HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION

2025 No. 143

Updates to Occupational Interest Profiles and High-Point Codes for the O*NET Program Using the O*NET 30.0 Database

Technical Memorandum

Prepared for: National Center for O*NET Development
313 Chapanoke Road, Suite 130
Raleigh, NC 27603

Prepared under: Subcontract Number (through RTI International):
1-312-0207142-41224L

Authors: Anoop A. Javalagi
Jeffrey A. Dahlke

Date: January 14, 2026

Updates to Occupational Interest Profiles and High-Point Codes for the O*NET Program Using the O*NET 30.0 Database

Table of Contents

List of Tables

Updates to Occupational Interest Profiles and High-Point Codes for the O*NET Program Using the O*NET 30.0 Database

Introduction

The Occupational Information Network (O*NET) is a system developed by the U.S. Department of Labor that provides information on over 900 occupations covering more than 55,000 jobs within the U.S. economy. This information is maintained in a comprehensive database. At the time we began this research, the most current version was the O*NET 30.0 Database (National Center for O*NET Development, 2025). To keep the database current, the National Center for O*NET Development (hereafter referred to as "the Center") is involved in a continual data collection process aimed at identifying and maintaining current information on the characteristics of workers and occupations. For years, the Human Resources Research Organization (HumRRO) has supported the Center's efforts to maintain the database. This report focuses on work to update Occupational Interest Profiles (OIPs) and High-Point Codes (HPCs) for 891 data-level occupations in the O*NET-Standard Occupational Classification (O*NET-SOC) 2019 taxonomy (Gregory et al., 2019), including 871 (97.76%) occupations with updated text content in the O*NET 30.0 Database compared to the O*NET 27.3 Database.

The Center has a longstanding history of conducting research to populate vocational interest information for the occupations included in the taxonomy (Rounds et al., 1999; Rounds et al., 2008; Rounds et al., 2013). Vocational interest information is an important part of the O*NET Program's support of educational planning, career exploration, career guidance, job search, and organizational placement. The O*NET Content Model defines interest information compatible with Holland's (1997) RIASEC model of personality types and work environments (Realistic, Investigative, Artistic, Social, Enterprising, and Conventional). This model serves as a foundation for interest information in the O*NET System due to its extensive use in applied settings and research.

The Center's current methodology for identifying OIPs and HPCs was developed and reported by Putka et al. (2023). This approach facilitates regular updates to O*NET's interest information by leveraging supervised machine learning models. These models generate accurate RIASEC interest ratings based on predictive relationships between (a) O*NET-SOC text data (e.g., occupation titles, descriptions, and task statements) and (b) human ratings of RIASEC dimensions. The current pipeline involves processing updated occupational text, organizing it into structured embeddings, and feeding it into pre-trained ensemble models to generate new predictions. Although the approach is primarily based on machine learning predictions, it also provides an opportunity for an expert (Dr. James Rounds) to review a targeted subset of occupations and offer recommendations on final ratings and HPCs.

The last update to the OIP and HPC information occurred in 2023 using data from the O*NET 27.1 and the O*NET 27.3 Databases (Putka et al., 2023), and the updated information was released with the O*NET 28.1 Database. In the present research, we revisited occupational interests using data updated by the O*NET Data Collection Program and available in the O*NET 30.0 Database. This effort marks the first time the Center has updated the machine-generated OIP data since the initial development of the models. In the following sections, we describe our technical approach, present a comparison of updated predictions based on text from O*NET 30.0 against previously published data available in O*NET 28.1, and describe the high-level outcomes of our updates.

Technical Approach

We followed the analysis strategy for generating OIPs and HPCs described in Putka et al. (2023). Below we offer a brief overview of our technical approach. For more detailed information, please refer to Putka et al. (2023).

Generation of RIASEC Predictions

We generated updated RIASEC interest ratings by applying the pre-trained ensemble prediction models developed in the 2023 research to updated text data from the O*NET 30.0 Database. The process involved the following steps:

- **Data Preparation:** We extracted and cleaned relevant text data for 891 data-level occupations. Of these occupations, 20 showed no text changes from the O*NET 27.3 Database (the previous input data used to generate interest information for the O*NET 28.1 Database) and retained their existing database values, leaving 871 occupations requiring new data processing. The cleaned data included all 891 occupation titles, descriptions, and task statements.
- **Feature Generation:** We converted the text data into quantitative embeddings using the "nli-distilroberta-base-v2" Sentence-BERT (SBERT) model (Reimers & Gurevych, 2019). This process transformed the unstructured text into numerical vectors capturing the semantic meaning of the occupational information.
- **Model Application:** We fed the generated embeddings into the final ensemble prediction models established by Putka et al. (2023). These models output predicted ratings for each of the six RIASEC dimensions.
- **High-Point Code Assignment:** Based on the predicted ratings, we assigned up to three HPCs for each occupation following the procedure developed by Rounds et al. (1999).

Identification of O*NET-SOCs Requiring Review

To ensure the quality and defensibility of the model-generated data, we identified a subset of occupations for expert review using specific, data-driven flagging criteria. We compared the new model predictions (based on O*NET 30.0 inputs) to the previously published interest data (from O*NET 28.1). This comparison served two purposes: (a) to evaluate the overall stability of the interest estimates across the database, and (b) to identify a subset of occupations for expert review using flagging criteria. Occupations were flagged for review if they met any of the following conditions:

- **Low Profile Correlation:** The correlation between the new predicted RIASEC profile and the previously published profile was less than 0.95, indicating a significant shift in the overall interest pattern.
- **High-Point Code Mismatch:** The predicted first HPC differed from the published first HPC.
- **Artistic Interest Threshold:** The predicted "Artistic" interest was ranked second, and its rating was within 0.50 points of the top-ranked interest.

Based on these criteria, we identified 39 (4.38%) of the 891 O*NET-SOCs that required review.

Review Process

HumRRO constructed a review worksheet listing the following information for each O*NET-SOC requiring review:

- The target O*NET-SOC code and title.
- The specific flagging criteria triggered by the occupation (e.g., Low Profile Correlation, HPC Mismatch, and/or Artistic Interest Threshold).
- The current published RIASEC ratings and HPCs (from the O*NET 28.1 Database).
- The new model-predicted RIASEC ratings and HPCs.

The expert's task was to evaluate the predicted interest profiles against the published profiles and the occupational content (e.g., tasks and descriptions). For each flagged occupation, the expert determined whether the model-predicted ratings and HPCs were accurate or if they required adjustment. If the expert deemed the predictions unsuitable, they provided revised ratings and HPCs according to their review of the occupation's information.

Summary of Updates

Following a review of the occupational data and model predictions for the 39 flagged occupations, the expert recommended adjustments to the HPCs and/or RIASEC ratings for 11 occupations (28.21% of the flagged occupations, 1.23% of all occupations). To evaluate the extent of changes introduced by our updates, we compared the final OIPs and HPCs for the 871 updated occupations against the previously published data from the O*NET 28.1 Database.

The results indicate a high degree of stability in the interest profiles, with the mean and median profile correlations between the updated and previous ratings both exceeding 0.99. The updates also resulted in refined HPCs for a subset of occupations. The interest ratings for all 871 updated occupations were refreshed based on their current text and, of those occupations, 832 (95.52%) had a visible change in their published OIP profile scores that are rounded to two decimal places, 20 (2.30%) had a change in their primary (i.e., highest rated) HPC, and 90 (10.33%) had a shift in their top three interests.

Table 1 provides a summary of the expert's revisions to the machine-generated predictions. In all 11 cases, the expert adjusted the order of the HPCs to better align with the occupational content. To maintain consistency between the OIPs and the HPCs, the underlying RIASEC ratings were also adjusted to match the revised HPC order.

For the remaining 28 flagged occupations, the expert concluded that the model-generated predictions were accurate and did not require revision. Consequently, the final OIP and HPC dataset consists of new model-generated ratings for 860 occupations, expert-adjusted ratings for 11 occupations, and retained ratings for 20 occupations with no input text changes.

We also examined the magnitude of rating shifts across the six RIASEC dimensions for the 871 occupations with text-based changes that underwent updates. As shown in Table 2, the mean absolute differences were small, ranging from 0.05 to 0.11. While the average difference across all occupations was near zero, the range of differences indicates that specific occupations experienced changes of up to 0.76 in absolute magnitude.

*Table 1. O*NET-SOCs with Expert Revisions to OIPs and HPCs*

O*NET-SOC Code	Title	HPC Revisions
11-9121.00	Natural Sciences Managers	EIC → IEC
13-1071.00	Human Resources Specialists	CES → ECS
15-1254.00	Web Developers	IC → CI
17-1022.00	Surveyors	RCI → CRI
17-2121.00	Marine Engineers and Naval Architects	IRC → RIC
27-1012.00	Craft Artists	RA → AR
29-1181.00	Audiologists	SI → IS
29-1218.00	Obstetricians and Gynecologists	SIR → ISR
29-1229.06	Sports Medicine Physicians	SRI → ISR
41-9041.00	Telemarketers	CES → ECS
53-6051.01	Aviation Inspectors	RCI → CRI

Note. For all occupations listed, the expert recommended adjusting the order of the High-Point Codes (HPCs). The underlying model-predicted RIASEC ratings were reordered to correspond with the revised HPC order.

*Table 2. Summary of Differences between Updated and Previous (O*NET 28.1) RIASEC Ratings for Occupations with Input Text Changes*

RIASEC Dimension	Mean Difference (SD of Difference)	Mean Absolute Difference	Range of Differences (Min to Max)
Realistic	-0.01 (0.13)	0.08	-0.76 to 0.54
Investigative	0.00 (0.09)	0.06	-0.42 to 0.48
Artistic	0.00 (0.10)	0.05	-0.50 to 0.67
Social	0.00 (0.08)	0.05	-0.51 to 0.44
Enterprising	0.01 (0.09)	0.06	-0.36 to 0.39
Conventional	-0.01 (0.16)	0.11	-0.76 to 0.71

Note. N = 871 occupations. Differences calculated as New Rating minus Old Rating. Mean difference refers to the average difference observed across all 871 occupations. Mean of Absolute Difference refers to the average of absolute difference observed across all 871 occupations. Range (Min to Max) refers to the range of differences (i.e., minimum observed difference to maximum difference) observed across all 871 occupations.

Table 3 shows the descriptive statistics and the frequency and percentage of the primary HPC for each of the six RIASEC interest dimensions for the 871 occupations that had input text changes. The Realistic dimension exhibited the highest mean rating (M = 4.55, SD = 1.94) and was the most frequent primary HPC (41.45%). The Conventional dimension followed closely

with a high mean rating (*M* = 4.52) and the lowest variability (*SD* = 1.01), serving as the primary HPC for 18.03% of occupations. The Investigative (*M* = 3.35, *SD* = 1.69), Social (*M* = 2.99, *SD* = 1.78), and Enterprising (*M* = 2.79, *SD* = 1.60) dimensions showed moderate mean ratings, with Social being the most frequent primary HPC among these three (14.81%). Finally, the Artistic dimension displayed the lowest mean rating (*M* = 1.96, *SD* = 1.22) and was the least frequent primary HPC, accounting for only 3.21% of occupations.

Table 3. Descriptive Statistics for Final RIASEC Ratings and Primary High-Point Codes

RIASEC Dimension	Mean (SD)	Primary HPC Count (Primary HPC %)
Realistic	4.55 (1.94)	361 (41.45%)
Investigative	3.35 (1.69)	106 (12.17%)
Artistic	1.96 (1.22)	28 (3.21%)
Social	2.99 (1.78)	129 (14.81%)
Enterprising	2.79 (1.60)	90 (10.33%)
Conventional	4.52 (1.01)	157 (18.03%)

Note. *N* = 871 occupations. Primary HPC Count refers to the number of occupations where the specific RIASEC dimension was the highest-rated interest.

Conclusion

We updated the OIPs and HPCs for 891 data-level O*NET-SOCs based on text data from the O*NET 30.0 Database. This included updating 871 occupations based on new data from the O*NET 30.0 Database and retaining 20 occupations with unchanged text from the O*NET 27.3 Database (the most recent database that was used to produce machine-generated interest information). Our analysis confirmed that the updated profiles remained highly consistent with previous versions while incorporating necessary refinements based on the latest occupational data. This update included changes to the RIASEC ratings and/or HPCs for 11 of the 39 O*NET-SOCs reviewed by the expert. We compiled the resulting OIPs and HPCs for the final 891 occupations for inclusion in the O*NET 30.2 Database. For internal counting and tagging purposes, a separate list of the 871 occupations with updated information is provided in the final data file.

References

Gregory, C., Lewis, P., Frugoli, P., & Nallin, A. (2019). *Updating the O*NET®-SOC taxonomy: Incorporating the 2018 SOC structure – summary and implementation*. National Center for O*NET Development. https://www.onetcenter.org/reports/Taxonomy2019.html

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Psychological Assessment Resources.

National Center for O*NET Development. (2022). O*NET 27.1 database. https://www.onetcenter.org/dictionary/27.1/excel/

National Center for O*NET Development. (2023). O*NET 27.3 database. https://www.onetcenter.org/dictionary/27.3/excel/

National Center for O*NET Development. (2023). O*NET 28.1 database. https://www.onetcenter.org/dictionary/28.1/excel/

National Center for O*NET Development. (2025). O*NET 30.0 database. https://www.onetcenter.org/dictionary/30.0/excel/

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Putka, D. J., Dahlke, J. A., Burke, M. I. (2023). *Using machine learning to develop occupational interest profiles and high-point codes for the O*NET system*. National Center for O*NET Development. https://www.onetcenter.org/reports/ML_OIPs.html

Reimers, N., & Gurevych, I. (2019). *Sentence-Bert: Sentence embeddings using Siamese BERT-networks. ArXiv:1908.10084 [Cs]*. http://arxiv.org/abs/1908.10084

Rounds, J., Armstrong, P. I., Liao, H.-Y., Rivkin, D., & Lewis, P. (2008). *Second generation occupational interest profiles for the O*NET system: Summary*. National Center for O*NET Development. https://www.onetcenter.org/reports/SecondOVP_Summary.html

Rounds, J., Smith, T., Hubert, L., Lewis, P., & Rivkin, D. (1999). *Development of occupational interest profiles (OIPs) for the O*NET system*. National Center for O*NET Development. https://www.onetcenter.org/reports/OIP.html

Rounds, J., Su, R., Lewis, P., & Rivkin, D. (2013). *O*NET® Interest Profiler short form psychometric characteristics: Summary*. National Center for O*NET Development. https://www.onetcenter.org/reports/IPSF_Psychometric.html