

O*NET Analyst Occupational Abilities Ratings: Wave 1 Results

Carrie L. Noble
Christopher Sager
Suzanne Tsacoumis
Joshua Updegraff
Jamie Donsbach

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

November 2003

Table of Contents

Introduction.....	1
Evaluation of Wave 1 Analyst Ratings.....	1
Wave 1 Recommended Data Flags	2
Wave 1 Interrater Agreement.....	4
Wave 1 Interrater Reliability: Across Constructs Within Occupations	6
Wave 1 Interrater Reliability: Across Occupations Within Constructs.....	8
Summary.....	9
References.....	10

List of Tables

Table 1. Recommended Data Flags	3
Table 2. Summary of Wave 1 Interrater Agreement Indices	5
Table 3. Wave 1 Interrater Reliability Coefficients.....	7

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 of time. 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 "waves." This report describes the results from the data collection process for the first wave of 54 occupations. Results for ratings collected on further waves of occupations will be reported in separate subsequent reports. For a description of the O*NET Data Collection Publication Schedule see www.onetcenter.org. The Standard Occupational Classification Codes and Titles for the O*NET occupations included in the Wave 1 are presented in Appendix A.

Evaluation of Wave 1 Analyst Ratings

As mentioned above, analysts provided ratings on importance and level of the 52 abilities for each of the 54 occupations in Wave 1. The mean, standard deviation and SE_M of the importance and level ratings were computed. These results are presented in Appendix B.

Three 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 two 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. Finally, 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.

It should be noted that after the next wave of ratings are completed, the total number of occupations will exceed 100. At this time, a reliability analysis will be computed that examines the extent to which raters agree about the order of and relative distance between 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*?

Wave 1 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 2 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. In this wave, there were 656 not relevant flags (see Table 1). To facilitate interpretation of these results, it should be noted that there are a total of 2,808 sets of ratings (54 occupations x 52 abilities) in the current wave. Given this, 23.36% (656/2808) of the ability ratings were flagged as not relevant. As can be noted in Table 1, the most common abilities identified as not relevant were Explosive Strength, Night Vision, Peripheral Vision, and Sound Localization. Given that these constructs capture fairly specific physical capabilities intuitively not required for many occupations, these results are not surprising.

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 SEM greater than .51 means that the upper and lower bounds of the confidence interval are more than 1 scale point away from the observed mean. The results of these two suppression criteria are presented in Table 1. As can be noted, there was only one instance (wrist-finger speed) where the mean importance rating was flagged for insufficient agreement, however there were 157 insufficient agreement flags for level ratings. The abilities that were flagged the most for the level criteria included: wrist-finger speed (n=21), far vision (n=16), flexibility of closure (n=14), visualization (n=13), and perceptual speed (n=12). Although 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 only reflected 5.59% of the 2,808 total ratings. These findings suggest a high level of agreement among the analysts. However, it may be prudent to provide additional training on the elements

with the highest number of flags, particularly wrist-finger speed since it was the only flag for importance.

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

Table 1. Recommended Data Flags Criteria

Element Name	Number of times ability is flagged as not relevant	Times ability is flagged because SEM importance > .51	Times ability is flagged because SEM level > .51
Oral Comprehension	0	0	0
Written Comprehension	0	0	0
Oral Expression	0	0	0
Written Expression	0	0	0
Fluency of Ideas	0	0	4
Originality	0	0	1
Problem Sensitivity	0	0	0
Deductive Reasoning	0	0	0
Inductive Reasoning	0	0	0
Information Ordering	0	0	0
Category Flexibility	0	0	0
Mathematical Reasoning	0	0	1
Number Facility	3	0	1
Memorization	0	0	3
Speed of Closure	0	0	4
Flexibility of Closure	0	0	14
Perceptual Speed	0	0	12
Spatial Orientation	36	0	1
Visualization	0	0	13
Selective Attention	0	0	0
Time Sharing	0	0	0
Arm-Hand Steadiness	9	0	3
Manual Dexterity	9	0	6
Finger Dexterity	0	0	0
Control Precision	6	0	4
Multilimb Coordination	13	0	0
Response Orientation	30	0	6
Rate Control	35	0	3
Reaction Time	27	0	6
Wrist-Finger Speed	26	1	21
Speed of Limb Movement	28	0	1
Static Strength	21	0	4
Explosive Strength	44	0	3
Dynamic Strength	28	0	4
Trunk Strength	8	0	2

Continued on next page

Table 1. Recommended Data Flags Criteria

Element Name	Number of times ability is flagged as not relevant	Times ability is flagged because SEM importance > .51	Times ability is flagged because SEM level > .51
Stamina	21	0	2
Extent Flexibility	22	0	1
Dynamic Flexibility	52	0	0
Gross Body Coordination	21	0	0
Gross Body Equilibrium	27	0	4
Near Vision	0	0	0
Far Vision	0	0	16
Visual Color Discrimination	2	0	5
Night Vision	44	0	3
Peripheral Vision	44	0	1
Depth Perception	11	0	1
Glare Sensitivity	41	0	2
Hearing Sensitivity	2	0	3
Auditory Attention	2	0	1
Sound Localization	44	0	1
Speech Recognition	0	0	0
Speech Clarity	0	0	0
TOTAL	656	1	157

Wave 1 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 Table 2. The columns labeled “Mean of M_s ” show the mean of the analyst mean importance and level ratings across the 52 abilities for each occupation.¹ 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.

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

Table 2. Summary of Wave 1 Interrater Agreement Indices

Occupation	Importance			Level		
	Mean of Ms	Median of SDs	Median of SE _{MS}	Mean of Ms	Median of SDs	Median of SE _{MS}
Sales Managers	2.35	0.53	0.19	1.99	0.76	0.27
Administrative Services Managers	2.22	0.52	0.18	1.85	0.67	0.24
Industrial Production Managers	2.48	0.52	0.18	2.38	0.73	0.26
Medical and Health Services Managers	2.35	0.46	0.16	2.13	0.64	0.23
Wholesale and Retail Buyers, Except Farm Products	2.18	0.49	0.17	1.93	0.64	0.23
Cost Estimators	2.30	0.59	0.21	2.05	0.71	0.25
Training and Development Specialists	2.38	0.52	0.18	2.10	0.75	0.27
Accountants	2.09	0.35	0.13	1.76	0.67	0.24
Architects, Except Landscape and Naval	2.32	0.52	0.18	2.13	0.74	0.26
Landscape Architects	2.49	0.46	0.16	2.28	0.52	0.18
Civil Engineers	2.45	0.64	0.23	2.55	1.06	0.38
Architectural Drafters	2.37	0.53	0.19	2.31	0.93	0.33
Biologists	2.24	0.46	0.16	2.22	0.82	0.29
Foresters	2.74	0.73	0.26	2.76	1.18	0.42
Urban and Regional Planners	2.36	0.46	0.16	2.20	0.71	0.25
Child, Family, and School Social Workers	2.03	0.46	0.16	1.68	0.59	0.21
Medical and Public Health Social Workers	2.17	0.46	0.16	1.81	0.71	0.25
Mental Health and Substance Abuse Social Workers	2.13	0.49	0.17	1.78	0.73	0.26
Art Directors	2.31	0.46	0.16	2.03	0.53	0.19
Graphic Designers	2.16	0.52	0.18	1.79	0.64	0.23
Dietitians and Nutritionists	2.15	0.00	0.00	1.81	0.52	0.18
Optometrists	2.53	0.59	0.21	2.59	0.83	0.30
Occupational Therapists	2.39	0.53	0.19	2.12	0.67	0.24
Physical Therapists	2.61	0.64	0.23	2.51	0.73	0.26
Recreational Therapists	2.41	0.71	0.25	2.13	0.96	0.34
Respiratory Therapists	2.67	0.71	0.25	2.71	0.92	0.33
Dental Hygienists	2.33	0.64	0.23	2.12	0.93	0.33
Diagnostic Medical Sonographers	2.72	0.52	0.18	2.73	0.76	0.27
Radiologic Technologists	2.67	0.52	0.18	2.57	0.59	0.21
Emergency Medical Technicians and Paramedics	3.13	0.64	0.23	3.24	0.71	0.25
Respiratory Therapy Technicians	2.70	0.67	0.24	2.56	0.92	0.33
Licensed Practical and Licensed Vocational Nurses	2.61	0.71	0.25	2.39	1.04	0.37
Medical Records and Health Information Technicians	2.27	0.53	0.19	1.85	0.83	0.30
Dental Assistants	2.51	0.59	0.21	2.27	0.71	0.25
First-Line Supervisors/Managers of Police and Detectives	3.03	0.53	0.19	3.26	0.53	0.19
Police Detectives	2.89	0.64	0.23	2.90	0.93	0.33
Police Patrol Officers	3.22	0.64	0.23	3.28	0.64	0.23
Hairdressers, Hairstylists, and Cosmetologists	2.46	0.53	0.19	2.01	0.53	0.19

Continued on next page

Table 2. Summary of Wave 1 Interrater Agreement Indices

Occupation	Importance			Level		
	Mean of M_s	Median of SD_s	Median of SE_{M_s}	Mean of M_s	Median of SD_s	Median of SE_{M_s}
First-Line Supervisors/Managers of Retail Sales Workers	2.38	0.52	0.18	2.18	0.59	0.21
Counter and Rental Clerks	2.12	0.53	0.19	1.53	0.71	0.25
Retail Salespersons	2.15	0.52	0.18	1.68	0.64	0.23
Insurance Sales Agents	2.24	0.35	0.13	1.86	0.52	0.18
Travel Agents	2.13	0.53	0.19	1.61	0.64	0.23
Receptionists and Information Clerks	2.13	0.52	0.18	1.58	0.52	0.18
Legal Secretaries	2.03	0.52	0.18	1.56	0.71	0.25
Medical Secretaries	2.21	0.53	0.19	1.72	0.74	0.26
Word Processors and Typists	2.09	0.53	0.19	1.54	0.53	0.19
Insurance Policy Processing Clerks	2.05	0.52	0.18	1.51	0.74	0.26
Office Clerks, General	2.08	0.49	0.17	1.67	0.53	0.19
Construction Carpenters	2.90	0.64	0.23	2.88	0.89	0.31
Construction and Building Inspectors	2.73	0.53	0.19	2.86	0.93	0.33
Electrical and Electronics Repairers, Commercial and Industrial Equipment	2.46	0.52	0.18	2.52	0.89	0.31
Heating and Air Conditioning Mechanics	2.98	0.64	0.23	3.03	0.83	0.3
Water and Liquid Waste Treatment Plant and System Operators	2.73	0.64	0.23	2.69	0.76	0.27

The importance ratings across all occupations had a median SD of .53 and a median SE_M of .19. The level ratings across occupations had a median SD of .74 and a median SE_M of .26. Overall, while the values are generally greater for the level than they are for the importance, these results indicate that the ratings made by the analysts were reasonably consistent for both scales.

Wave 1 Interrater Reliability: Across Constructs Within Occupations

To examine the interrater reliability of the Wave 1 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 that the median ICC (3, k) be .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 Table 3. The data revealed high levels of interrater reliability across the 54 Wave 1 occupations. Specifically, the mean ICC for importance ratings for the abilities across the occupations was .96 ($SD = .02$). The mean ICC for the level ratings was .95 ($SD = .03$). 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 just barely higher than the mean 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 forces a “0” for level if the ability is rated not important. This will be monitored when analyzing the data collected in future waves.

Table 3. Wave 1 Interrater Reliability Coefficients

Occupation	Importance	Level
Sales Managers	0.97	0.96
Administrative Services Managers	0.97	0.97
Industrial Production Managers	0.97	0.97
Medical and Health Services Managers	0.98	0.98
Wholesale and Retail Buyers, Except Farm Products	0.97	0.97
Cost Estimators	0.97	0.97
Training and Development Specialists	0.97	0.97
Accountants	0.98	0.98
Architects, Except Landscape and Naval	0.98	0.97
Landscape Architects	0.98	0.98
Civil Engineers	0.96	0.96
Architectural Drafters	0.97	0.97
Biologists	0.97	0.97
Foresters	0.92	0.89
Urban and Regional Planners	0.98	0.97
Child, Family, and School Social Workers	0.98	0.98
Medical and Public Health Social Workers	0.98	0.97
Mental Health and Substance Abuse Social Workers	0.98	0.97
Art Directors	0.98	0.98
Graphic Designers	0.97	0.97
Dietitians and Nutritionists	0.98	0.99
Optometrists	0.96	0.97
Occupational Therapists	0.97	0.97
Physical Therapists	0.96	0.96
Recreational Therapists	0.93	0.93
Respiratory Therapists	0.93	0.95
Dental Hygienists	0.94	0.95
Diagnostic Medical Sonographers	0.96	0.96
Radiologic Technologists	0.96	0.97
Emergency Medical Technicians and Paramedics	0.88	0.88
Respiratory Therapy Technicians	0.94	0.95
Licensed Practical and Licensed Vocational Nurses	0.93	0.92
Medical Records and Health Information Technicians	0.96	0.95
Dental Assistants	0.95	0.95
First-Line Supervisors/Managers of Police and Detectives	0.94	0.91
Police Detectives	0.94	0.92

Continued on next page

Table 3. Wave 1 Interrater Reliability Coefficients

Occupation	Importance	Level
Police Patrol Officers	0.93	0.93
Hairdressers, Hairstylists, and Cosmetologists	0.96	0.96
First-Line Supervisors/Managers of Retail Sales Workers	0.96	0.97
Counter and Rental Clerks	0.94	0.94
Retail Salespersons	0.95	0.96
Insurance Sales Agents	0.99	0.98
Travel Agents	0.97	0.97
Receptionists and Information Clerks	0.97	0.97
Legal Secretaries	0.96	0.96
Medical Secretaries	0.97	0.96
Word Processors and Typists	0.96	0.96
Insurance Policy Processing Clerks	0.97	0.97
Office Clerks, General	0.97	0.96
Construction Carpenters	0.89	0.86
Construction and Building Inspectors	0.94	0.93
Electrical and Electronics Repairers, Commercial and Industrial Equipment	0.94	0.94
Heating and Air Conditioning Mechanics	0.92	0.91
Water and Liquid Waste Treatment Plant and System Operators	0.90	0.88

Wave 1 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 between 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 not used to evaluate the raters during the Wave 1 data collection process because it cannot be calculated until a reasonable number of occupations have been rated by the analysts. Since there will be data on more than 100 occupations following the completion of the Wave 2 data collection process, these reliability coefficients will be computed holding each analyst out. If an analyst shows a pattern of this type of reliability coefficient improving when his/her ratings were held out, we will remove that analyst's ratings from the analysis. It is also important to note that this reliability is dependent on the sample of occupations being rated. That is, all else being equal, this $ICC(3, k)$ based on ratings of a sample of homogeneous occupations will be lower than this $ICC(3, k)$ based on ratings of a sample of

heterogeneous occupations. It is important to keep this point in mind when interpreting the reliability results for the first 100 occupations and subsequent sets of occupations.

Summary

The main findings of the analysis of Wave 1 analyst ratings were as follows:

- The not-relevance and suppression criteria did not generate any results reflecting poorly on the quality of Wave 1 ratings.
- 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.
- All ICCs 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 Wave 1 did not vary greatly from one occupation to the next.

Given these results, it appears as though the analysts were well trained and generally understand the abilities and associated definitions. Agreement was high and there's clear evidence regarding the quality of the data. Nevertheless, it may be beneficial to closely examine the definitions, as well as the training, associated with the abilities that were flagged more often than others for having a $SE_M > .51$ for level (e.g., wrist-finger speed, far vision, flexibility of closure). It's possible that additional clarification could reduce the observed variance for those abilities.

References

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