FR-10-29

O*NET® Analyst Occupational Skills Ratings: Analysis Cycle 11 Results

Suzanne Tsacoumis
Shannon Willison

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

Prepared under: **Contract Number: 2108-009**

December 2010

Human Resources Research Organization
66 Canal Center Plaza, Suite 700 • Alexandria, Virginia 22314
www.humrro.org • Phone: (703) 549-3611 - Fax: (703) 519-9661

ET ANALYST OCCUPATIONAL SKILLS RATINGS: ANALYSIS CYCLE 11 RESULTS

Table of Contents

List of Tables

O*NET ANALYST OCCUPATIONAL SKILLS RATINGS: ANALYSIS CYCLE 11 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 areas. Importance and level information regarding the abilities and skills associated with these occupations is being collected from occupational analysts. It should be noted that there are theoretical or philosophical reasons for preferring one rater group to the other for collecting different types of data. For example, incumbents are generally more familiar with the day-to-day duties of their job; therefore, they are the best source of information regarding tasks and GWAs. In contrast, it is likely that trained analysts understand the ability and skill constructs better than incumbents and therefore should provide the ability and skills data (Tsacoumis, 2007). Granted, it is imperative that the occupational analysts have detailed occupation information in order to rate the ability and skill constructs. It has also been suggested that some incumbents deliberately inflate their ratings to influence policy decisions regarding, for example, compensation and training (Harvey, 1991; 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 provide the ability and skill information in the O*NET database.

This report focuses on the skill results only. Skills reflect proficiencies that are developed through training or experience. The 35 O*NET skills cover performance applicable to a broad range of jobs in the world's economy. These are grouped into seven categories: content, process, social, complex problem solving, technical, systems, and resource management.

To facilitate the skill rating process, occupational analysts are provided relevant occupational information. Trained occupational analysts are responsible for rating the importance and level of the 35 skills for each of the O*NET occupations. More specifically, eight trained occupational analysts provided ratings for each occupation. For a description of the entire analyst data collection process, including the preparation and distribution of the occupational data, the steps associated with the ratings process, and the collection and management of the skill ratings, see *O*NET Analyst Occupational Skills Ratings: Procedures* (Willison & Tsacoumis, 2010).

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 eleventh analysis cycle of 120 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 11 are presented in Appendix A.

Evaluation of Cycle 11 Analyst Ratings

As mentioned above, occupational analysts provided ratings on importance and level of the 35 skills for each of the 120 occupations in Cycle 11. The mean, standard deviation, and SE_M of the importance and level ratings were computed. These results are shown in Appendix B.

Four sets of analyses were performed to evaluate the ratings that occupational analysts provided. First, we focused on identifying the data that may be difficult to interpret based on limited agreement among raters or because there is an indication that the skill level rating is not relevant for a specific occupation. Thus, a set of recommended suppression criteria was established which flagged: (a) a skill level rating as not relevant to an occupation because of low importance ratings, (b) a skill with too little agreement in importance ratings across raters for a particular occupation, and (c) a skill 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 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 35 skill 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 11 Recommended Data Flags

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 an skill was flagged as not relevant for a particular occupation if two or fewer of the eight occupational 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, in the Cycle 11 data, the level ratings for the Installation were considered not relevant for Advertising and Promotions Managers (11-2011.00) as well as Cooks, Restaurant (35-2014.00) because Installation was not considered important for the performance of these two occupations. In this cycle, there were 594 not relevant flags (see Table 1). To facilitate interpretation of these results, it should be noted that there are 4,200 sets of ratings (120 occupations x 35 skills) in the current cycle. Given this, 14.14% (594/4,200) of the skill ratings were flagged as not relevant.

Table 1 shows the number of not-relevant flags for skill level. The abilities with the most flags in Cycle 11 include Installation, Repairing, Equipment Maintenance, Equipment Selection, and Operations Analysis. With the exception of Operations Analysis, each of these abilities has received large numbers of flags in previous cycles. Given that these constructs capture fairly specific physical capabilities intuitively not required for many occupations, these results are not surprising. Nevertheless, the drastic increase in the number of flags associated with Operations Analysis is surprising. In addition, there was a reasonable increase in the flags for Science. When researching the analysts' ratings of the skills, we noticed that they seem to have adopted restrictive definitions for Operations Analysis and Science. As a result, we conducted an additional training session during which all analysts incorporated broader definitions of these constructs

The remaining two criteria involve the recommended suppression of identifying any skill 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. There were no instances in Cycle 11 where the mean importance rating was flagged for insufficient agreement. In fact, no importance ratings received flags for an SE_M greater than .51 since Cycle 3. The results of the suppression criteria for level for Cycles 1-11 are presented in Table 2. There were 182 insufficient agreement flags for level ratings. The number of flags indicating insufficient agreement with level ratings in Cycle 11 represents an increase in flags compared to the previous cycle, but it is not the highest percentage across all cycles.

In Cycle 11, the skills that were flagged the most for the level criteria were: Operations Analysis (n = 35), Technology Design (n = 22), and Programming (n = 20). This is consistent with previous cycles in that the same three skills received the most flags in Cycle 10. However, proportionally, a greater percentage of the occupations were flagged for these three constructs in Cycle 11 than in Cycle 10.

Overall, when looking at the number of occupations flagged to the total number of occupations in the cycle, 13 skills increased in the number of level flags from Cycle 10 to Cycle 11. Granted, some of these increases were small, such as the one flag for Management of Personnel Resources. The largest increase was in Operations Analysis with 29% (n = 35) of the occupations flagged in Cycle 11 compared to 17% in Cycle 10 (n = 33). There was also a fairly large increase in Science with 13% (n=16) of the occupations being flagged for level in Cycle 11 compared to 4% (n = 7) in Cycle 10. The percentage of level flags associated with the remaining 22 skills either remained the same as in Cycle 10 or decreased slightly.

Since Cycle 3, flags for skill level ratings with high SE_M s have remained under 5% for each Cycle. Although Cycle 11 showed a slight overall increase in the percentage of level flags compared to previous cycles, the findings suggest there remains a high level of agreement among the occupational analysts. Nevertheless, we will conduct additional training on the definitions of Operations Analysis and Science since these results suggest that the analysts may not have a uniform understanding of those constructs. 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. Number of Times Skill Level Flagged as Not Relevant

Element Name	Cycle 1 (N = 12)	Cycle 2 (N = 44)	Cycle 3 (N = 72)	Cycle 4 (N = 92)	Cycle 5 (N = 88)	Cycle 6 (N = 99)	Cycle 7 (N = 101)	Cycle 8 (N = 100)	Cycle 9 (N = 31)	Cycle 10 (N = 192)	Cycle 11 (N = 120)
Reading Comprehension	0	0	0	0	0	0	0	0	0	0	0
Active Listening	0	0	0	0	0	0	0	0	0	0	0
Writing	0	0	0	0	0	0	0	0	0	0	0
Speaking	0	0	0	0	0	0	0	0	0	0	0
Mathematics	0	4	3	1	3	4	2	8	0	1	0
Science	6	26	34	10	27	28	32	41	12	27	43
Critical Thinking	0	0	0	0	0	0	0	0	0	0	0
Active Learning	0	0	0	0	0	0	0	0	0	0	0
Learning Strategies	0	3	2	2	0	2	2	6	1	0	0
Monitoring	0	0	0	0	0	0	0	0	0	0	0
Social Perceptiveness	0	0	0	0	0	0	0	0	0	0	0
Coordination	0	0	0	0	0	0	0	0	0	0	0
Persuasion	0	1	0	0	0	0	0	0	0	0	0
Negotiation	0	0	0	0	0	0	0	0	0	0	0
Instructing	0	1	2	1	2	3	2	3	0	0	0
Service Orientation	0	0	0	0	0	0	0	0	0	0	0
Complex Problem Solving	0	0	0	0	0	0	0	0	0	0	0
Operations Analysis	0	6	12	1	7	7	7	14	6	7	62
Technology Design	2	28	38	53	38	39	17	21	6	14	16
Equipment Selection	6	29	37	67	47	47	32	28	8	113	67
Installation	11	40	59	87	74	77	69	77	25	167	100
Programming	3	35	50	59	48	64	40	52	17	14	25
Quality Control Analysis	1	6	6	30	19	10	4	10	0	30	15
Operations Monitoring	3	14	10	29	9	4	1	4	0	0	1
Operation and Control	3	10	14	56	29	21	13	11	3	59	38
Equipment Maintenance	10	33	36	70	56	53	34	34	14	132	78
Troubleshooting	7	25	27	49	28	23	12	13	1	58	31
Repairing	11	37	41	71	55	55	37	36	13	138	83
Systems Analysis	0	8	10	5	0	5	3	3	1	1	1
Systems Evaluation	0	6	9	1	0	2	0	4	1	0	0
Judg. and Dec. Making	0	0	0	0	0	0	0	0	0	0	0
Time Management	0	0	0	0	0	0	0	0	0	0	0
M. of Financial Resources	4	19	41	26	22	39	9	13	5	22	22
M. of Material Resources	2	9	23	8	13	16	11	12	2	19	12
M. of Personnel Resources	0	0	1	1	0	0	1	2	0	0	0
Total Flags out of all possible ratings	16.43% (69/420)	22.08% (340/1540)	18.06% (455/2520)	19.47% (627/3220)	15.49% (477/3080)	14.40% (499/3465)	9.28% (328/3535)	11.20% (392/3500)	10.60% (115/1085)	11.93% (802/6720)	14.14% (594/4200)

Table 2. Level Flags Due to Large SE_M

Element Name	Cycle 1 (N = 12)	Cycle 2 (N = 44)	Cycle 3 (N = 72)	Cycle 4 (N = 92)	Cycle 5 (N = 88)	Cycle 6 (N = 99)	Cycle 7 (N = 101)	Cycle 8 (N = 100)	Cycle 9 (N = 31)	Cycle 10 (N = 192)	Cycle 11 (N = 120)
Reading Comprehension	0	0	0	0	0	0	0	0	0	0	0
Active Listening	0	0	0	0	0	0	0	0	0	0	0
Writing	0	0	0	0	0	0	0	0	0	0	0
Speaking	0	0	0	0	0	0	0	0	0	0	0
Active Learning	0	0	0	1	1	1	1	1	0	0	0
Learning Strategies	0	1	6	0	0	0	0	0	0	0	0
Monitoring	0	0	3	0	1	1	0	0	0	0	0
Social Perceptiveness	0	1	1	0	0	0	0	0	0	0	0
Instructing	0	2	4	2	1	1	1	3	0	0	0
Complex Problem Solving	0	0	1	0	1	0	0	1	0	0	0
Judg. and Dec. Making	0	0	1	0	0	0	0	0	0	0	0
Time Management	0	0	0	0	0	0	0	0	0	0	0
M. of Personnel Resources	1	2	2	0	2	0	0	0	1	0	1
Service Orientation	1	3	2	0	0	1	0	0	0	0	3
Systems Evaluation	4	12	9	2	8	1	0	2	0	0	3
Persuasion	0	3	8	1	1	3	1	3	1	0	6
Critical Thinking	0	0	0	0	0	0	0	0	0	1	0
Coordination	0	0	1	0	0	0	1	1	0	1	0
Systems Analysis	0	5	3	3	2	2	1	1	0	1	2
Mathematics	1	1	1	7	2	7	3	2	0	1	5
Negotiation	0	1	4	0	0	0	1	6	0	2	0
Repairing	0	2	3	0	3	2	1	1	0	4	1
Operations Monitoring	0	2	1	0	2	2	1	0	0	4	3
Equipment Maintenance	0	2	3	0	1	4	1	1	1	6	2
M. of Material Resources	1	10	5	1	5	2	5	2	2	7	3
Operation and Control	0	4	0	0	1	2	3	0	1	7	5
M. of Financial Resources	2	1	5	19	6	5	1	3	2	7	7
Science	0	0	4	6	4	5	2	4	5	7	16
Troubleshooting	0	4	3	1	3	2	1	0	4	8	3
Installation	1	3	10	2	8	8	18	8	4	18	16
Equipment Selection	2	8	7	5	6	4	8	2	5	19	15
Quality Control Analysis	3	5	7	29	12	11	2	2	3	21	14
Technology Design	4	2	6	5	14	10	10	4	6	27	22
Programming	3	2	6	7	14	7	8	1	3	30	20
Operations Analysis	0	3	4	14	11	14	14	14	3	33	35
Total Flags out of all possible ratings	5.48% (23/420)	5.13% (79/1540)	4.37% (110/2520)	3.26% (105/3220)	3.54% (109/3080	2.74% (95/3465)	2.38% (84/3535)	1.77% (62/3500)	3.78% (41/1085)	3.04% (204/6720)	4.33% (182/4200)

Cycle 11 Interrater Agreement

Interrater agreement was computed to examine the level of absolute agreement among the occupational 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 skill *Reading Comprehension* required to perform a particular occupation. To look at agreement, we calculated the standard deviation (*SD*) of ratings across occupational analysts for a given construct and scale for each occupation and the SE_M of these ratings. For both indices, lower values indicate 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 occupational analyst mean importance and level ratings across the 35 skills 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 35 skills 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 35 skills 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 .64 and a median SE_M of .23. These results for importance and level reveal that raters agreed similarly in this cycle compared to previous cycles. Overall, while the values are generally greater (indicating less agreement) for level than they are for importance, the results indicate that the ratings made by the occupational analysts were consistent for both scales.

Cycle 11 Interrater Reliability: Across Constructs Within Occupations

To examine the interrater reliability of the Cycle 11 ratings we calculated the interclass correlations (ICC [3, *k*]; Shrout & Fleiss, 1979) among the occupational 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 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 presented in Appendix D. The data revealed high levels of interrater reliability across the 120 Cycle 11 occupations. Specifically, the mean and median ICC for importance ratings for the skills across the occupations was .94 and .95 (*SD* = .04), respectively. The mean and median ICC for the level ratings were also .94 and .95, respectively (*SD* = .04). The reliability for both the importance and level ratings exceeded the median target coefficient value of .80. Results also indicate that for the most part, 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 skill is rated not important. Overall, the results support a good level of agreement in the occupational analysts' ratings.

Cycle 11 Interrater Reliability: Across Occupations Within Constructs

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

Another effective way to evaluate the reliability of the occupational 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 *Reading 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 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 occupations in Cycles 1 through 11 and results are presented in Table 3. The reliability analyses are based on 951 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 for the skills, the median ICC(C,8) across the construct ratings for importance was .86 (M = .86, SD = .05) and for level was .89 (M = .88, SD = .05). This indicates that on the whole, the reliabilities achieved the target level. The majority of the skills had high ICC(C,8) reliabilities for both importance and level. In fact, there were 8 skill levels with reliabilities equal to or greater than .90 for both importance and level (e.g., Science). However, there are some low reliabilities to note.

Table 3. Interrater Reliabilities and Standard Errors of Measurement for Skills Across Occupations in Cycles 1 through 11

		Cycles 1-11 (N = 951)					
		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	0.38	0.83	0.18	0.51	0.89	0.21
3	Writing	0.52	0.90	0.19	0.64	0.93	0.22
4	Speaking	0.48	0.88	0.17	0.56	0.91	0.21
5	Mathematics	0.42	0.85	0.22	0.53	0.90	0.30
6	Science	0.68	0.94	0.22	0.67	0.94	0.34
7	Critical Thinking	0.40	0.84	0.17	0.47	0.88	0.22
8	Active Learning	0.36	0.82	0.21	0.49	0.89	0.27
9	Learning Strategies	0.46	0.87	0.21	0.56	0.91	0.26
10	Monitoring	0.29	0.77	0.18	0.40	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	0.33	0.80	0.21	0.33	0.80	0.31
14	Negotiation	0.38	0.83	0.20	0.36	0.82	0.28
15	Instructing	0.53	0.90	0.19	0.48	0.88	0.27
16	Service Orientation	0.43	0.86	0.20	0.35	0.81	0.27
17	Complex Problem Solving	0.34	0.80	0.21	0.47	0.88	0.23
18	Operations Analysis	0.44	0.86	0.26	0.49	0.89	0.39
19	Technology Design	0.28	0.76	0.22	0.35	0.81	0.34
20	Equipment Selection	0.57	0.91	0.21	0.56	0.91	0.32

		Cycles 1-11 (N = 951)					
		Importance			Level		
	Skill	ICC(C,1)	ICC(C,8)	s_E	ICC(C,1)	ICC(C,8)	s_E
21	Installation	0.49	0.88	0.17	0.48	0.88	0.26
22	Programming	0.42	0.85	0.20	0.50	0.89	0.29
23	Quality Control Analysis	0.49	0.89	0.25	0.53	0.90	0.34
24	Operations Monitoring	0.61	0.92	0.22	0.58	0.92	0.30
25	Operation and Control	0.69	0.95	0.21	0.69	0.95	0.28
26	Equipment Maintenance	0.77	0.96	0.16	0.80	0.97	0.22
27	Troubleshooting	0.64	0.93	0.20	0.67	0.94	0.29
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	0.90	0.29
30	Systems Evaluation	0.38	0.83	0.23	0.49	0.88	0.32
31	Judg. and Dec. Making	0.36	0.82	0.18	0.50	0.89	0.22
32	Time Management	0.29	0.77	0.17	0.39	0.84	0.21
33	M. of Financial Resources	0.43	0.86	0.21	0.49	0.89	0.34
34	M. of Material Resources	0.35	0.81	0.22	0.39	0.84	0.34
35	M. of Personnel Resources	0.42	0.85	0.20	0.43	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 skill ICC(C,8) reliabilities were found for Technology Design, Monitoring, Time Management, and Coordination, all of which had importance reliabilities under .80. The construct with the lowest level reliability was Coordination; this was the only skill that had a level reliability lower than .80. Even though these skills had the lowest reliabilities compared to other skills, the reliabilities were still considerably high; the lowest reliability for both importance and level was .76.

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.

Summary

The main findings of the analysis of Cycle 11 analyst ratings were as follows:

- About 14% of the skill ratings were flagged because the construct was considered not important for performance. This is an increase from the most recent cycles, but not the highest it has ever been (22%). The constructs that were flagged were very similar to those flagged in previous cycles and conceptually it is understandable that these constructs may be considered not-relevant for the given occupations.
- Zero importance ratings and about four percent of the level ratings were flagged based on a SE_M greater than .51. For the level ratings, this is higher than previous cycles but is not the highest it has ever been.

- There was strong interrater agreement for this cycle 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 occupational analysts rank ordered the skills within each occupation similarly on both importance and level.
- Index interrater reliability calculated at the end of Cycle 11 was high and 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 skill similarly on both importance and level.

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.

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.

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.