of the abilities with the most flags (n = 4), showed a decrease in Cycle 7 (n = 1). Two other abilities showed a one-point decline in the number of flags in Cycle 7 (multi-limb coordination and explosive strength); the number of flags for each of the remaining 50 abilities either remained the same or increased in Cycle 7. Among the 25 abilities that showed increases in the number of flags, the magnitudes of the increases ranged from as few as one to as many as 13. We explored the possibility that the increase of flags may be due to a larger concentration of occupations from one job family; approximately 40% of Cycle 7 was comprised of Production Occupations. However, only 32 out of 119 flags were from production occupations. Also, it is unlikely that the flag increase is due to raters misunderstanding the constructs, as 15 of the 16 analysts had previous rating experience. For these reasons, it is not entirely clear why the number of flags increased for Cycle 7, but we will continue to monitor this trend and conduct follow up training for targeted constructs.

Table 2. Ability Flags Due to Large SE_M

Element Name	Frequency SE_M Importance > 51						Frequency SE_M Level > 51					
	Cycle 1&2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 1&2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7
Oral Comprehension	0	0	0	0	0	0	0	0	0	0	0	0
Written Comprehension	0	0	0	0	0	0	0	0	0	0	0	0
Oral Expression	0	0	0	0	0	0	0	0	0	0	0	0
Written Expression	0	0	0	0	0	0	0	0	0	0	0	1
Fluency of Ideas	0	0	0	0	0	0	7.5	10	1	0	0	1
Originality	0	0	0	0	0	0	2	8	0	0	0	1
Problem Sensitivity	0	0	0	0	0	0	0	1	0	1	0	0
Deductive Reasoning	0	0	0	0	0	0	0	2	0	0	0	0
Inductive Reasoning	0	0	0	0	0	0	0.5	1	0	0	0	0
Information Ordering	0	0	0	0	0	0	0.5	1	0	1	0	0
Category Flexibility	0	0	0	0	0	0	1	10	0	1	0	0
Mathematical Reasoning	0	0	0	0	0	0	4	3	1	1	0	0
Number Facility	0	0	0	0	0	0	8	10	9	3	1	2
Memorization	0	0	0	0	0	0	10.5	18	1	5	3	5
Speed of Closure	0	0	0	0	0	0	18	29	5	10	4	4
Flexibility of Closure	1	0	0	0	0	0	21.5	35	22	5	1	1
Perceptual Speed	0	0	0	0	0	0	13.5	15	9	3	0	1
Spatial Orientation	0.5	0	0	0	0	0	5	6	1	1	1	4
Visualization	0	0	0	0	0	0	16	26	16	6	4	1
Selective Attention	0	0	0	0	0	0	1	6	0	2	0	1
Time Sharing	0	0	0	0	0	0	3	7	0	1	0	0
Arm-Hand Steadiness	0	0	0	0	0	0	2.5	3	0	0	0	0
Manual Dexterity	0	0	0	0	0	0	7	9	2	4	0	0
Finger Dexterity	0	0	0	0	0	0	10	9	0	3	0	0
Control Precision	0	0	0	0	0	0	4.5	8	4	5	1	1
Multilimb Coordination	0	0	0	0	0	0	4	5	1	5	1	0
Response Orientation	0	0	0	0	0	0	7	11	4	3	1	5
Rate Control	0	0	0	0	0	0	2.5	6	0	3	1	1
Reaction Time	0	0	0	0	0	0	12.5	19	4	4	3	13
Wrist-Finger Speed	0.5	0	0	0	0	0	27	7	1	2	2	15
Speed of Limb Movement	0.5	0	0	0	0	0	2.5	13	2	1	1	7

Continued on next page

Table 2. (Continued)

Element Name	Frequency SE_M Importance > 51						Frequency SE_M Level > 51					
	Cycle 1&2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 1&2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7
Static Strength	0	0	0	0	0	0	5	12	4	0	0	3
Explosive Strength	0.5	0	0	0	0	0	3	6	0	1	3	2
Dynamic Strength	0	0	0	0	0	0	5.5	9	2	2	2	2
Trunk Strength	0	0	0	0	0	0	1.5	0	0	0	0	0
Stamina	0	0	0	0	0	0	2.5	3	1	1	0	1
Extent Flexibility	0	0	0	0	0	0	7	14	0	5	4	10
Dynamic Flexibility	0	0	0	0	0	0	4	0	0	0	0	1
Gross Body Coordination	0	0	0	0	0	0	0	2	1	1	0	1
Gross Body Equilibrium	0	0	0	0	0	0	2	5	1	1	1	1
Near Vision	0	0	0	0	0	0	0	0	2	0	0	0
Far Vision	0	0	0	0	0	0	15	20	3	9	0	1
Visual Color Discrimination	0	0	0	0	0	0	10.5	18	7	4	1	2
Night Vision	0	0	0	0	0	0	3.5	1	0	0	3	3
Peripheral Vision	0	0	0	0	0	0	1.5	3	0	2	1	6
Depth Perception	0	0	0	0	0	0	0.5	8	2	1	0	0
Glare Sensitivity	0	0	0	0	0	0	2	9	1	0	0	2
Hearing Sensitivity	0	0	0	0	0	0	4.5	10	5	4	1	2
Auditory Attention	0	0	0	0	0	0	5	23	0	2	6	10
Sound Localization	0.5	0	0	0	0	0	5	8	4	3	2	5
Speech Recognition	0	0	0	0	0	0	4	3	4	2	1	3
Speech Clarity	0	0	0	0	0	0	1	6	0	1	0	0
TOTAL	0% (7/9360)	0% (0/5668)	0% (0/5200)	0% (0/4732)	0% (0/5200)	0% (0/5252)	5.81% (544/9360)	7.82% (443/5668)	2.31% (120/5200)	2.30% (109/4732)	.94% (49/5200)	2.27% (119/5252)

The previous six cycles showed an encouraging trend reflecting a decrease in the percentage of ability level ratings receiving flags. Although Cycle 7 does not uphold this trend, it should be noted that the total number of level flags reflected only 2.27% of the 5,252 total ratings. This is an improvement from Cycles 2 and 3 and is consistent with Cycles 4 and 5. These findings suggest there remains a high level of agreement among the analysts. However, it continues to be advisable to monitor the elements that have been noted as problematic and to address these issues with analysts in future cycles.

The detailed results of the recommended data flags and suppression criteria are depicted by the shaded cells in the results presented in Appendix B.

Cycle 7 Interrater Agreement

Interrater agreement was computed to examine the level of absolute agreement among the analysts in ratings within a construct for a particular occupation. For example, these indices identified the extent to which eight raters provided the same rating regarding the level of the ability *Written Comprehension* required to perform a particular occupation. To look at the agreement, we calculated the standard deviation (*SD*) of ratings across analysts for a given construct and scale for each occupation and the SE_M of these ratings. For both indices, lower values indicate higher agreement, and vice versa.

A summary of these results is shown in Appendix C. The columns labeled "Mean of *Ms*" show the mean of the analyst mean importance and level ratings across the 52 abilities for each occupation.[1] The columns labeled "Median of *SDs*" show the median of the *SDs* associated with each mean importance and level rating across the 52 abilities for each occupation. Finally, the columns labeled "Median of SE_Ms" show the median of the SE_Ms 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 .60 and a median SE_M of .21. The level ratings across occupations had a median *SD* of .74 and a median SE_M of .26. These results for importance and level reveal that raters agreed slightly less in Cycle 7 than they did in Cycle 6; however, these values are relatively consistent with agreement in Cycles 1-6. Overall, while the values are generally greater for the level than they are for the importance, the results indicate that the ratings made by the analysts were consistent for both scales.

Cycle 7 Interrater Reliability: Across Constructs Within Occupations

To examine the interrater reliability of the Cycle 7 ratings we calculated the interclass correlations (*ICC* [3, *k*]; Shrout & Fleiss, 1979) among the analysts' ratings to look at consistency across constructs within occupations. As mentioned previously, this calculation examines 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* (3, *k*) of .80 or greater. The value of .80 is judged to be a good rule-of-thumb that has been used previously in the O*NET context (e.g., McCloy, Waugh, & Medsker, April 1998).

[1] While the mean is not a measure of agreement, it can affect the potential range of the *SD* and SE_M.

The results of these analyses are presented in Appendix D. The data revealed high levels of interrater reliability across the 101 Cycle 7 occupations. Specifically, both the mean and median ICC for importance ratings for the abilities across the occupations was .92 (*SD* = .04). The mean and median ICC for the level ratings were .92 and .91 (*SD* = .04), respectively. The reliability for both the importance and level ratings exceeded the median target coefficient value of .80. Results also indicate that occupations with the lowest reliability coefficients for importance had the lowest values for level ratings. This may be due to the skip pattern which forces a "0" for level if the ability is rated not important. Overall, the results support a good level of agreement in the analysts' ratings.

Cycle 7 Interrater Reliability: Across Occupations Within Constructs

Another effective way to evaluate the reliability of the analyst's ratings is to look at the consistency 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, Shrout and Fleiss' (1979) *ICC*(3, *k*) is calculated for each construct on each scale (instead of for each occupation on each scale as described above). For example, each of the 52 ability importance scale ratings will have a reliability value. The target level of interrater reliability for this coefficient is that the median *ICC*(3, *k*) across the construct ratings for a particular domain on a particular scale be .80 or greater (e.g., the median reliability across 52 ability level ratings should be at least .80). The value of .80 is judged to be a good rule-of-thumb that has been used in the O*NET context before (e.g., McCloy, Waugh, & Medsker, April 1998).

This reliability analysis was conducted on all 680 occupations from Cycles 1 through 7 and results are presented in Table 3. Note that one occupation was rated in two different cycles, therefore, the reliability analyses are based on 681 rating targets. 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 .87 (*SD* = .11) and for level was .90 (*SD* = .08), indicating that on the whole, the reliabilities achieved the target level. However, there are some low reliabilities to note.

Table 3. Interrater Reliabilities and Standard Errors of Measurement Across Occupations in Cycles 1 through 7

		Cycles 1 through 7 (*N* = 681)					
		Importance			Level		
	Ability	ICC(C,1)	ICC(C,8)	s_E	ICC(C,1)	ICC(C,8)	s_E
1	Oral Comprehension	0.38	0.83	0.19	0.48	0.88	0.21
2	Written Comprehension	0.49	0.88	0.19	0.60	0.92	0.22
3	Oral Expression	0.47	0.88	0.18	0.53	0.90	0.21
4	Written Expression	0.48	0.88	0.20	0.63	0.93	0.24
5	Fluency of Ideas	0.40	0.84	0.22	0.46	0.87	0.30
6	Originality	0.49	0.88	0.21	0.54	0.90	0.27
7	Problem Sensitivity	0.35	0.81	0.19	0.49	0.89	0.23
8	Deductive Reasoning	0.34	0.80	0.19	0.51	0.89	0.23

Continued on next page

Table 3. (Continued)

		Cycles 1 through 7 (N = 681)					
		Importance			Level		
	Ability	ICC(C,1)	ICC(C,8)	s_E	ICC(C,1)	ICC(C,8)	s_E
9	Inductive Reasoning	0.41	0.85	0.19	0.54	0.90	0.24
10	Information Ordering	0.20	0.67	0.20	0.33	0.80	0.23
11	Category Flexibility	0.20	0.67	0.20	0.28	0.76	0.27
12	Mathematical Reasoning	0.46	0.87	0.23	0.56	0.91	0.31
13	Number Facility	0.38	0.83	0.24	0.47	0.88	0.34
14	Memorization	0.19	0.65	0.24	0.23	0.71	0.36
15	Speed of Closure	0.23	0.70	0.26	0.27	0.75	0.38
16	Flexibility of Closure	0.21	0.69	0.27	0.27	0.74	0.36
17	Perceptual Speed	0.27	0.75	0.26	0.26	0.73	0.33
18	Spatial Orientation	0.55	0.91	0.19	0.55	0.91	0.27
19	Visualization	0.40	0.84	0.24	0.42	0.86	0.36
20	Selective Attention	0.16	0.60	0.21	0.18	0.63	0.26
21	Time Sharing	0.17	0.62	0.23	0.20	0.67	0.29
22	Arm-Hand Steadiness	0.72	0.95	0.20	0.71	0.95	0.27
23	Manual Dexterity	0.70	0.95	0.20	0.66	0.94	0.31
24	Finger Dexterity	0.44	0.86	0.24	0.44	0.86	0.31
25	Control Precision	0.72	0.95	0.20	0.69	0.95	0.30
26	Multilimb Coordination	0.70	0.95	0.20	0.70	0.95	0.28
27	Response Orientation	0.63	0.93	0.19	0.66	0.94	0.29
28	Rate Control	0.71	0.95	0.18	0.69	0.95	0.26
29	Reaction Time	0.70	0.95	0.19	0.70	0.95	0.31
30	Wrist-Finger Speed	0.42	0.85	0.22	0.41	0.85	0.35
31	Speed of Limb Movement	0.58	0.92	0.18	0.58	0.92	0.27
32	Static Strength	0.70	0.95	0.19	0.74	0.96	0.28
33	Explosive Strength	0.37	0.83	0.13	0.38	0.83	0.21
34	Dynamic Strength	0.59	0.92	0.19	0.63	0.93	0.27
35	Trunk Strength	0.62	0.93	0.20	0.65	0.94	0.27
36	Stamina	0.66	0.94	0.18	0.65	0.94	0.26
37	Extent Flexibility	0.72	0.95	0.18	0.72	0.95	0.30
38	Dynamic Flexibility	0.23	0.70	0.11	0.25	0.73	0.17
39	Gross Body Coordination	0.62	0.93	0.19	0.65	0.94	0.25
40	Gross Body Equilibrium	0.63	0.93	0.17	0.62	0.93	0.25
41	Near Vision	0.15	0.58	0.19	0.32	0.79	0.25
42	Far Vision	0.38	0.83	0.23	0.32	0.79	0.34
43	Visual Color Discrimination	0.40	0.84	0.24	0.43	0.86	0.34
44	Night Vision	0.54	0.90	0.16	0.54	0.90	0.23
45	Peripheral Vision	0.59	0.92	0.15	0.57	0.92	0.22
46	Depth Perception	0.61	0.93	0.20	0.61	0.93	0.28
47	Glare Sensitivity	0.65	0.94	0.15	0.67	0.94	0.23
48	Hearing Sensitivity	0.46	0.87	0.24	0.45	0.87	0.33
49	Auditory Attention	0.35	0.81	0.23	0.38	0.83	0.34
50	Sound Localization	0.51	0.89	0.16	0.53	0.90	0.25
51	Speech Recognition	0.19	0.65	0.22	0.23	0.71	0.30
52	Speech Clarity	0.36	0.82	0.20	0.39	0.83	0.29

Note. These ICCs indicate how consistently raters rated occupations on a given ability.

s_E = Standard error of measurment = Observed score variance times the square root of one minus ICC(C,8).

The lowest ICC(C,8) reliabilities were found for Near Vision, Selective Attention and Time Sharing; none of the reliabilities for these constructs had importance reliabilities over .62. The level reliabilities for Selective Attention and Time Sharing were also low, failing to exceed .67. Even though these are not the only abilities with low variation in the importance or required level across jobs, low variance is a contributing factor to low reliability. Also, the reliabilities for Selective Attention and Time Sharing are consistent with the reliabilities for these constructs in the previous two cycles.

In addition, Speech Recognition, Memorization, Category Flexibility, Information Ordering, and Flexibility of Closure all had ICC(C,8) reliabilities for importance that were less than .70 but ICC(C,8) reliabilities for level that were .70 or greater. These differences in reliabilities for importance and level likely reflect high agreement but lack of variability in the ratings of these constructs across occupations on importance yet high agreement and high variation in the ratings of these constructs across jobs on level.

Despite a few cases of low reliability, 85% of the ability constructs had high ICC(C,8) reliabilities for both importance and level (i.e., $\geq .70$). In fact, there were eight constructs with ICC(C,8) reliabilities of .95 for importance and similarly high reliabilities for level (e.g., Arm-Hand Steadiness and Reaction Time).

Comparisons with interrater reliabilities obtained from all previous cycles indicate that for most elements, the ICC(C,8) reliability estimates from Cycle 7 were consistent (within .01) of those from previous cycles. The only exceptions were small decreases (.04 or less) in reliability for Dynamic Flexibility, Sound Localization, and Speech Recognition. Although there were no large increases to note, it is important to realize that increases in the size of reliability coefficients are limited because of the relatively large coefficients already obtained on the majority of constructs.

Keep in mind that 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 across constructs of .80 or greater, which was achieved for both importance and level (.87 and .90, respectively). These results suggest that there was a good level of agreement among the raters with respect to the order and relative distance among occupations on particular constructs for importance and level.

Summary

The main findings of the analysis of Cycle 7 analyst ratings were as follows:

- The results generated from the not-relevant and suppression criteria were consistent with results from previous cycles.
- Due to the small increase in the percentage of abilities flagged for level ratings based on a SE_M greater than .51, elements once considered problematic in earlier cycles will continue to be targeted in analyst training.

- Interrater agreement for this cycle proved to be better than the standard of SE_Ms less than .51, even though raters in cycle 7 appeared to agree less than they did in cycle 6 (based on absolute agreement indices).
- All within-occupation ICC reliabilities were well above the target value of .80 (McCloy, Waugh, & Medsker, April 1998). These high levels of interrater reliability indicate that the analysts rank ordered the abilities within each occupation similarly on both importance and level.
- Index interrater reliability calculated at the end of Cycle 7 did not vary greatly from one occupation to the next.
- The importance and level median across-occupation ICC reliabilities were above the target value of .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 as though the analysts were well trained and understand the abilities and associated definitions. Review training for returning analysts and, if required, new analyst training will continue to occur prior to each new cycle. Special focus in training will be given to problem elements to maintain a low percentage of ability flags. Agreement was high and there is clear evidence regarding the quality of the data.

References

Byrum, C.N., & Tsacoumis, S. (2006a). *O*NET analyst occupational abilities ratings: Analysis Cycle 6 results.* Alexandria, VA: Human Resources Research Organization.

Byrum, C.N., & Tsacoumis, S. (2006b). *O*NET analyst occupational abilities ratings: Analysis Cycle 5 results.* Alexandria, VA: Human Resources Research Organization.

Byrum, C.N., & Tsacoumis, S. (2005). *O*NET analyst occupational abilities ratings: Analysis Cycle 4 results.* Alexandria, VA: Human Resources Research Organization.

Donsbach, J., Tsacoumis, S., Sager, C., & Updegraff, J. (2003). *O*NET analyst occupational abilities ratings: Procedures* (DFR-03-22). Alexandria, VA: Human Resources Research Organization.

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.

McCloy, R., Waugh, G., & Medsker, G. (1998, April). *Determining the occupational reinforcer patterns for O*NET occupational units*. Alexandria, VA: Human Resources Research Organization.

Noble, C.L., Sager, C., Tsacoumis, S., Updegraff, J. & Donsbach, J. (2003).*O*NET analyst occupational abilities ratings: Wave 1 results.* Alexandria, VA: Human Resources Research Organization.

Noble, C.L., & Tsacoumis, S. (2004). *O*NET analyst occupational abilities ratings: Analysis Cycle 2 results.* Alexandria, VA: Human Resources Research Organization.

Noble, C.L., & Tsacoumis, S. (2005). *O*NET analyst occupational abilities ratings: Analysis Cycle 3 results.* Alexandria, VA: Human Resources Research Organization.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

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.