FR10-03

Human Resources Research Organization

O*NET® Analyst Occupational Skills Ratings: Cycles 1 - 10 Results

Suzanne Tsacoumis
Shannon Willison
Laurie E. Wasko

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 2010

66 Canal Center Plaza, Suite 700 • Alexandria, VA 22314 • 703.549.3611 • fax. 703.549.9025
www.humrro.org

Executive Summary

The Occupational Information Network (O*NET®) is a comprehensive system developed by the U.S. Department of Labor that provides information for 965 occupations within the U.S. economy. In order 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 (GWA), knowledge, education and training, work styles, and work context. Importance and level information regarding the abilities and skills associated with each occupation are collected from occupational analysts. This report presents the results of the occupational analyst ratings for the 35 skills in the O*NET content model for 831 unique occupations.

To evaluate the ratings that occupational analysts provided, we performed three sets of analyses focused on computing measures of interrater agreement and interrater reliability. To evaluate interrater agreement, we calculated the standard deviation (*SD*) of ratings across analysts for a given construct and scale for each occupation, as well as standard error of the mean (SE_M) of these ratings. There was strong interrater agreement for all cycles. To examine the interrater reliability of each cycle's ratings, we calculated the intraclass correlations (ICC [3, *k*]; Shrout & Fleiss, 1979) among analysts' ratings to look at consistency *across constructs within occupations* and consistency *across occupations within constructs*. High levels of interrater reliability indicated that the occupational analysts rank ordered the *skills within each occupation* similarly as well as the *occupations within each skill* similarly on both importance and level.

Additional analyses raised a concern that a disproportionate number of occupations received "not relevant" ratings for two skills: Science and Operations Analysis. After supplementary targeted training, occupational analysts re-rated the two skills for all 831 occupations. The revised ratings were more consistent with general expectations and retained high reliability estimates.

Finally, three distinct criteria were established to flag the skill data that affected the presentation of data within the publicly available O*NET Online. First, the level rating of a skill was flagged if it was not relevant for a particular occupation. The percentage of skill ratings flagged because the skill was considered not important for performance of that occupation was relatively low (14.10%). The remaining two criteria involve the recommended suppression of any importance or level mean rating that had a standard error of the mean (SE_M) greater than .51. Across all cycles, there were a small number of skills (3.14%) flagged for level ratings and an even smaller number of skills (0.03%) flagged for importance ratings.

Overall these results provide clear evidence to the quality of the data. Review training for experienced occupational analysts and, if required, new occupational analyst training will continue to occur prior to each new cycle.

O*NET ANALYST OCCUPATIONAL SKILLS RATINGS: CYCLES 1 - 10 RESULTS

Table of Contents

List of Tables

O*NET ANALYST OCCUPATIONAL SKILLS RATINGS:
CYCLES 1 – 10 RESULTS

Introduction

The Occupational Information Network (O*NET®) is a comprehensive system developed by the U.S. Department of Labor that provides information for 965 occupations within the U.S. economy. This information is maintained in a comprehensive database which was developed to replace the Dictionary of Occupational Titles (DOT) (U.S. Department of Labor, 1991). In order 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 (GWA), knowledge, education and training, work styles, and work context. Importance and level information regarding the abilities and skills associated with each occupation are collected from occupational analysts.

There are theoretical and 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, and thus are the best source of information regarding tasks and GWAs. In contrast, skills require inferences that are more abstract in nature, so trained occupational analysts are likely to understand skill constructs better than incumbents (Morgeson and Campion, 1997; Tsacoumis, 2007) and therefore it follows that they should provide these data. In these instances, it is imperative that the occupational analysts have detailed occupation information in order to rate the skill constructs. It has also been suggested that some incumbents deliberately inflate their ratings to influence policy decisions such as those associated with compensation and training, whereas analysts are less prone to making these types of errors (Harvey, 1991; Moregeson & Campion, 1997; Morgeson, Delaney-Klinger, Mayfield, Ferrara, & Campion, 2004). Skill ratings may be particularly vulnerable to such effects given that they are more abstract and thus more difficult to verify than more observable descriptors such as job tasks (Morgeson & Campion, 1997; Morgeson et al., 2004). Given these considerations, occupational analysts as opposed to incumbents provided the skill information in the O*NET database.

This report presents the results of the analyst ratings for the 35 skills in the O*NET content model (http://www.onetcenter.org/content.html) for 831 unique occupations. Skills reflect proficiencies that are developed through training or experience. These are grouped into seven categories: content, process, social, complex problem solving, technical, systems, and resource management. To facilitate the skills rating process, analysts were provided relevant occupational information from incumbents. Eight trained analysts were responsible for rating the importance and level of the 35 skills for each of the O*NET occupations. 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 skill ratings, see *O*NET Analyst Occupational Skill Ratings: Procedures* (Willison & Tsacoumis, 2010).

To ensure a controlled data collection and management process, occupational data is collected in groups or “analysis cycles.” This report describes the results of the skills data collection process for the first ten cycles. Note that many occupations were repeated in two different cycles, but only the most recent data are provided in this report. For example, 53-6041.00 Traffic Technicians was rated in Cycles 5 and 9, but only Cycle 9 data for this occupation are included in the report. 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. The O*NET-SOC Codes and Titles included in O*NET Analysis Cycles 1 through 10 are presented in Appendix A.

Evaluation of Cycles 1 - 10 Analyst Ratings

As mentioned above, analysts provided ratings on importance and level of the 35 skills for each of the 831 unique occupations in Cycles 1-10. The mean, standard deviation, and SE_M of the importance and level ratings were computed. These results are presented in Appendix B.

To evaluate the ratings that analysts provided, we performed three sets of analyses focused on computing measures of interrater agreement and interrater reliability. Poor agreement or reliability estimates might indicate confusion about the constructs, potentially due to either the nature of the construct definition or to rater training. Specifically, the second analysis involved computing the interrater agreement among the eight analysts for a given skill. Next, the interrater reliability of the raters was computed to determine the similarity of the ratings with regard to the order of and relative distance among *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 rated the 35 skill constructs with regard to the (a) level of skill required to perform the occupation, or (b) relative importance of the skill to performance in a particular occupation. The analyses of the skill constructs were computed separately. Finally, another interrater reliability estimate was computed to examine the consistency of ratings across occupations within constructs. This estimate of interrater reliability was computed for each skill to determine the consistency with which raters rank-ordered occupations with regard to the importance/level of a given skill.

Cycles 1 - 10 Interrater Agreement

For each cycle, interrater agreement was computed to examine the level of absolute agreement among analysts when rating the importance or level of a skill for a particular occupation. For example, this index identifies the extent to which eight raters provided the same rating regarding the level of *Persuasion* required to perform in the occupation of *Dental Hygienist*. Specifically, we calculated the standard deviation (*SD*) of ratings across analysts for a given construct and scale for each occupation, as well as standard error of the mean (SE_M) of these ratings. For both indices (*SD* and SE_M), lower values indicate higher agreement and higher values indicate lower agreement.

The agreement analysis results for each cycle are provided in Appendix C. The columns labeled "Mean of *Ms*" show the mean of the eight analysts mean importance and level ratings across the 35 skills for each occupation.[1] The columns labeled "Median of *SDs*" show the median of the *SD*s associated with each mean importance and level rating across the skills for each occupation. Finally, the columns labeled "Median of SE_Ms" show the median of the SE_Ms

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

associated with each mean importance and level rating across the skills for each occupation, once again, within each cycle.

Table 1 provides a summary of the agreement analysis across all skills and occupations within a cycle. As can be noted, there was more agreement for the importance ratings than the level ratings; however, results indicate that the analyst ratings were consistent for both scales.

Table 1. Summary of Interrater Agreement Statistics by Cycle

		Importance		Level	
Cycle	*N*	Median *SD*	Median SE_M	Median *SD*	Median SE_M
1	12	0.53	0.19	0.74	0.26
2	44	0.53	0.19	0.74	0.26
3	73	0.53	0.19	0.74	0.26
4	92	0.52	0.18	0.64	0.23
5	88	0.52	0.18	0.71	0.25
6	99	0.52	0.18	0.71	0.25
7	101	0.53	0.19	0.74	0.26
8	100	0.53	0.19	0.71	0.25
9	31	0.53	0.19	0.74	0.26
10	192	0.52	0.18	0.64	0.23

Note. N = Number of occupations per cycle.

When comparing across the ten cycles, agreement indices for importance ratings were similar, with median *SD*s equal to either .52 or .53 and median SE_Ms equal to either .18 or .19. The agreement indices for the level ratings across the ten cycles showed slightly more variability, with median *SD*s ranging from .64 to .74 and median SE_Ms ranging from .23 to .26. Of particular notice was the median *SD* of .64 for level ratings in Cycles 4 and 10; these values were the lowest median *SD*s for level from all of the cycles, indicating the greatest amount of agreement on the level ratings was achieved during Cycles 4 and 10. In Cycle 4, this may be due to the preponderance of education or teaching related occupations. Approximately 75% of the occupations in this cycle were in the Education, Training, and Library Job Family. In Cycle 10, most of the occupations (84%) were rated in previous cycles, almost half of which were in Cycle 9. The similarity among the occupations in Cycle 4 and the recent familiarity of the occupations in Cycle 10 may have contributed to the increased interrater agreement.

Cycles 1 - 10 Interrater Reliability: Across Constructs Within Occupations

To examine the interrater reliability of each cycle's ratings, we calculated the intraclass correlations (ICC [3, *k*]; Shrout & Fleiss, 1979) among 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 35 skills 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 in multiple contexts, including O*NET (e.g., Clement, Chauvot, Philipp, & Ambrose, 2003; McCloy, et al., 1999; Rase & Tognetti-Stuff, 1983).

The results of these analyses are shown in Appendix D. The data revealed high levels of interrater reliability across the occupations in each of the ten cycles. Table 2 provides a summary of this information. For example, the mean and median ICC across all 88 occupations in Cycle 5 was .95 for both importance and level. Similar ICCs were observed across all cycles, ranging from a mean ICC of .93 (e.g., for both importance and level in Cycles 7 and 9) to a mean ICC of .97 (for both importance and level ratings in Cycle 4).

Table 2. Summary of Interrater Reliability Across Constructs by Cycle

		Importance			Level		
Cycle	*N*	Mean	Median	SD	Mean	Median	SD
1	12	.95	.95	.01	.95	.95	.02
2	44	.95	.95	.02	.94	.94	.03
3	73	.94	.95	.03	.94	.94	.03
4	92	.97	.98	.02	.97	.97	.02
5	88	.95	.95	.02	.95	.95	.02
6	99	.94	.95	.03	.95	.95	.03
7	101	.93	.93	.03	.93	.92	.03
8	100	.93	.94	.03	.94	.94	.03
9	31	.93	.93	.02	.93	.93	.02
10	192	.95	.96	.03	.95	.96	.05

Note. N = Number of occupations per cycle.

Results indicate the reliability for the importance and level ratings well exceeded the median target coefficient value of .80 for all of the ten cycles. Variability around the mean was also similar between importance and level ratings across the ten cycles, although there were two cycles where there was slightly more variability around the mean in the level ratings than in the importance ratings (e.g., Cycles 2 and 10). Similar to the interrater agreement indices presented in the previous section, the highest level of interrater reliability for both importance and level ratings was reached in Cycle 4 (where the mean and median ICC was .97). Lastly, trends in the data show that overall, occupations with the lowest reliability coefficients for importance also tended to have the lowest values for level ratings. This may be due to the skip pattern which forces a "0" for level if the skill is rated not important. Overall, the results support a strong amount of agreement in the analysts' ratings.

Cycles 1 - 10 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 skill *Critical Thinking*? To make this evaluation, we calculated Shrout and Fleiss' (1979) *ICC*(3, *k)* for each construct on each scale (instead of for each occupation on each scale as described above). For example, each of the 35 skill 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 35 skill 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 multiple contexts, including O*NET (e.g., Clement, Chauvot, Philipp, & Ambrose, 2003; McCloy, et al., 1999; Rase & Tognetti-Stuff, 1983).

This reliability analysis was conducted for skills on all 831 occupations from Cycles 1 - 10 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. Overall, the median ICC(C,8) across the construct ratings for importance was .86

(*M* = .86, *SD* = .06) and for level was .89 (*M* = .88, *SD* = .05). This indicates that on the whole, the reliabilities achieved the target level. Reliability estimates for the majority of the skills were in the .80s, and six skills had reliabilities equal to or greater than .90 for both importance and level. However, there are some low reliabilities to note.

Table 3. Interrater Reliabilities and Standard Errors of Measurement Across Occupations

		Cycles 1-10 (*N* = 831)					
		Importance			Level		
	Skill	ICC(C,1)	ICC(C,8)	s_E	ICC(C,1)	ICC(C,8)	s_E
1	Reading Comprehension	0.47	0.88	0.19	0.61	0.93	0.23
2	Active Listening	0.40	0.84	0.18	0.51	0.89	0.21
3	Writing	0.51	0.89	0.19	0.63	0.93	0.22
4	Speaking	0.51	0.89	0.17	0.56	0.91	0.21
5	Mathematics	0.42	0.85	0.22	0.52	0.90	0.31
6	Science	0.54	0.90	0.17	0.52	0.89	0.30
7	Critical Thinking	0.40	0.84	0.17	0.47	0.87	0.22
8	Active Learning	0.35	0.81	0.21	0.49	0.88	0.27
9	Learning Strategies	0.45	0.87	0.21	0.56	0.91	0.26
10	Monitoring	0.30	0.77	0.18	0.39	0.84	0.24
11	Social Perceptiveness	0.38	0.83	0.19	0.42	0.86	0.25
12	Coordination	0.33	0.79	0.17	0.31	0.78	0.25
13	Persuasion	0.35	0.81	0.21	0.34	0.81	0.31
14	Negotiation	0.39	0.83	0.20	0.38	0.83	0.28
15	Instructing	0.52	0.90	0.19	0.48	0.88	0.27
16	Service Orientation	0.43	0.86	0.20	0.37	0.82	0.27
17	Complex Problem Solving	0.34	0.81	0.21	0.46	0.87	0.23
18	Operations Analysis	0.25	0.73	0.20	0.28	0.76	0.34
19	Technology Design	0.28	0.75	0.22	0.36	0.82	0.33
20	Equipment Selection	0.57	0.91	0.21	0.56	0.91	0.31
21	Installation	0.46	0.87	0.17	0.45	0.87	0.26
22	Programming	0.43	0.86	0.20	0.51	0.89	0.28
23	Quality Control Analysis	0.50	0.89	0.25	0.53	0.90	0.34
24	Operations Monitoring	0.62	0.93	0.22	0.60	0.92	0.29
25	Operation and Control	0.70	0.95	0.21	0.71	0.95	0.28
26	Equipment Maintenance	0.77	0.96	0.17	0.80	0.97	0.22
27	Troubleshooting	0.65	0.94	0.20	0.68	0.94	0.28
28	Repairing	0.79	0.97	0.16	0.82	0.97	0.21
29	Systems Analysis	0.45	0.87	0.22	0.54	0.91	0.29
30	Systems Evaluation	0.38	0.83	0.23	0.49	0.88	0.33
31	Judg. and Dec. Making	0.36	0.82	0.18	0.49	0.89	0.22
32	Time Management	0.31	0.78	0.17	0.38	0.83	0.22
33	M. of Financial Resources	0.45	0.87	0.21	0.51	0.89	0.33
34	M. of Material Resources	0.36	0.81	0.22	0.39	0.84	0.34
35	M. of Personnel Resources	0.43	0.86	0.20	0.44	0.86	0.27

Note. These ICCs indicate how consistently raters rated (rank ordered) occupations on a given skill.
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) was found for Operations Analysis (.73). This skill also had the lowest level ICC(C,8) as well (.76). Besides this skill, only four other skills had importance reliabilities less than .80 and one other skill had a level reliability less than .80.

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 (.86 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.

Additional Considerations

The reliability results provide strong evidence that there was a high level of agreement and consistency among the analysts' ratings; however, one additional investigation was necessary to evaluate the data. Concurrent with the occupational analyst cycles, skill rating data were also collected from job incumbents; it was the incumbent skill ratings that were initially published on O*NET Online. For reasons stated in the introduction, skill information in the online database was updated with ratings from occupational analysts beginning in 2008. Therefore, a comparison between occupational analyst skill ratings and incumbent skill ratings on the same occupations is pertinent. Ratings from the two sources were not expected to match exactly but the comparison did raise concerns for two skills: Science and Operations Analysis. These constructs had the largest disparities between the two rating sources and also seemed inconsistent with general expectations. For example, based on the occupational analyst ratings, Science was considered relevant for approximately 20% of the occupations in the US economy. This figure seemed low considering that science is a fundamental curriculum within the US education system.

In response to these concerns, the occupational analysts were re-trained to ensure that the definitions of Science and Operations Analysis were interpreted in a manner consistent with the Department of Labor's original intent. After training, analysts re-rated the two constructs for all 831 occupations in Cycles 1-10. These ratings replaced the original ratings in the final database published on O*NET Online. The remainder of this report evaluates the final database that resulted from the Science and Operations Analysis re-rating process.

Revised Science and Operations Analysis Ratings

The revised ratings were more consistent with general expectations for both constructs. For example, Science was considered relevant for 71% of the 831 occupations. Also, disparities between the incumbent and occupational analyst ratings for Science and Operations Analysis were more consistent with the differences found between the two rater groups for other skills.

To ensure that the new ratings for the revised Science and Operations Analysis skills were consistent with the previously high reliability estimates we looked at the consistency across occupations within these two constructs. Similar to the evaluation above, we calculated Shrout and Fleiss' (1979) *ICC*(3, *k)* for each construct on each scale. Again, 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 is .80 or greater.

Table 4. Interrater Reliabilities and Standard Errors of Measurement Across Occupations for Revised Science and Operations Analysis

		Revised Science and Operations Analysis (*N* = 831)					
		Importance			Level		
	Skill	ICC(C,1)	ICC(C,8)	s_E	ICC(C,1)	ICC(C,8)	s_E
6	Science	0.68	0.94	0.21	0.67	0.94	0.33
18	Operations Analysis	0.41	0.85	0.26	0.48	0.88	0.39

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

These results suggest that there was a good level of agreement among the raters, with the reliabilities exceeding the target of .80. Note that the ICCs for the revised skills tend to be higher than the ICCs for the original ratings presented earlier in this report, but the SE_Ms for the revised ratings also tend to be comparable or higher than the SE_Ms for the original ratings. All else being equal, we wouldn't expect this to happen; higher reliability should lead to lower SE_Ms. There appears to be more observed variance across occupations in the revised ratings compared to original ratings. This likely occurred because many of the original ratings were 1, 0 (importance rating of 1 forces a level rating of zero), whereas for the revised skill ratings, most occupations received an importance of 2 or higher. The higher importance ratings allowed for more variability in the ratings overall, particularly in level.

Cycles 1 - 10 Recommended Data Flags

The final section of this report includes analyses for identifying recommended data flags for the final published data. Three distinct criteria were established to flag the skill data. All three flags affect the presentation of data within the publicly available O*NET Online (online.onetcenter.org). First, the level rating of a skill 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 a skill is considered not relevant when that construct is not important for the performance of the particular occupation. For example, the level ratings for Mathematics were considered not relevant for Actors (27-1011.00) as well as Massage Therapists (31-9011.00) because Mathematics was not considered important for the performance of these two occupations. Across all 831 occupations, there were 4,104 instances where the level of a skill within a particular occupation was flagged as not relevant (see Table 4). To facilitate interpretation of these results, it should be noted that there are 29,085 sets of ratings (831 occupations x 35 skills) across all cycles. Given this, 14.10% (4,104/29,085) of the skill ratings were flagged as not relevant.

Table 5 shows the number of times the level of a skill was flagged as not relevant. As can be noted, Equipment Selection, Installation, Programming, Equipment Maintenance, and Repairing received the highest number of flags for skills overall. These results are not surprising given that most of these constructs represent fairly specific technical skills intuitively not required for many occupations.

Table 5. Number of Times Skill Level Flagged as Not Relevant

Element Name	Cycles 1-10
Reading Comprehension	0
Active Listening	0
Writing	0
Speaking	0
Mathematics	26
Science	243
Critical Thinking	0
Active Learning	0
Learning Strategies	18
Monitoring	0
Social Perceptiveness	0
Coordination	0
Persuasion	1
Negotiation	0
Instructing	14
Service Orientation	0
Complex Problem Solving	0
Operations Analysis	67
Technology Design	256
Equipment Selection	414
Installation	686
Programming	382
Quality Control Analysis	116
Operations Monitoring	74
Operation and Control	219
Equipment Maintenance	472
Troubleshooting	243
Repairing	494
Systems Analysis	36
Systems Evaluation	23
Judg. and Dec. Making	0
Time Management	0
M. of Financial Resources	200

Table Continued on Next Page

Element Name	Cycles 1-10
M. of Material Resources	115
M. of Personnel Resources	5
Total Flags out of all possible ratings	14.10% (4104/29085)

Note. Number of occupations = 831.

The remaining two criteria involve the recommended suppression of 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 $\mathrm{SE_M}$ 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 importance and level are presented in Table 6. As can be noted, there were very few instances where the mean importance rating was flagged for insufficient agreement; less than 1% of the mean importance ratings were flagged.

Table 6. Importance and Level Flags Due to Large SE_M in Cycles 1 - 10

Element Name	$\mathrm{SE_M}$ > .51 Importance	$\mathrm{SE_M}$ > .51 Level
Reading Comprehension	0	0
Active Listening	0	0
Writing	0	0
Speaking	0	0
Mathematics	0	25
Science	0	37
Critical Thinking	0	1
Active Learning	0	5
Learning Strategies	0	7
Monitoring	0	5
Social Perceptiveness	0	2
Coordination	0	4
Persuasion	0	21
Negotiation	0	14
Instructing	0	14
Service Orientation	0	7
Complex Problem Solving	0	3
Operations Analysis	0	110
Technology Design	0	88
Equipment Selection	1	66
Installation	2	80

Table Continued on Next Page

Element Name	$SE_M > .51$ Importance	$SE_M > .51$ Level
Programming	0	81
Quality Control Analysis	3	95
Operations Monitoring	0	12
Operation and Control	0	18
Equipment Maintenance	2	19
Troubleshooting	0	26
Repairing	0	16
Systems Analysis	0	18
Systems Evaluation	0	38
Judg. and Dec. Making	0	1
Time Management	0	0
M. of Financial Resources	0	51
M. of Material Resources	0	40
M. of Personnel Resources	0	8
TOTAL	0.03% (8/29085)	3.14% (912/29085)

Note. Number of occupations = 831.

In comparison to the observed agreement among the importance ratings, there was a slightly higher percentage of level rating flags due to insufficient agreement. Across all cycles, there were 912 insufficient agreement flags for level ratings, which is 3.14% (912 divided by 29,085). The skills that were flagged the most for level criteria include: Operations Analysis, Quality Control Analysis, Technology Design, Programming, and Installation. These elements will be monitored in future cycles and discussed during analyst training to ensure the analysts interpret the constructs similarly.

Overall, these results provide strong evidence that there was a high level of agreement among the analysts when rating both importance and level of the 35 skills. Nevertheless, we will monitor the constructs that have the highest number of flags and discuss them when preparing analysts to make ratings in future cycles.

Summary

The main findings of the analysis skills ratings from Cycles 1 – 10 were as follows:

- There was strong interrater agreement for all cycles as evidenced by the overall low medians of SE_Ms.
- All within-occupation ICC reliabilities were well above the target value of .80. These high levels of interrater reliability indicate that the analysts rank ordered the skills within each occupation similarly on both importance and level.

- 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 skill similarly on both importance and level.
- Interrater reliability calculated on the revised skill ratings was similarly high (exceeding the target value of .80) as the reliability calculated on the ratings at the end of Cycle 10.
- The percentage of skill ratings flagged because the skill was considered not important for performance was relatively low (14.10%).
- Across all cycles, there were a small number of skills (3.14%) flagged for level ratings and an even smaller number of skills (0.03%) flagged for importance ratings based on a SE_M greater than .51.

Given these results, it appears as though the analysts were well trained and understand the skills 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.

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.

Harvey, R. J. (1991). Job analysis. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 71-163). Palo Alto, CA: Consulting Psychologists Press.

McCloy, R., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999). *Determining the occupational reinforcer patterns for O*NET occupational units*. Raleigh, NC: National Center for O*NET Development.

Morgeson, F. P., & Campion, M. A. (1997). Social and cognitive sources of potential inaccuracy in job analysis. *Journal of Applied Psychology, 82*, 627-655.

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.

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.

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.

Willison, S. & Tsacoumis, S. (2010). *O*NET analyst occupational skill ratings: Procedures*. (FR-08-70). Alexandria, VA: Human Resources Research Organization.

U.S. Department of Labor (1991). Dictionary of occupational titles. (Rev. 4th ed.). Washington, DC: U.S. Government Printing Office.