FR 06-034a

HumRRO

Human Resources Research Organization

O*NET Analyst Occupational Abilities Ratings: Analysis Cycle 5 Results

Carrie N. Byrum
Suzanne Tsacoumis

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

Prepared under: Contract Number: 2108-009

May 2006

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 5 RESULTS

Table of Contents

Introduction **1**

Evaluation of Cycle 5 Analyst Ratings **2**
Cycle 5 Recommended Data Flags 2
Cycle 5 Interrater Agreement 7
Cycle 5 Interrater Reliability: Across Constructs Within Occupations 7
Cycle 5 Interrater Reliability: Across Occupations Within Constructs 8

Summary **10**

References **12**

List of Tables

Table 1. Number of Times Ability Level Flagged as Not Relevant 3
Table 2. Ability Flags Due to Large SE_M 5
Table 3. Interrater Reliabilities and Standard Errors of Measurement Across Cycle 1, 2, 3, 4, and 5 Occupations 8

O*NET ANALYST OCCUPATIONAL ABILITIES RATINGS: ANALYSIS CYCLE 5 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 over 900 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. 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 fifth analysis cycle of 91 occupations. Results for Cycle 1 are presented in Noble, Sager, Tsacoumis, Updegraff, & Donsbach (2003). Cycle 2 and Cycle 3 results are presented in Noble & Tsacoumis (2004) and Noble & Tsacoumis (2005), respectively. Finally, Cycle 4 results can be found in Byrum & Tsacoumis (2005). 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 5 are presented in Appendix A.

Evaluation of Cycle 5 Analyst Ratings

As mentioned above, analysts provided ratings on importance and level of the 52 abilities for each of the 91 occupations in Cycle 5. 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 within a particular occupation. That is, this analysis provides information regarding the consistency across raters in terms of how they rate the relative importance of the 52 ability constructs to performance in a particular occupation. Finally, another interrater reliability estimate 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 5 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 5 data, the level ratings for the Arm-Hand Steadiness ability were considered not relevant for Property, Real Estate, and Community Association Managers (11-9141.00) as well as Industrial-Organizational Psychologists (19-3032.00) because Arm-Hand Steadiness was not considered important for the performance of these two occupations. In this cycle, there were 1,027 not relevant flags (see Table 1). To facilitate interpretation of these results, it should be noted that there are 4,732 sets of ratings (91 occupations x 52 abilities) in the current cycle. Given this, 21.70% (1,027/4,732) 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 1-4 results. The abilities with the most flags in Cycle 5 include Dynamic Flexibility, Explosive Strength, Night Vision, Peripheral Vision, Spatial Orientation, 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.

It is interesting to note that the increase in flags for particular physical abilities that occurred in Cycle 4 did not continue in Cycle 5. This provides some support for the notion that the concentration of occupations in the Education, Training, and Library Job Family accounted for the unusual increase in flags for physical abilities that occurred in Cycle 4. In fact, the Cycle 5 occupations were distributed across 20 of the 23 job families.

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

Element Name	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
Oral Comprehension	0	0	0	0	0
Written Comprehension	0	0	0	0	0
Oral Expression	0	0	0	0	0
Written Expression	0	0	0	0	0
Fluency of Ideas	0	2	0	0	0
Originality	0	7	2	0	0
Problem Sensitivity	0	0	0	0	0
Deductive Reasoning	0	0	0	0	0
Inductive Reasoning	0	0	0	0	0
Information Ordering	0	0	0	0	0
Category Flexibility	0	0	0	0	0
Mathematical Reasoning	0	6	4	1	3
Number Facility	3	5	0	1	1
Memorization	0	1	0	0	0
Speed of Closure	0	2	3	0	0
Flexibility of Closure	0	2	0	0	0
Perceptual Speed	0	1	1	0	0
Spatial Orientation	36	48	66	81	54
Visualization	0	6	3	0	2
Selective Attention	0	0	0	0	0
Time Sharing	0	0	0	0	0
Arm-Hand Steadiness	9	14	11	49	15
Manual Dexterity	9	19	9	54	16
Finger Dexterity	0	6	3	0	1
Control Precision	6	19	13	48	16
Multilimb Coordination	13	31	23	50	25
Response Orientation	30	72	50	66	39
Rate Control	35	88	57	73	43
Reaction Time	27	65	40	66	39
Wrist-Finger Speed	26	50	54	76	50
Speed of Limb Movement	28	57	49	65	47

Continued on next page

Element Name	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
Static Strength	21	38	33	56	36
Explosive Strength	44	104	90	93	85
Dynamic Strength	28	61	46	65	42
Trunk Strength	8	16	23	29	30
Stamina	21	42	38	58	38
Extent Flexibility	22	47	36	64	40
Dynamic Flexibility	52	104	102	98	90
Gross Body Coordination	21	46	36	58	38
Gross Body Equilibrium	27	67	53	61	44
Near Vision	0	0	0	0	0
Far Vision	0	4	3	0	0
Visual Color Discrimination	2	18	7	2	1
Night Vision	44	99	83	83	58
Peripheral Vision	44	85	79	82	55
Depth Perception	11	21	24	35	12
Glare Sensitivity	41	93	68	84	48
Hearing Sensitivity	2	39	32	16	3
Auditory Attention	2	10	4	1	2
Sound Localization	44	95	83	84	54
Speech Recognition	0	0	0	0	0
Speech Clarity	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)

The remaining two criteria involve the recommended suppression of identifying any ability mean or level importance 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 are presented in Table 2. As can be noted, there were no instances where the mean importance rating was flagged for insufficient agreement. There were 109 insufficient agreement flags for level ratings, 25 of these flagged constructs also had ability level ratings flagged as not relevant (22.94% of 109). It should be noted that the number of flags indicating insufficient agreement remained proportionally the same between Cycle 4 and Cycle 5. As such, Cycle 4 had 100 occupations and 120 level flags for insufficient agreement (2.31%) while in Cycle 5 there were only 91 occupations and 109 flags for level (2.30%).

In Cycle 5, the abilities that were flagged the most for the level criteria included: Speed of Closure (n=10), Far Vision (n=9), and Visualization (n=6). In many cases, the abilities with the most flags in Cycle 5 also received many flags in the previous four cycles. However, there are a few points of note. First, there was a substantial drop in the number of flags for Flexibility of Closure (n=5) and Visualization (n=6). In the previous four cycles, the Flexibility of Closure ability received no fewer than 14 flags and Visualization received no fewer than 13 flags. It is possible that a number of factors may have influenced these ratings. First, the experience,

training, and feedback received by experienced raters may have contributed to increasing agreement on these constructs. Second, the introduction of four new analysts to the project and the targeted-training on problematic constructs they received may have had a substantial and lasting effect on analyst agreement.

Another interesting observation is that, after receiving an unexpected 23 flags in Cycle 3, Auditory Attention did not receive any flags in Cycle 4. This ability was closely observed in Cycle 5 and, because it received only 2 flags, targeted training on this ability does not seem warranted.

Table 2. Ability Flags Due to Large SE_M

Element Name	Frequency SE_M Importance > .51					Frequency SE_M Level > .51				
	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
Oral Comprehension	0	0	0	0	0	0	0	0	0	0
Written Comprehension	0	0	0	0	0	0	0	0	0	0
Oral Expression	0	0	0	0	0	0	0	0	0	0
Written Expression	0	0	0	0	0	0	0	0	0	0
Fluency of Ideas	0	0	0	0	0	4	11	10	1	0
Originality	0	0	0	0	0	1	3	8	0	0
Problem Sensitivity	0	0	0	0	0	0	0	1	0	1
Deductive Reasoning	0	0	0	0	0	0	0	2	0	0
Inductive Reasoning	0	0	0	0	0	0	1	1	0	0
Information Ordering	0	0	0	0	0	0	1	1	0	1
Category Flexibility	0	0	0	0	0	0	2	10	0	1
Mathematical Reasoning	0	0	0	0	0	1	7	3	1	1
Number Facility	0	0	0	0	0	1	15	10	9	3
Memorization	0	0	0	0	0	3	18	18	1	5
Speed of Closure	0	0	0	0	0	4	32	29	5	10
Flexibility of Closure	0	2	0	0	0	14	29	35	22	5
Perceptual Speed	0	0	0	0	0	12	15	15	9	3
Spatial Orientation	0	1	0	0	0	1	9	6	1	1
Visualization	0	0	0	0	0	13	19	26	16	6
Selective Attention	0	0	0	0	0	0	2	6	0	2
Time Sharing	0	0	0	0	0	0	6	7	0	1
Arm-Hand Steadiness	0	0	0	0	0	3	2	3	0	0
Manual Dexterity	0	0	0	0	0	6	8	9	2	4
Finger Dexterity	0	0	0	0	0	0	20	9	0	3
Control Precision	0	0	0	0	0	4	5	8	4	5
Multilimb Coordination	0	0	0	0	0	0	8	5	1	5
Response Orientation	0	0	0	0	0	6	8	11	4	3

Continued on next page

Element Name	Frequency SE_M Importance > .51					Frequency SE_M Level > .51				
	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
Rate Control	0	0	0	0	0	3	2	6	0	3
Reaction Time	0	0	0	0	0	6	19	19	4	4
Wrist-Finger Speed	1	0	0	0	0	21	33	7	1	2
Speed of Limb Movement	0	1	0	0	0	1	4	13	2	1
Static Strength	0	0	0	0	0	4	6	12	4	0
Explosive Strength	0	1	0	0	0	3	3	6	0	1
Dynamic Strength	0	0	0	0	0	4	7	9	2	2
Trunk Strength	0	0	0	0	0	2	1	0	0	0
Stamina	0	0	0	0	0	2	3	3	1	1
Extent Flexibility	0	0	0	0	0	1	13	14	0	5
Dynamic Flexibility	0	0	0	0	0	3	5	0	0	0
Gross Body Coordination	0	0	0	0	0	0	0	2	1	1
Gross Body Equilibrium	0	0	0	0	0	4	0	5	1	1
Near Vision	0	0	0	0	0	0	0	0	2	0
Far Vision	0	0	0	0	0	16	14	20	3	9
Visual Color Discrimination	0	0	0	0	0	5	16	18	7	4
Night Vision	0	0	0	0	0	3	4	1	0	0
Peripheral Vision	0	0	0	0	0	1	2	3	0	2
Depth Perception	0	0	0	0	0	1	0	8	2	1
Glare Sensitivity	0	0	0	0	0	2	2	9	1	0
Hearing Sensitivity	0	0	0	0	0	3	6	10	5	4
Auditory Attention	0	0	0	0	0	1	9	23	0	2
Sound Localization	0	1	0	0	0	1	9	8	4	3
Speech Recognition	0	0	0	0	0	0	8	3	4	2
Speech Clarity	0	0	0	0	0	0	2	6	0	1
TOTAL	0% (1/2808)	0% (6/6552)	0% (0/5668)	0% (0/5200)	0% (0/4732)	5.59% (157/2808)	5.91% (387/6552)	7.82% (443/5668)	2.31% (120/5200)	2.30% (109/4732)

While the frequency of flagging an ability level rating was higher than the importance rating, it should be noted that the total number of level flags reflected only 2.30% of the 4,732 total ratings. In addition, this value continues to reflect a decrease in the percentage of ability level ratings receiving flags across the first three cycles. These findings suggest there remains a high level of agreement among the analysts and that constructs that earlier appeared problematic may not require additional training. Moreover, the homogeneity of occupations that was found in Cycle 4 was not found in Cycle 5; Cycle 5 was comprised of occupations that represented multiple O*NET job families. This indicates that the reduced number of flags for level should not be attributed to the similarity among the occupations. However, it continues to be advisable to monitor the elements that were previously problematic to determine if additional training is warranted.

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 5 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 .52 and a median SE_M of .18. The level ratings across occupations had a median *SD* of .71 and a median SE_M of .25. These results for importance indicate a slight decrease in agreement over those found in the four previous cycles. Furthermore, the results for level represent an even larger decrease in agreement when compared to the results from earlier cycles. 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. However, it is important to comment that the introduction of new analysts may have contributed to the small decrease in interrater agreement.

Cycle 5 Interrater Reliability: Across Constructs Within Occupations

To examine the interrater reliability of the Cycle 5 ratings we calculated the interclass correlations ICC [3, *k*]; Shrout & Fleiss, 1979) among the analyst's 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).

The results of these analyses are presented in Appendix D. The data revealed high levels of interrater reliability across the 91 Cycle 5 occupations. Specifically, the mean ICC for importance ratings for the abilities across the occupations was .95 (*SD* = .03). The mean ICC for the level ratings was .95 (*SD* = .04). The reliability for both the importance and level ratings exceeded the target coefficient value of .80. Interrater reliability did not vary greatly across occupations and the mean coefficient for importance ratings was identical to the mean coefficient for level ratings. 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

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

forces a "0" for level if the ability is rated not important. This will be monitored when analyzing the data collected in future cycles.

Cycle 5 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 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*) must be 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 type of reliability was first used to evaluate the raters after combining results of Cycle 1 and Cycle 2 data collection because it requires a reasonable number of occupations. With the completion of Cycle 4, there were 389 occupations included in the reliability analysis. The current reliability analysis was conducted on all 480 occupations from Cycles 1 through 5 and results are presented in Table 3. 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. The lowest ICC(C,8) reliabilities were found for Speech Recognition, Memorization, Selective Attention, and Time Sharing; none of the reliabilities for these constructs had reliabilities over .68 on either level or importance. Furthermore, the reliabilities for these constructs had either remained the same or declined slightly when compared to the reliabilities found with the occupations from Cycles 1 through 4. These reliabilities may be due to low variation in the importance or the required level of these abilities across jobs or disagreement among raters.

Table 3. Interrater Reliabilities and Standard Errors of Measurement Across Cycle 1, 2, 3, 4, and 5 Occupations

Ability		Cycles 1 through 5 (*N* = 480)					
		Importance			Level		
		ICC(C,1)	ICC(C,8)	s_E	ICC(C,1)	ICC(C,8)	s_E
1	Oral Comprehension	0.33	0.80	0.18	0.46	0.87	0.21
2	Written Comprehension	0.48	0.88	0.19	0.61	0.93	0.22
3	Oral Expression	0.44	0.86	0.18	0.50	0.89	0.20
4	Written Expression	0.47	0.88	0.20	0.63	0.93	0.24
5	Fluency of Ideas	0.40	0.84	0.22	0.43	0.86	0.31
6	Originality	0.49	0.88	0.21	0.53	0.90	0.28
7	Problem Sensitivity	0.33	0.80	0.19	0.48	0.88	0.24
8	Deductive Reasoning	0.30	0.78	0.19	0.49	0.88	0.24

Continued on next page

Ability		Cycles 1 through 5 (N = 480)					
		Importance			Level		
		ICC(C,1)	ICC(C,8)	s_E	ICC(C,1)	ICC(C,8)	s_E
9	Inductive Reasoning	0.39	0.83	0.20	0.53	0.90	0.25
10	Information Ordering	0.19	0.66	0.21	0.33	0.80	0.24
11	Category Flexibility	0.20	0.67	0.21	0.28	0.76	0.28
12	Mathematical Reasoning	0.47	0.88	0.23	0.57	0.91	0.32
13	Number Facility	0.38	0.83	0.24	0.47	0.88	0.36
14	Memorization	0.15	0.59	0.24	0.21	0.68	0.37
15	Speed of Closure	0.23	0.70	0.26	0.27	0.74	0.40
16	Flexibility of Closure	0.22	0.70	0.28	0.27	0.75	0.39
17	Perceptual Speed	0.26	0.73	0.27	0.24	0.72	0.36
18	Spatial Orientation	0.57	0.91	0.18	0.56	0.91	0.27
19	Visualization	0.40	0.84	0.25	0.42	0.85	0.38
20	Selective Attention	0.16	0.60	0.21	0.17	0.62	0.26
21	Time Sharing	0.19	0.65	0.23	0.19	0.65	0.30
22	Arm-Hand Steadiness	0.69	0.95	0.21	0.69	0.95	0.28
23	Manual Dexterity	0.68	0.94	0.21	0.63	0.93	0.32
24	Finger Dexterity	0.41	0.85	0.25	0.43	0.86	0.33
25	Control Precision	0.68	0.94	0.20	0.65	0.94	0.32
26	Multilimb Coordination	0.68	0.94	0.20	0.66	0.94	0.29
27	Response Orientation	0.62	0.93	0.18	0.63	0.93	0.29
28	Rate Control	0.66	0.94	0.16	0.65	0.94	0.24
29	Reaction Time	0.69	0.95	0.18	0.68	0.94	0.31
30	Wrist-Finger Speed	0.40	0.84	0.20	0.38	0.83	0.35
31	Speed of Limb Movement	0.59	0.92	0.17	0.57	0.91	0.26
32	Static Strength	0.71	0.95	0.19	0.74	0.96	0.27
33	Explosive Strength	0.42	0.85	0.12	0.44	0.86	0.20
34	Dynamic Strength	0.65	0.94	0.17	0.65	0.94	0.26
35	Trunk Strength	0.63	0.93	0.20	0.64	0.93	0.27
36	Stamina	0.68	0.95	0.18	0.65	0.94	0.25
37	Extent Flexibility	0.74	0.96	0.17	0.75	0.96	0.29
38	Dynamic Flexibility	0.29	0.77	0.10	0.32	0.79	0.16
39	Gross Body Coordination	0.66	0.94	0.18	0.67	0.94	0.25
40	Gross Body Equilibrium	0.67	0.94	0.15	0.64	0.93	0.23
41	Near Vision	0.16	0.61	0.19	0.35	0.81	0.25
42	Far Vision	0.38	0.83	0.24	0.32	0.79	0.37
43	Visual Color Discrimination	0.40	0.84	0.24	0.42	0.85	0.36
44	Night Vision	0.60	0.92	0.13	0.56	0.91	0.22
45	Peripheral Vision	0.63	0.93	0.13	0.58	0.92	0.20
46	Depth Perception	0.60	0.92	0.20	0.58	0.92	0.29
47	Glare Sensitivity	0.68	0.94	0.13	0.69	0.95	0.21
48	Hearing Sensitivity	0.47	0.88	0.23	0.45	0.87	0.33
49	Auditory Attention	0.33	0.79	0.23	0.36	0.82	0.34
50	Sound Localization	0.54	0.91	0.14	0.54	0.90	0.23

Continued on next page

Ability	Cycles 1 through 5 (N = 480)					
	Importance			Level		
	ICC(C,1)	ICC(C,8)	s_E	ICC(C,1)	ICC(C,8)	s_E
51 Speech Recognition	0.14	0.57	0.23	0.19	0.65	0.31
52 Speech Clarity	0.32	0.79	0.20	0.38	0.83	0.29

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

Moreover, Information Ordering, Category Flexibility, and Near Vision had ICC(C,8) reliabilities for importance that also did not exceed .67 but had reliabilities for level that were greater than or equal to .76. 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 and high agreement and high variation in the ratings of these constructs across jobs on level. Furthermore, a number of abilities demonstrated small decreases in ICC(C,8) reliability after the addition of the Cycle 5 occupations. These decreases likely occurred because the low variability in importance ratings for constructs across occupations that existed for occupations from Cycle 1 through 3 was exacerbated after the introduction of new analysts to the rating process.

However, comparisons with interrater reliabilities obtained from Cycle 1, 2, 3, and 4 data indicate that for some elements ICC(C,8) reliability improved with the addition of Cycle 5 data. For example, the construct Dynamic Flexibility demonstrated increases in ICC(C,8) reliability of .06 and .08 for importance and level, respectively. 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. Median ICC(C,8) reliabilities for importance and level were .87 and .89, 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 5 analyst ratings were as follows:

- The not-relevance and suppression criteria did not generate any results reflecting poorly on the overall quality of the Cycle 5 ratings.
- The stability in the percentage of abilities flagged for level ratings due to a SE_M greater than .51 across Cycle 4 and Cycle 5 indicates that elements once considered problematic in earlier cycles may not require additional training. The similarities of Cycle 4 occupations do not appear to have unduly influenced interrater agreement.
- While interrater agreement was higher for importance than for level ratings, overall results indicate that the ratings made by the analysts were consistent for both scales across occupations. Cycle 5 results revealed a small decrease in analyst agreement for

both scales from that of the previous four cycles. It is possible that this decrease may be due to the introduction of new analysts.

- 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 5 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. Agreement was high and there is clear evidence regarding the quality of the data. Nevertheless, slight decreases in agreement may have resulted from the introduction of new analysts.

References

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