Research Article – Extended

Personality Traits Systematically Explain the Semantic Arrangement of Occupational Preferences

Jumpei Yamashita[1,2], Ritsuko Iwai[2,3], Haruo Oishi[1], and Takatsune Kumada[2]

[1]NTT Access Network Service Systems Laboratories, Nippon Telegraph and Telephone Corporation, Musashino-shi, Tokyo, Japan
[2]Graduate School of Informatics, Kyoto University, Kyoto-shi, Kyoto, Japan
[3]Guardian Robot Project, RIKEN Information R&D and Strategy Headquarters (R-IH), RIKEN, Seika-cho, Kyoto, Japan

Abstract: Understanding occupational preferences through Big Five personality traits offers a crucial insight into the socio-psychological profiles of working individuals, extending beyond mere occupational behaviors. Previous research, however, has not conclusively shown that the broad, situation-general Big Five traits can systematically account for occupational preferences as outlined by the existing RIASEC model. The RIASEC framework's reliance on theory-driven, preselected occupational scenarios may hinder this explanation. In this study, we initially employed data-driven, exploratory methods to identify and validate occupational preference factors from thousands of participants' responses to a wide array of occupational titles. Subsequently, we explored the connections between the Big Five traits and these newly identified preference factors. Our analysis revealed a coherent and systematic relationship between data-driven occupational preferences and the Big Five traits, formulating the Hexagonal Openness–Extraversion–Agreeableness model of occupational personality traits. This model facilitates a broader understanding of individuals' work-related personalities from a comprehensive social-psychological viewpoint.

Keywords: occupational preferences, vocational interests, personality, Big Five, data-driven

Explaining the semantic arrangement of occupational preferences, or the RIASEC, known as "vocational interests" (Holland, 1997), by personality traits, or the Big Five (Goldberg, 1990; McCrae & Costa, 1987; McCrae & Costa, 2008), is the first step in describing the social and psychological personas of working people who have various occupational preferences. While studying occupational preferences has significant explanatory power for occupational behavior, such as suitability for jobs (Ehrhart & Makransky, 2007), its explanations regarding broad socio-psychological phenomena are inadequate. Similarly, the Big Five model can explain a broad spectrum of socio-psychological behaviors/statuses (Ang et al., 2006; De Hoogh et al., 2005; Gerber et al., 2011; Gutiérrez et al., 2005; Jensen-Campbell & Malcolm, 2007; Roccas et al., 2002), but it lacks explanatory power for occupational behavior (Ehrhart & Makransky, 2007). Bridging this gap will enhance our understanding of occupational behavior as a socio-psychological phenomenon. This approach could facilitate the placement of workers in suitable occupations by considering occupational factors and a range of everyday factors, including social or psychological habits among individuals within specific professions.

Optimizing congruence between people's occupational preferences and their actual occupations, or person–vocation fit, helps improve their occupational performance and job satisfaction (Hoff et al., 2020; Holland, 1997; Nye et al., 2012). One of the most influential methods of examining occupational preferences, the RIASEC, estimates those preferences mainly from a person's responses, such as "like" or "dislike," to occupational titles, such as researcher and artist (Holland, 1985; Holland et al., 1997). Occupational titles serve as distinctive "entities" within specific working situations that individuals may affiliate externally. The RIASEC initially segmented occupational preferences for these titles (entities) into six distinct types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (Table 1; see also Holland, 1997). Subsequent research has theoretically positioned these six preferences around a circle, following the order R–I–A–S–E–C, to reflect the semantic similarity in the characteristics of the occupations they represent (cf. "circumplex" models; Hogan, 1983;

Table 1. RIASEC occupational preference types based on Holland (1997)

	Preferring activities	Occupational examples
Realistic	Manipulating things, tools, and machines	Mechanics and carpenters
Investigative	Systematically and creatively conducting research	Scientists
Artistic	Creating works of art unsystematically	Artists
Social	Interacting with others for training or therapy	Teachers and counselors
Enterprising	Interacting with others to achieve the organization's or economic objectives	Sales people
Conventional	Routinely working following systematized procedures	Clerks

Holland, 1997). In contemporary measures of fit, the alignment between workers' preferences and actual occupations is measured along these distinct occupational characteristics "dimensions" (cf. The "congruence index"; Nye et al., 2012). A close correspondence between an individual's preferences and their occupational role in terms of these dimensions suggests a favorable person–vocation fit, whereas a notable discrepancy signifies a less optimal match.

Personality traits are people's distinctive internal characteristics of thoughts, feelings, and behaviors (Allport & Odbert, 1936). One of the most influential taxonomies of personality traits, the Big Five, describes people's coherent and stable everyday characteristics in terms of five traits identified by data-driven procedures: Openness to Experience (Openness), Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Goldberg, 1990; McCrae & Costa, 1987; McCrae & Costa, 2008). Contrary to situational "entities," the Big Five personality traits are internal "dimensions," where a combination of five distinct levels characterizes each individual's stable personality profile. The Big Five traits are derived from individuals' choices among adjectives that span five broad dimensions of personality, reflecting characteristics such as being "analytical" or "talkative" across various situations. The data-driven and situation-transcendent nature of the Big Five allows for the prediction of a wide array of socio-psychological behaviors/statuses, such as political ideology (Gerber et al., 2011), cultural intelligence (Ang et al., 2006), leadership (De Hoogh et al., 2005), interpersonal styles/relationships (Jensen-Campbell & Malcolm, 2007; McCrae & Costa, 1989), values (Roccas et al., 2002), and subjective well-being (Gutiérrez et al., 2005).

While several studies have investigated the relationship between the Big Five and RIASEC preferences (Larson et al., 2002), previous research has yet to show that the Big Five can systematically explain the circular arrangement of RIASEC occupational preferences. Meta-analysis has suggested only five substantial correlations (Larson et al., 2002): between the Openness and Investigative type (r = 0.28), Openness and Artistic type (r = 0.48), Extraversion and Social type (r = 0.31), Extraversion and Enterprising type (r = 0.41), and Agreeableness and Social type (r = 0.19). There are no substantial correlations between the Big Five and Realistic and Conventional types. This sparse pattern of correlations has led recent researchers to suggest that the Big Five and occupational preferences are distinct constructs (Hurtado et al., 2019).

Nevertheless, some missing links between the Big Five traits, which reflect broad everyday personalities, and the RIASEC preferences, which reflect specific occupational characteristics, may be due to the RIASEC's theoretical configuration (cf. Deng et al., 2007; Tracey & Rounds, 1995). When the RIASEC was constructed, the occupational titles (situational entities) to which participants would respond were selected a priori by experts based on theoretically assumed occupational characteristics (dimensions; Deng et al., 2007; Holland, 1997). Consequently, recent studies have shown that the RIASEC model has limited explanatory power beyond the original, restricted range of occupational titles (Deng et al., 2007). Other studies have suggested arbitrariness (degrees of freedom) in constructing the preference model (Tracey & Rounds, 1995). Indeed, research has indicated that certain preference types or factors, each associated with distinct characteristics (dimensions), may possess greater explanatory power than the RIASEC model (Tracey & Rounds, 1995). The theoretical underpinnings and occupation-specific focus of the RIASEC model may obscure its connections with the data-driven Big Five traits, which dimensionally capture a broader range of factors applicable beyond specific occupational contexts.

Our approach utilizes a data-driven, exploratory method to identify occupational preferences, thereby uncovering the broad, situation-general personality dimensions underlying the specific entities of occupational titles. This methodology, which we describe as an "entity-based, data-driven approach," aims to bridge the gap between specific situational entities like occupational titles and broad, situation-general dimensions of individual differences such as the Big Five traits – a connection that was overlooked in traditional approaches. This strategy is based on the premise that analyzing a wide array of psychological responses to a variety of occupational titles in society (i.e., diverse entities), rather than relying on a set of preselected occupational titles (i.e., preselected entities) that highlight

specific occupational characteristics (i.e., domain-specific situational dimensions), can reveal preference factors that are indicative of broad, data-driven dimensions of individual differences beyond particular situations, such as the Big Five personality traits. These dimensions, in turn, reflect factors relevant to occupational preferences and socio-psychological phenomena.

The implementation of this approach is straightforward. It involves selecting a broad range of occupational titles from various societal sectors without adhering to any preexisting hypotheses. We then gather extensive responses to these titles from diverse working individuals. The subsequent analysis uses statistical methods like exploratory factor analysis (EFA).

However, this approach is not without its challenges. One key issue is that different individuals' subjective understanding of these roles can influence the variation in responses across a broad spectrum of occupational titles. This variability, which does not necessarily indicate preference differences, must be carefully considered in the analysis.

We may solve this problem by using crowdsourcing to distribute occupational titles to many participants and selecting only titles most participants can conceptualize. Methods such as factor analysis and crowdsourcing, which are commonplace today, were not standard when the RIASEC model was initially constructed (cf. the historical roots of the RIASEC inventory in Strong, 1935); however, recent attempts to supplement the RIASEC inventory with missing occupational titles (Tracey & Rounds, 1996; Tracey, 2002) have not fully adopted these data-driven procedures.

This study explored the extent to which the Big Five personality traits can account for the configuration of an entity-based, data-driven model of occupational preferences. Our investigation is structured around four objectives, with Study I addressing the first objective and Study II covering the subsequent three. The initial objectives aim to (1) extract data-driven occupational preferences through exploratory analysis and (2) validate these preferences in a confirmatory manner (Analysis for Objectives 1 and 2). The latter objectives focus on (3) exploring the relationship between the Big Five and the extracted occupational preferences and (4) validating this relationship confirmatively (Analysis for Objectives 3 and 4). Consistent with the enduring view that occupational preferences are manifestations of personality (Holland, 1997; Larson et al., 2002), we hypothesize that the Big Five may systematically elucidate the circular organization of entity-based, data-driven occupational preferences, albeit with potentially modest effects. Such elucidation could harmonize diverse workers and occupations by considering a variety of socio-psychological behaviors and statuses in daily life – elements that may have been previously overlooked in assessing person–vocation fit using specific occupational characteristics.

Study 1

In Study 1, the analysis for Objective 1 explored entity-based, data-driven occupational preferences. Specifically, we collected 3,024 working people's responses to 247 occupational titles. Because the possible variations in subjective understanding of occupational titles may be problematic for our purpose, we selected items (occupational titles) that most participants successfully conceptualized through crowdsourcing. We then conducted a factor analysis on these responses.

Method

Participants

Participants responded to an online survey administered by Cross Marketing Inc. in Japan in February 2021. In order to include a wide range of working people engaged in part-time to full-time occupations, we recruited participants who worked for three or more days per week. A total of 3,024 workers met the criterion and participated in Study 1. Approximately 380 men and 380 women were assigned to each of the four age categories (20–29, 30–39, 40–49, 50–59) for 3,024 participants – all received payment for their participation. Recruitment of participants and survey procedures were approved by the Ethics Committee, Graduate School of Informatics of Kyoto University (KUIS–EAR–2019–005). Participants provided consent to participate in the study as part of their survey responses. Note that a standard EFA criterion for the appropriate sample size is a subject-to-variable ratio of 4:1 or 5:1 (Streiner, 1994; Floyd & Widaman, 1995). In the current study, the number of participants was 12 times the number of items (3,024 participants for 247 items), which is considered "good" (Comrey & Lee, 2013).

Measures

The questionnaire consisted of 247 occupational items, which were selected as follows. We consulted the Fourth Revised Classification of Occupations, edited by the Ministry of Health, Labor and Welfare (The Japan Institute for Labour Policy and Training, 2011), as the source of a recent comprehensive list of occupations in Japan. We excluded 44 subdivisions (occupational titles), such as "others in this category," from the list of 892 titles, and we then selected titles that a wide variety of workers in contemporary Japan could conceptualize. To accomplish this, we engaged Yahoo Crowdsourcing to conduct a preliminary

survey in which participants were asked to answer "yes" or "no" to the question "Can you conceptualize what this job entails?" for 16 occupational titles randomly selected from 848 titles. The preliminary survey was repeated for all occupational titles until five participants in each of two age conditions (ranges 20–40 years and 41–60 years) and two gender conditions (male and female) had responded. A total of 427 items were extracted, for which three or more of the five participants in all four conditions responded that they could conceptualize the occupational content. From the 427 titles, two psychologists who had experience conducting questionnaire research but were unfamiliar with the items on the RIASEC-based questionnaires selected 247 occupations. The criteria were to select those occupations that all working people could conceptualize and the most representative one if several conceptually common occupations were included. Finally, we modified a few words that participants might find challenging to conceptualize even after the above selections.

In the primary survey, participants were asked whether they fit the 247 items, considering their personality and aptitude (cf. Holland, 1997). Because responding to 247 items was burdensome, we collected their responses using a 2-point scale (yes/no) instead of a statistically more desirable 7-point scale (cf. Matell & Jacoby, 1971).

Analysis for Objective 1: Exploratory Extraction of Data-Driven Occupational Preferences

We identified preference factors for occupational titles. For this purpose, we conducted EFA on participants' responses to these titles. Data were analyzed in the R environment (R Core Team, 2013). To determine whether the data for responses to the items were appropriate for factor analysis, we calculated the Kaiser–Meyer–Olkin (KMO) value and conducted Bartlett's Test of Sphericity using the EFAtools package (Steiner & Grieder, 2020). We also determined the optimal number of factors using the Kaiser criterion (eigenvalue greater than "one" rule), minimum average partial (MAP) procedure (Velicer, 1976), and Bayesian information criterion (BIC) (Schwarz, 1978).

EFA was performed using the Psych package (Revelle, 2017). We performed EFA on polychoric correlations because the response data were obtained on an ordinal 2-point scale (Holgado-Tello et al., 2010; Rhemtulla et al., 2012). We examined the internal structure of occupational preferences using unweighted least squares (ULS) with the Promax rotation method. The univariate distributions of occupational items were asymmetric because most responses did not indicate fit to the occupational title, resulting in a nonnormal distribution of the observed variables. Because ULS does not assume a normal distribution, it can robustly handle asymmetric data, as in the current study (Lorenzo-Seva & Ferrando, 2021). To alleviate the problem of matrix positive definitiveness, we smoothed the matrix by eigenvalue decomposition (Bock et al., 1988; Bollen & Long, 1993). Note that the strict satisfaction of positive definiteness is not critical for ULS (Bollen & Long, 1993). We evaluated the internal consistency of the items comprising the factors using Cronbach's α, which is appropriate for the binary responses in the present data; a value of 0.7 or higher was considered acceptable (Cortina, 1993).

If a factor included a typical occupation for a RIASEC type, we noted this type on the factor. We referred to a report by the Japan Institute for Labour Policy and Training (Matsumoto et al., 2012) to identify the top 10% of the most typical RIASEC types for each occupation in contemporary Japan.

Results

We obtained 2,796 valid samples after excluding 228 participants who indicated they were unfit for any occupation. Because all occupational items were indicated as a fit by at least one participant, no occupational items were excluded. Participants responded that they were fit for an average of 17.7 (SD = 15.6) of the 247 occupational items.

Analysis for Objective 1: Exploratory Extraction of Data-Driven Occupational Preferences

The KMO value (0.928) and Bartlett's Test of Sphericity (χ^2 = 219,784.1, df = 30381, $p < .001$) showed that the occupational items were appropriate for factor analysis. The Kaiser criterion, MAP, and BIC results suggested 15-, 15-, and 18-factor solutions, respectively. Following the common proposal of the MAP and Kaiser criteria, we selected a 15-factor solution.

Table 2 shows the results of the EFA. Because of the large number of items, we describe only the items with factor loadings of 0.4 or above (Williams et al., 2010). The 15 factors emerged with a cumulative variance explanation rate of 0.56 (see details in the Electronic Supplementary Material, ESM 1, Tables E1 and E2). The commonalities of all occupational items ranged from 0.31 to 0.76. Cronbach's α was > 0.7, that is, acceptable, for 11 factors, excluding the Helping, Writing, Cooking, and Ordinal noise factors. Note that Factor 15, named Ordinal noise, was semantically uninterpretable, but its factor loadings were in apparent descending order according to the order in which the items were presented to the participants.

Discussion

We extracted 14 data-driven occupational preference factors that were semantically valid as the preliminary

Table 2. Exploratory factor analysis results

Factor	Factor name	Number of items	Example occupations	RIASEC type
1	Mechanical	44	Designing, assembling, inspecting, replacing broken parts, and repairing various machines and equipment	R (partial)
2	Routine physical	34	Routine physical work, such as delivering mail and packages, reading electricity and gas meters, and cleaning buildings	C (partial)
3	Procedural	19	Work following procedures for processing raw materials and ingredients, cooking, and inspecting finished food products (e.g., retort pouch foods)	C (partial)
4	Servicing	20	Customer services, such as providing tourist information or selling products	S (partial) E (partial)
5	Artistic	18	Creating and designing paintings, posters, web pages, interior spaces, advertising texts, and theatrical productions	A
6	Intellectual	13	Occupations requiring a high level of professional knowledge, such as dealing with laws, setting company business policies, and providing medical diagnosis and treatment	I (partial) Others
7	Helping	12	Occupations assisting in the medical care and treatment of patients or caring for residents of care facilities or infants in hospitals, nursing homes, and kindergartens	S (partial)
8	Crafting	14	Occupations that make personal items, such as tailoring kimonos and clothes and making wooden or leather products by hand	R (partial) A (partial)
9	Physical	15	Occupations involving physical work, such as driving buses, trucks, trains, ships, and other vehicles	C (partial)
10	Clerical	11	Clerical work such as accepting administrative applications, presenting procedures necessary for applications, and collecting, organizing, and preparing application documents	C (partial)
11	Selling	3	Sales and marketing work	E (partial)
12	Writing	2	Interviewing, writing, or editing articles	E (partial)
13	Teaching	6	Educating and teaching students in elementary, junior high, and high schools	S (partial)
14	Cooking	4	Preparing, boiling, simmering, baking, or cooking food from ingredients and adding seasoning and serving style to dishes in a restaurant	Others
15	Ordinal noise	–	–	–

Objective 1. Because this was an exploratory investigation, it is necessary to evaluate the validity and reliability of the factor structure through confirmatory analysis (i.e., the preliminary Objective 2). Further, the data quality may be slightly suspect because the large number of occupational titles led us to avoid a 7-point response scale, and the Ordinal factor suggests that participants' engagement with the questionnaire decreased with each successive item. Because of these problems, the reliability score (Cronbach's α) of the items per factor may not be sufficient for Factor 7, Helping; Factor 12, Writing; and Factor 14, Cooking.

Study 2

In Study 2, the analysis of Objective 2 evaluated the validity and reliability of the entity-based, data-driven occupational preferences using the originally developed Occupational Personality Trait Inventory (OPTI). We expected that the reduced participant burden and increased sensitivity of the OPTI (i.e., fewer items and a 7-point scale) would ensure sufficient validity and reliability for assessing all the factors.

In our primary aim, the analysis for Objective 3 scrutinized how Big Five personality traits correlate with the OPTI preference factors. We hypothesized that the OPTI factors, identified through a data-centric approach, would lessen the emphasis on occupational specificity, thereby clarifying the broad, situation-general dimensions of the Big Five. Explaining how the OPTI configurations align with the Big Five could comprehensively characterize the socio-psychological profiles of workers, transcending mere occupational situations.

Subsequently, the investigation for Objective 4 introduced and substantiated a theoretical framework of occupational preferences as informed by the Big Five traits, specifically by creating the Hexagonal Openness–Extraversion–Agreeableness model for "occupational personality traits" (refer to Figure 1). This phase involved a detailed quantitative analysis of the model by calculating the conceptual distances between the traits of Openness, Extraversion, Agreeableness, and the OPTI factors. We

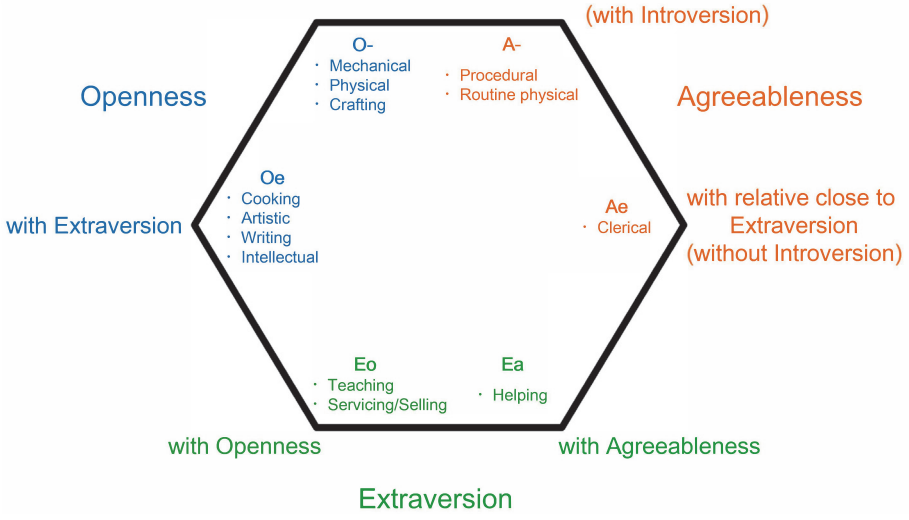

Figure 1. Hexagonal Openness-Extraversion-Agreeableness model of occupational personality traits (Hexagonal O-E-A model). The Openness (O) factor group includes occupations that intellectually, technically, or artistically handle concepts or things. Among them, occupations with relatively high interpersonal interaction are defined as the Openness with Extraversion (Oe) factor subgroup (i.e., Intellectual, Writing, Artistic, and Cooking factors). In contrast, the remaining are the Openness (O-) factor subgroups (i.e., Mechanical, Physical, and Crafting factors). The Extraversion (E) factor group includes socially stimulating occupations centered on interpersonal interactions. Those occupations that include a relatively intellectual or technical aspect are the Extraversion with Openness (Eo) factor subgroup (i.e., Teaching, Servicing/Selling factors). In contrast, those that include the relative caregiving aspect are the Extraversion with Agreeableness (Ea) factor subgroup (i.e., Helping factor). The Agreeableness (A) factor group includes occupations performed routinely according to instructions. Those occupations with relatively high interpersonal interaction are the Agreeableness with Extraversion (Ae) factor subgroup (i.e., Clerical factor), while the remaining are the Agreeableness (A-) factor subgroup (i.e., Procedural and Routine physical factors).

anticipated that the spatial arrangement of each factor within the model would conform to its theoretical positioning, thus validating our proposed framework.

Method

Participants

Workers (n = 4,166) participated in an online survey in Japan in February 2022. Approximately 520 men and 520 women were assigned to each of the four age categories (20–29, 30–39, 40–49, 50–59), for a total of 4,166 participants.

Measures

As a measurement inventory, we developed and used the OPTI, which consisted of only 68 items that most typically represent 13 preference factors (cf. Table 3). The OPTI asked participants to rate their fitness for each occupational item on a 7-point scale, considering their personality and aptitude (cf. Holland, 1997).

In developing the OPTI, we selected six items from those with factor loadings of 0.4 or above in Analysis for Objective 1 for each factor. We selected only three items from factors for which fewer than six items passed the above criterion. We made four exceptions in this item selection. First, Factor 10 (Selling) was combined with Factor 4 (Servicing) because Factor 4 (Servicing) already included salespeople. Second, Factor 11 (Writing) had only two items, so we added the title "magazine journalist," which had a factor loading of 0.37, to make three items. Third, we moved preschool teachers from Factor 12 (Teaching) to Factor 7 (Helping) to separate the semantic implications of Factors 7 and 12. Fourth, we only used five items for Factor 9 (Physical) because a technical error led to an incorrect item.

We also used the Trait Descriptors Personality Inventory (TDPI; Iwai et al., 2019) to estimate the degrees of the Big Five traits. The participants rated their degree of fit to 20 questions on a 7-point scale (four questions for each trait). Iwai et al. (2019) developed the TDPI to overcome the lack of Japanese instruments encompassing relatively few items and confirmed factor structures. The TDPI demonstrated a robust correlation with existing Japanese personality questionnaires, including the Ten-Item Personality Inventory (TIPI-J; r = .58–.68; Iwai et al., 2019). The Cronbach's α coefficients between questionnaire items ranged from .66 to .82 (Iwai et al., 2019). Subsequently, the structural validity of the five-factor model was confirmed in a different, relatively large sample of participants (n = 17,751; Iwai et al., 2018).

Analysis for Objective 2: Confirmatory Validation of Data-Driven Occupational Preferences

Analysis for Objective 2 aimed to evaluate the factor structure of the OPTI through confirmatory analysis.

Table 3. Confirmatory factor analysis matrix of occupational preference

	Factor loading	R^2	α	CR	AVE		Factor loading	R^2	α	CR	AVE		Factor loading	R^2	α	CR	AVE
Mechanical (R)						Artistic (A)						Physical (C)					
Bicycle repairer	0.91	.82	0.95	.95	.77	Graphic designer	0.84	.70	.92	.92	.66	Route bus driver	0.89	.80	.93	.93	.72
Aircraft assembler	0.94	.88				Web designer	0.85	.72				Train driver	0.92	.85			
Railcar repairer	0.93	.87				Painter, calligrapher	0.81	.66				Dump truck driver	0.85	.73			
Car mechanic	0.86	.74				Copywriter	0.87	.76				Truck driver	0.75	.56			
Mechanical design engineer	0.79	.63				Interior designer	0.74	.55				Small boat operator	0.81	.66			
Electrical engineer	0.81	.65				Producer	0.76	.58									
Routine physical (C)						Intellectual (I)						Clerical (C)					
Postman	0.80	.64	0.92	.92	.66	Lawyer	0.84	.70	.92	.92	.66	Post office counter clerk	0.80	.64	.89	.89	.57
Home-delivery service worker	0.88	.78				Prosecutor	0.87	.76				General affairs clerk	0.73	.54			
Street and park cleaner	0.83	.70				Company executive	0.81	.65				Bank counter clerk	0.86	.74			
Car/bicycle parking lot manager	0.74	.55				Member of parliament	0.81	.66				Medical clerk	0.74	.54			
Building cleaning staff	0.75	.56				Doctor	0.80	.64				Librarian	0.57	.33			
Water or gas meter reader	0.85	.72				College teacher	0.72	.52				Labor and social security attorney	0.81	.66			
Procedural (C)						Helping (S)						Writing (E)					
Frozen processed food cooking and manufacturing worker	0.86	.74	0.94	.94	.73	Nurse	0.81	.65	.90	.90	.59	Newspaper/magazine/book editor	0.91	.82	.91	.91	.78
Ice-cream cooking and manufacturing worker	0.87	.76				Babysitter	0.67	.45				Newspaper/broadcast reporter	0.93	.87			
Ham, bacon, sausage cooking and manufacturing worker	0.92	.84				Midwife	0.86	.73				Magazine reporter	0.80	.64			
Lunch box and side dish cooking and manufacturing worker	0.74	.55				Facility caregiver	0.73	.54									
Food inspector	0.81	.65				Dental hygienist	0.90	.81				Teaching (S)					
Boil-in-the-bag foods cooking and manufacturing worker	0.90	.81				Nursery school/kindergarten teacher	0.62	.38				High school teacher	0.96	.92	.94	.95	.86
Servicing/Selling (S/E)						Crafting (R/A)						Junior high school teacher	0.95	.90			
Tourist guide	0.67	.45	0.9	.90	.59	Kimono tailor	0.87	.76	.94	.94	.72	Elementary school teacher	0.86	.74			
Bus tour conductor	0.82	.67				Dyeing and finishing worker	0.87	.77									
Hotel front desk clerk	0.69	.47				Woodworker/woodcarver	0.86	.74				Cooking (–)					
Car salesperson	0.86	.74				Leather/leather goods manufacturer	0.78	.61				Chinese restaurant cook	0.92	.85	.94	.94	.84
Bartender	0.75	.56				Wooden furniture and fittings manufacturer	0.82	.68				Chef	0.92	.85			
Real-estate salesperson	0.81	.66				Clothes repairer	0.89	.78				Japanese restaurant cook	0.91	.83			

We conducted confirmatory factor analysis (CFA) on the responses to the OPTI. We then evaluated the CFA model's goodness of fit and the validity and reliability of the confirmatory factors.

CFA was performed using the Lavaan package (Rosseel, 2012). We used unweighted least squares with robust standard errors to estimate robust goodness-of-fit indices for CFA on nonnormal variables (Brosseau-Liard & Savalei, 2014). The robust version of goodness-of-fit indices included the comparative fit index (CFI), Tucker–Lewis nonnormed fit index (NNFI), and the root-mean-square error of approximation (RMSEA). The CFI and NNFI indices are considered good at 0.95 (Bryant & Yarnold, 1995), and the RMSEA is considered suitable at values below 0.06 (Hu & Bentler, 1999). The internal consistency of the items comprising each factor was tested with the criterion of ordinal $\alpha > .7$ (Zumbo et al., 2007). Convergent validity within each factor was tested with composite reliability (CR) > .6 (Bagozzi & Yi, 1988). Discriminant validity between factors was tested, at least between non-similar factors, by confirming that the average variance extracted (AVE) with a value of more than .5 was larger than the squared inter-factor correlation (Fornell & Larcker, 1981). We did not require strict discriminant validity between all factors because previous studies have shown that similar occupational preferences are often strongly correlated (Holland, 1997).

Validity is often tested by spatially arranging occupational preferences regarding their similarities (cf. the circular R–I–A–S–E–C arrangements; Holland, 1997; Tracey & Rounds, 1995). Thus, we also conducted multidimensional scaling (MDS; Kruskal & Wish, 1978) using the Python scikit-learn library (Pedregosa et al., 2011) to visualize the relationships among the 13 factors using a 13×13 inter-factor correlation in CFA (Mount et al., 2005). Using data that indicates the degree of similarity between variables (e.g., factor correlations), MDS attempts to locate points representing each of the variables (e.g., factors) in low-dimensional space so that the proximities between points represent similarities between them. To determine the optimal number of dimensions, we used Kruskal's stress index as a "badness-of-fit" index of MDS, with larger values indicating a poorer solution (Kruskal, 1964).

Analysis for Objective 3: Exploratory Investigation of the Relationship Between the Big Five and Data-Driven Occupational Preferences

Analysis for Objective 3 sought an arrangement of the OPTI factors by the Big Five traits. We conducted multiple linear regression (MLR) analysis on Big Five scores as the explanatory variables and OPTI scores as the objective variables. Using basic R functions, MLR was repeatedly performed on the Big Five and 13 OPTI factor scores (R Core Team, 2013). We analyzed the standardized multiple regression coefficients (b^*s) across 13 MLR models, each corresponding to an OPTI factor, to determine the independent influence of the Big Five personality traits on each OPTI factor. Given the large sample size, we focused only on significant b^*s at $p < .0001$. We then depicted the relationships of OPTI factors with the Big Five in the OPTI arrangement space (the MDS space).

Analysis for Objective 4: Confirmatory Validation of the Relationship Between the Big Five and Data-Driven Occupational Preferences

Based on the previous MLR analyses, we proposed the Hexagonal O–E–A model (cf. Figure 1), which posits that the OPTI factors are, at least in part, expressions of Big Five traits. This model results in the classification of the 13 OPTI factors into three groups, primarily associated with "Openness (O-)," "Extraversion (E-)", and "Agreeableness (A-)." Each group is further divided into two subgroups based on their secondary ties with the above three traits (i.e., "-o," "-e," and "-a"). These six preference factors are connected in a manner that the alphabetical codes of the subgroups match, resulting in a hexagonal arrangement model of occupational preferences based on the Big Five.

Analysis for Objective 4 evaluated the quantitative validity of the proposed model. We calculated the distances of the Openness, Extraversion, Agreeableness, and the 13 OPTI factors when the relationships between the two sets of factors (i.e., the three traits and the OPTI factors) were maximized (i.e., simulating the repeated MLR analyses). The adjacent OPTI factors in the model should be closer to each other than the others. Moreover, the distance between the Big Five and the OPTI factors should be as predicted in the model.

For this purpose, we conducted canonical correlation analysis (CCA) using the scikit-learn Python library (Pedregosa et al., 2011), following previous studies (Mount et al., 2005). CCA maps factors in the latent coordinate space where the correlation between two sets of factors is maximized. The current CCA generated two canonical variables: one linearly combining the three traits and the other linearly combining the OPTI factors. These two canonical variables were constructed to maximize the correlation between them (i.e., canonical correlation). These generations of canonical variables were repeated after eliminating the relationship once established; orthogonal canonical correlations were repeatedly generated. Considering the three canonical correlations as three orthogonal axes in the coordinate space, we calculated the distances between each factor (i.e., the three traits and the OPTI factors). Here, we first calculated the center coordinates of Openness, Extraversion, and Agreeableness traits. Regarding the OPTI factors, we calculated the center coordinates of multiple factors grouped into each point in our hexagonal model.

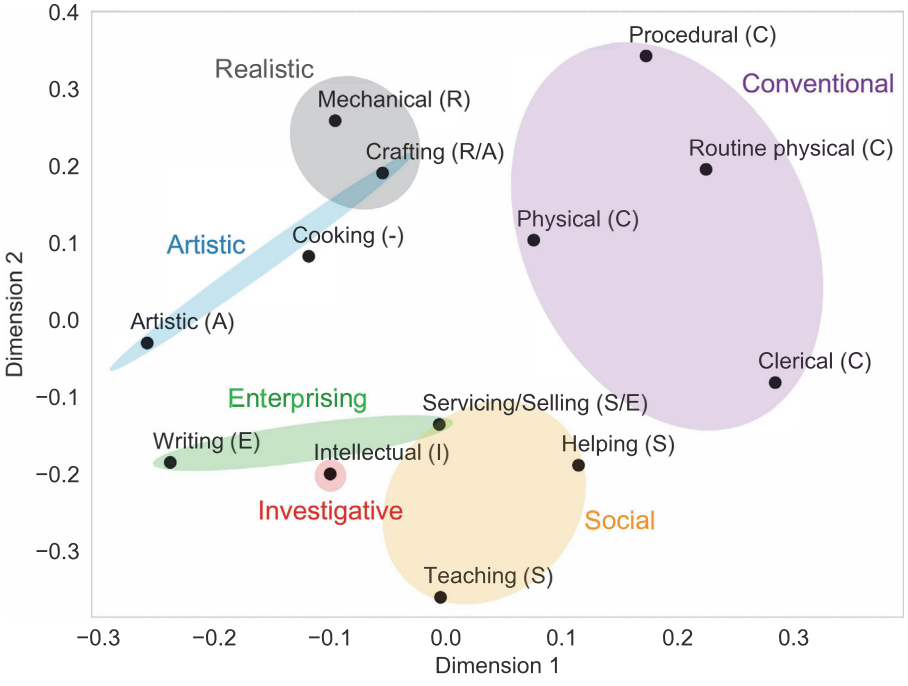

Figure 2. MDS results of OPTI factor correlation (RIASEC).

Results

Participants (n = 182) indicated they were fit for all occupations to the same degree. After these 182 participants were excluded, the data from 55 participants who responded to all the Big Five traits to the same degree were also excluded. Finally, we obtained 3,929 valid samples.

Analysis for Objective 2: Confirmatory Validation of Data-Driven Occupational Preferences

All the indices confirm the CFA model's goodness of fit. The KMO value (0.976) and Bartlett's Test of Sphericity (χ^2= 231,703.2, df = 2346, p < .001) showed that the occupational items were appropriate for factor analysis. The goodness-of-fit results were as follows: CFI (0.989) and NNFI (0.988) were greater than 0.95 (good), and RMSEA (0.05) was lower than 0.06 (good).

Overall, the results suggest the validity and reliability of the 13 OPTI factors. The results of the factor loading, the squared multiple correlation coefficient (R^2), α, CR, and AVE are shown in Table 3. Ordinal αs were > .7 for all factors, suggesting the internal consistency of items. CR for all items was greater than .6, suggesting appropriate convergent validity. Regarding the discriminant validity, AVE was > .5, and AVE for almost all factors was larger than the squared inter-factor correlations (see also ESM Table S3). Of the 78 comparisons, only three AVEs were not greater than the squared inter-factor correlations: the AVE of Servicing/Selling (S/E) was not greater than the squared inter-factor correlations between Servicing/Selling (S/E) and Intellectual (I) and between Servicing/Selling (S/E) and Helping (S); and the AVE of Helping (S) was not greater than the squared inter-factor correlation between Helping (S) and Servicing/Selling (S/E).

In the MDS analysis, we obtained stress indices of 3.23 for the one-, 0.38 for the two-, 0.12 for the three-, and 0.05 for the four-dimensional analyses on the inter-factor correlation. Because a significant stress reduction occurred when the model changed from one-dimensional to two-dimensional, the two-dimensional model was judged optimal (Kruskal, 1964). The results of the MDS are shown in Figure 2. Preferences for the same type in the RIASEC were generally close, that is, highly similar. On the other hand, the inter-type arrangements were not arranged circularly in the order R–I–A–S–E–C.

Analysis for Objective 3: Exploratory Investigation of the Relationship Between the Big Five and Data-Driven Occupational Preferences

We show the b^*s of the MLR models in Table 4. To facilitate the interpretation, we reordered the OPTI factors as follows. First, factors were divided into groups according to which Big Five traits had the largest absolute b^* value. There were three groups of factors, named the "Openness," "Extraversion," and "Agreeableness" factor groups. The OPTI factors were sorted based on each group's significant, second-largest absolute b^* value. Finally, the OPTI factors were sorted based on each group's significant, third-largest absolute b^* value. We further named subgroups by their positions within the group.

Table 4. Regression results of Big Five traits and OPTI factors

	With Extraversion									Openness			With Unconscientiousness								
	Intellectual (I)			Writing (E)			Artistic (A)			Cooking (–)			Crafting (R/A)			Physical (C)			Mechanical (R)		
	*b**	*SE*	*p*-value	*b**	*SE*	*p*-value	*b**	*SE*	*p*-value	*b**	*SE*	*p*-value	*b**	*SE*	*p*-value	*b**	*SE*	*p*-value	*b**	*SE*	*p*-value
Openness	**0.320**	0.016	<.0001	**0.295**	0.017	<.0001	**0.325**	0.017	<.0001	**0.196**	0.017	<.0001	**0.246**	0.017	<.0001	**0.200**	0.017	<.0001	**0.274**	0.017	<.0001
Conscientiousness	−0.035	0.016	.026	−0.059	0.016	<.001	−0.062	0.016	<.001	−0.049	0.017	.003	**−0.078**	0.017	<.0001	**−0.135**	0.017	<.0001	**−0.137**	0.017	<.0001
Extraversion	**0.140**	0.016	<.0001	**0.136**	0.017	<.0001	**0.088**	0.017	<.0001	**0.071**	0.018	<.0001	−0.061	0.018	<.001	0.045	0.017	.009	−0.059	0.017	<.001
Agreeableness	0.014	0.016	.38	0.001	0.016	.965	−0.002	0.016	.923	0.047	0.017	.005	0.059	0.017	<.001	0.049	0.017	.003	0.033	0.017	.045
Neuroticism	**−0.099**	0.015	<.0001	−0.048	0.016	.002	−0.035	0.016	.023	−0.039	0.016	.018	−0.028	0.016	.082	**−0.104**	0.016	<.0001	**−0.107**	0.016	<.0001
R^2	.169			.135			.131			.063			.060			.067			.078		

	With Openness			Extraversion			With Agreeableness				With Openness			Agreeableness			With Introversion		
	Teaching (S)			Servicing/Selling (S/E)			Helping (S)				Clerical (C)			Routine physical (C)			Procedural (C)		
	*b**	*SE*	*p*-value	*b**	*SE*	*p*-value	*b**	*SE*	*p*-value		*b**	*SE*	*p*-value	*b**	*SE*	*p*-value	*b**	*SE*	*p*-value
Openness	**0.183**	0.017	<.0001	**0.143**	0.017	<.0001	0.066	0.017	<.001	Openness	**0.148**	0.017	<.0001	**0.107**	0.018	<.0001	**0.077**	0.018	<.0001
Conscientiousness	−0.014	0.016	.369	−0.037	0.016	.02	0.004	0.016	.82	Conscientiousness	0.033	0.017	.051	**−0.097**	0.017	<.0001	−0.048	0.017	.004
Extraversion	**0.223**	0.017	<.0001	**0.246**	0.017	<.0001	**0.182**	0.017	<.0001	Extraversion	−0.025	0.018	.147	**−0.095**	0.018	<.0001	**−0.126**	0.018	<.0001
Agreeableness	**0.070**	0.016	<.0001	**0.068**	0.016	<.0001	**0.129**	0.017	<.0001	Agreeableness	**0.150**	0.017	<.0001	**0.141**	0.017	<.0001	**0.144**	0.017	<.0001
Neuroticism	−0.052	0.016	<.001	−0.044	0.016	.005	−0.015	0.016	.35	Neuroticism	0.031	0.016	.055	0.016	0.016	.323	0.036	0.017	.028
R^2	.141			.129			.085			R^2	.058			.038			.034		

Note. Bold numbers indicate b*s that reached a significance level of .0001.

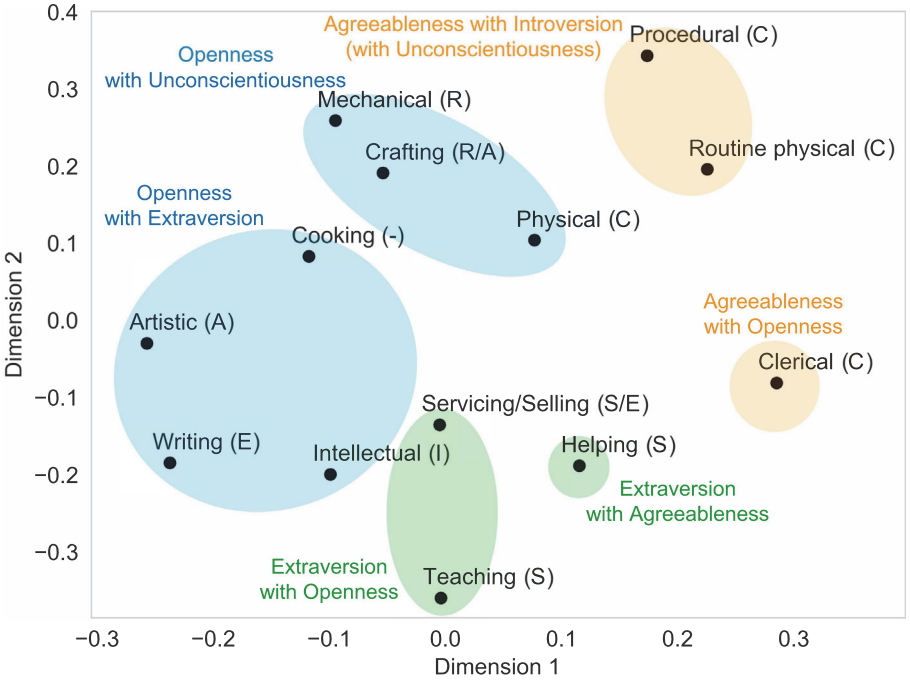

Figure 3. MDS results of OPTI factor correlation (Big Five).

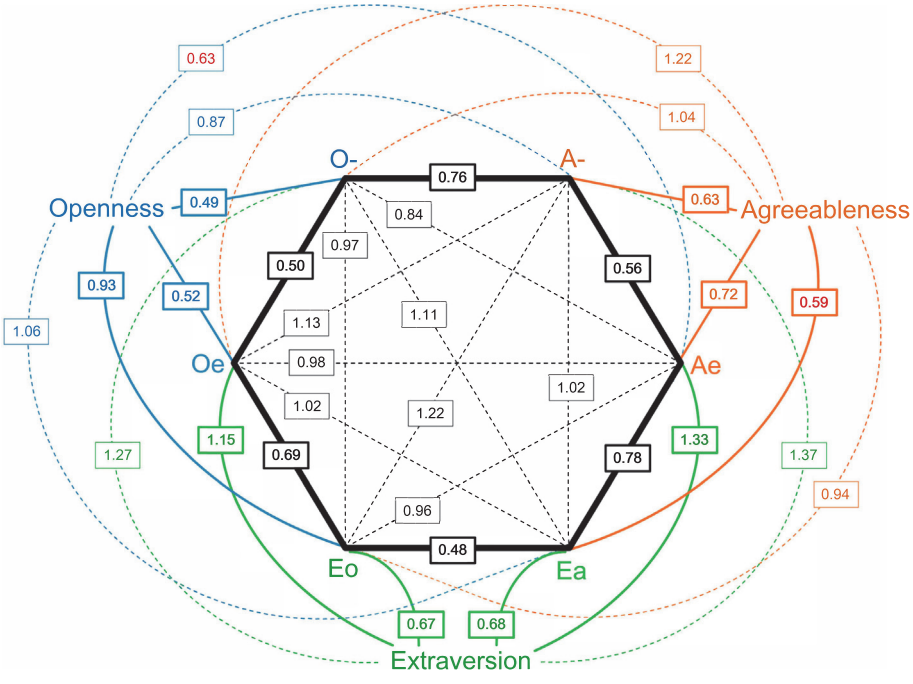

Figure 4. Quantitative evaluation of Hexagonal O-E-A model. The thick solid lines represent the distances assumed close. The thin dashed lines represent the distances assumed far. The red characters indicate distances that did not follow the model.

The results suggest that the Intellectual (I), Writing (E), Artistic (A), and Cooking (-) factors were primarily and positively influenced by Openness, followed by Extraversion (i.e., the "Openness with Extraversion" factor subgroup). The Mechanical (R), Physical (C), and Crafting (R/A) factors were primarily and positively influenced by Openness, followed negatively by Conscientiousness (i.e., the "Openness with Unconscientiousness" factor subgroup). The Teaching (S), Servicing/Selling (S/E), and Helping (S) factors were primarily and positively influenced by Extraversion. While the Teaching (S) and Servicing/Selling (S/E) factors were secondarily and positively influenced by Openness (i.e., the "Extraversion with Openness" factor subgroup), the Helping (S) factor was secondarily and positively influenced by Agreeableness (i.e., the "Extraversion with Agreeableness" factor subgroup). The Clerical (C),

Procedural (C), and Routine physical (C) factors were primarily and positively influenced by Agreeableness. The Clerical (C) factor was secondarily and positively influenced by Openness (the "Agreeableness with Openness" factor subgroup), and the Procedural (C) and Routine physical (C) factors were secondarily and negatively influenced by Extraversion (the "Agreeableness with Introversion" factor subgroup). The Agreeableness with Introversion factor subgroup included an occupation factor weakly negatively influenced by Conscientiousness (i.e., Routine physical). We superimposed these results on the mapping in two-dimensional space obtained by the MDS results (cf. Analysis for Objective 2), as shown in Figure 3. ESM Table S4 shows the raw correlation matrix between the Big Five and OPTI factor scores.

Analysis for Objective 4: Confirmatory Validation of the Relationship Between the Big Five and Data-Driven Occupational Preferences

The results suggest that the Hexagonal O–E–A model has quantitative validity. The three orthogonal canonical correlation coefficients were .48, .43, and .22. The distances across Openness, Extraversion, Agreeableness, and six subgroups of occupational preference factors are shown in Figure 4. The distance between each factor subgroup of preference was closer for those adjacent at the edge of the hexagon but farther for those not adjacent. Overall, the distances between the three traits and factor subgroups of preferences were closest for those primarily linked, moderate for those secondarily linked, and distant for those not linked in the model. The only apparent exception was that the Agreeableness with Extraversion (Ae) factor subgroup of preference, including the Clerical factor of preference, was close to the Openness.

Discussion

For the preliminary Objective 2, we first confirmed the validity and reliability of the entity-based, data-driven preference factors. We conclude that these preferences (OPTI factors) were successfully extracted. These preferences may have more diverse factors while being more convergent within each factor than the RIASEC. Consistent with previous studies in the United States (Deng et al., 2007), these results suggest that RIASEC-type modeling of occupational preferences still needs to be finer for a broad range of occupational titles in contemporary Japan. Indeed, an interpretable, semantic arrangement of OPTI factors regarding RIASEC-based characteristics' dimensions is not available for the diverse occupational titles used in the current study. Therefore, an effective alternative may be an arrangement of the OPTI factors by the Big Five, describing the broader socio-psychological personas of people with various occupational preferences.

For the primary Objective 3, we revealed an interpretable personality explanation of the arrangement of OPTI factors. The OPTI factors in the Openness group had a positive linear relationship with Openness. This group includes occupational preferences centered on technically or artistically handling concepts and things. These occupational preferences may be associated with a desire for broad technical or artistic knowledge. Some preferences may also require creativity. Despite the possible role of other moderators (cf. General Discussion, Limitations and Future Directions), the positive effects of Openness on these preferences are consistent with the concept of Openness (McCrae & Costa, 2008) as measuring an aspect of intelligence that seeks diverse knowledge (DeYoung et al., 2005) and creativity (King et al., 1996). Of these, the Intellectual, Writing, Artistic, and Cooking factors, which seem to include interpersonal interactions in the work to a certain degree, were secondarily and positively influenced by Extraversion (the Openness with Extraversion subgroup). This is consistent with the basic tendency of Extraversion for social stimulation (McCrae & Costa, 2008). On the other hand, the Mechanical, Physical, and Crafting factors, which do not seem to include interpersonal interactions, were not influenced by Extraversion (the Openness with Unconscientiousness subgroup). Interestingly, Conscientiousness had a negative effect on these factors.

The OPTI factors in the Extraversion group had a positive linear relationship with Extraversion. This group includes occupational preferences centered on interpersonal interactions, consistent with the basic tendency of Extraversion (McCrae & Costa, 2008). Of these, the Teaching and Servicing/Selling factors, which include those requiring academic knowledge or negotiation techniques in economic activities, were positively influenced by Openness (the Extraversion with Openness subgroup), consistent with the intellectual aspect of Openness (DeYoung et al., 2005; McCrae & Costa, 2008). On the other hand, the Helping factor, which includes caregiving and welfare, was relatively strongly and positively influenced by Agreeableness (the Extraversion with Agreeableness subgroup). This finding is consistent with an aspect of Agreeableness that measures the degree to which a person is caring (McCrae & Costa, 2008).

The OPTI factors in the Agreeableness group had a positive linear relationship with Agreeableness. This group includes occupational preferences centered on routine tasks performed according to operational procedures. Following operational procedures may be associated with Agreeableness to instructions given by superiors and others in a social context. Of these, Clerical, which sometimes includes relatively technical desk work, was relatively strongly and

positively influenced by Openness (the Agreeableness with Openness subgroup), consistent with the intellectual aspect of Openness (DeYoung et al., 2005; McCrae & Costa, 2008). Compared to the remaining Procedural and Routine physical factors, the Clerical factor is characterized by the absence of negative influence by Extraversion. This finding is consistent with the possible social interactions in clerical occupations. On the other hand, the Procedural and Routine physical factors were negatively influenced by Extraversion (the Agreeableness with Introversion subgroup). Some workers may prefer following a set procedure because they do not prefer new interpersonal interactions (Holland, 1997).

Moreover, the MDS results suggest that the connection between the factor groups is semantically continuous. First, part of the Openness group, secondarily and positively influenced by Extraversion (the Openness with Extraversion subgroup), is adjacent to the Extraversion preference. Second, the Extraversion with Openness subgroup is adjacent to the Openness group. Third, the Extraversion with Agreeableness subgroup is adjacent to the Agreeableness group. Although the Agreeableness with Openness subgroup, which is adjacent to the Extraversion group, was not positively influenced by Extraversion, this part of the Agreeableness group is at least relatively close to the Extraversion group, compared to the other part of the Agreeableness preference, which was negatively influenced by Extraversion (Agreeableness with Introversion subgroup). Besides the three traits, the Openness with Unconscientiousness subgroup is adjacent to the part of the Agreeableness group that partly includes an occupation factor negatively influenced by Conscientiousness. The only exception is that the Agreeableness with Openness subgroup is not adjacent to the Openness group. This proximity may be due to the selective response to the intellectual but closed aspect of Openness measurements (e.g., "Are you analytical?"; Webster & Kruglanski, 1994).

For primary Objective 4, we proposed and validated the Hexagonal O–E–A model incorporating the above MLR and MDS results. Our model may allow us to assess person–vocation fit using the interpretable, socio-psychological persona similarities between the OPTI factors. For example, excluding other possible moderators, people with a high preference for the Eo factor may be best suited for the Teaching and Selling/Servicing factor occupations but also moderately suited for Oe factor occupations (i.e., Intellectual, Writing, Artistic, and Cooking factor occupations) in that they share intellectual or creative activity, as well as the Ea factor occupations (i.e., Helping) in that they share interpersonal activity. Importantly, in contrast to the RIASEC preferences, these tendencies to activities tied to the Big Five can be generalizable to other social or psychological domains.

General Discussion

This study suggests that the Big Five traits can systematically explain an entity-based, data-driven version of occupational preferences (i.e., OPTI factors) compared to the occupational theory-driven RIASEC preferences. Analysis for Objectives 1 and 2 successfully obtained 13 OPTI factors and validated them as data-driven occupational preferences. Moreover, Analysis for Objectives 3 and 4 proposed and validated the Hexagonal O–E–A model. The proposed model assumes that Openness, or the degree of intellectual or creative activity; Extraversion, or the degree of interpersonal activity; and Agreeableness, or the degree of following others, explain the OPTI factors. Furthermore, the preferences associated with these three traits could be arranged circularly by intermediate preferences associated with two. We conclude that the Hexagonal O–E–A model provides a Big Five explanation of the arrangement of occupational preferences.

The significance of using interpretable personality structures, mainly circular structures like circumplex models, to explain broad behaviors such as occupational preferences has been well-established (Hogan, 1983). For instance, Broughton et al. (1991) demonstrated how occupational preferences can be systematically explained using the circumplex model of interpersonal styles. In this model, variables like behavioral styles or occupational preferences are organized circularly along two orthogonal dimensions: "Dominance," linked to Extraversion, and "Love," linked to Agreeableness (Gurtman, 2009; McCrae & Costa, 1989; Wiggins, 1979). Broughton et al. used a selection of occupational titles related to the interpersonal circumplex. They proposed a model that allows in-depth interpretation of these preselected titles, such as associating business executives with a dominant style and bank tellers with a submissive style.

However, our model adds a significant dimension by including Openness, which enables a more comprehensive explanation of various occupations, a factor not incorporated in the previous study. Similar to expanding the interpersonal circumplex to encompass other Big Five domains (Hofstee et al., 1992), our study broadens the scope of personality-based occupational preferences beyond mere interpersonal dimensions. In our model shown in Figure 1, Agreeableness can be represented as a vector oriented 0° counter-clockwise from the horizontal rightward direction and Extraversion as a vector at 240°. Consequently, as outlined in prior research (McCrae & Costa, 1989), the "Love" dimension, indicative of a "warm-agreeable" style, spans between 240° and 0° (e.g., Ea preference, including Helping), starting from a "cold-hearted" style between 60° and 180° (e.g., O- preference, including Mechanical). Similarly, "Dominance" spans between 180° and 240°, representing an "assured-dominant" style (e.g., Oe and Eo preferences, including Intellectual and Teaching), and

extends from an "unassured-submissive" style between 0° and 60° (e.g., Ae and A- preferences, including Clerical and Procedural). Although these interpretations align with previous studies (Broughton et al., 1991) and seem intuitively reasonable, they may be incomplete. Relying solely on "cold-hearted" or "assured-dominant" interpersonal styles without considering Openness, which may embody aspects like intellectuality or creativity, may not fully capture preferences related to Openness, including those for Intellectual, Teaching, or Mechanical occupations.

Given that the Big Five explain a broad range of socio-psychological statuses, the proposed model may further expand the understanding and utilization of occupational preferences. The Hexagonal O–E–A model provides an interpretable circular arrangement of occupational preferences concerning the Big Five. Accordingly, we can adapt our model to propose, test, and utilize various hypotheses about the relationship between people with different occupational preferences and those statuses. For example, individuals with high Openness, which includes a tendency to seek new knowledge (DeYoung et al., 2005; McCrae & Costa, 2008), may have a strong preference for research and teaching. Of these, those with high Openness but also with Extraversion, which includes a tendency to seek interaction with others (McCrae & Costa, 2008), may have a stronger preference for teaching than research. These preferences for research and teaching, characterized by Openness and Extraversion, may be understood concerning the socio-psychological persona associated with these two traits. For example, it has been suggested that Openness is positively correlated with liberalism, while Extraversion is negatively correlated (Gerber et al., 2011). In that case, it may be possible to understand research preference by associating it with a liberal persona and teaching preference by associating it with a not-so-liberal persona. These examples illustrate our aspiration that the proposed model will aid in comprehending occupational preferences and a broad range of socio-psychological behaviors and statuses encountered daily. Our entity-based, data-driven approach fosters harmony among diverse workers and occupations (i.e., situational entities) by considering socio-psychological factors rooted in fundamental individual differences (i.e., situation-general dimensions), such as personality traits or intelligence. These factors might have been previously overlooked when focusing primarily on occupational characteristics (i.e., domain- and situation-specific dimensions).

Limitations and Future Directions

The first limitation of this study is the potential for limited generalizability across different cultures (cf. Hurtado et al., 2019). Our research was conducted in Japan, and while the theoretical framework, particularly the RIASEC model's validity in representing specific occupations, is expected to be generalizable to Japan, the United States, and other countries, it may not apply universally. Studies with large samples have indicated that RIASEC's validity in Japan is nearly as high as in the U.S., Iceland, and Israel (Long et al., 2006; Rounds & Tracey, 1996), despite a smaller study suggesting slightly reduced validity (Tracey et al., 1997). However, because RIASEC-based preferences were not directly measured in the participants in this study, we cannot empirically determine whether the assumptions based on RIASEC were applicable throughout the current survey procedures. Furthermore, cultural differences might influence our entity-based, data-driven occupational preference items. We compared Japanese occupational titles in the OPTI with their U.S. counterparts in the O*NET database (see ESM Table S5), finding that most occupations (60.5 out of 68) had U.S. equivalents. Six occupations without direct counterparts seem present in the U.S. but lack specific categorization in the database. The remaining 1.5 occupations (e.g., kimono tailor, calligrapher) might be unique to Japan, indicating cultural distinctions. However, the impact of cultural differences on occupational items in other countries remains unclear.

Nevertheless, we anticipate that Openness, Extraversion, and Agreeableness traits could systematically explain data-driven occupational preferences in each country, especially the U.S. Our assumption that the Big Five, underlying broad everyday behaviors/statuses should correlate with psychological responses to broader occupational entities is likely culture-independent. The Big Five have demonstrated overall validity across cultures (McCrae & Costa, 2008). Moreover, in the U.S., Openness, Extraversion, and Agreeableness significantly influence RIASEC preferences (Larson et al., 2002). Entity-based, data-driven occupational preferences in the U.S. might differ from RIASEC-based modeling (Deng et al., 2007). Therefore, the theoretical basis for Big Five-based modeling of data-driven occupational preferences for broader entities seems valid, at least in the U.S. However, this remains a hypothesis and necessitates further investigation in other countries.

The relationship between the Big Five and data-driven occupational preferences has empirical backing, but detailed interpretations (i.e., the process of how different personalities lead to specific preferences) are currently hypothetical. Unfortunately, investigating this hypothesis within our model or modifying/extending the model based on theory is beyond the scope of the current study. Moreover, basic trait tendencies, like intellectuality and other factors, including self- and occupational image differences, could be moderators. For example, individuals high in Openness might have open self-images towards various

occupations, influencing their preferences. Future studies should consider controlling for this effect. Occupational prestige may also affect preferences, particularly in the U.S. (Tracey, 2002). Despite the cross-cultural validity of the planar preference circumplex, the three-dimensional U.S. model incorporating a prestige axis has shown less validity in Japan (Tracey et al., 1997), suggesting potential differences in the impact of prestige. The role of traits or model revisions incorporating prestige, especially regarding intellectual or creative occupations, often perceived as prestigious, warrants further examination.

Methodologically, our inventory is limited as it does not cover all occupations; we included only those familiar to many people. While we used standard titles (e.g., financial and insurance clerk) as proxies for specialized titles (e.g., accountant) not in the OPTI, future research should aim for a more comprehensive list. Participants' imperfect understanding of occupations presents another limitation. Despite efforts to mitigate this issue, participants' partial and biased knowledge could influence results. Future studies could address these biases by using more extensive, more diverse populations and more advanced analyses to consider varying levels of occupational knowledge. Additionally, since occupational selections were subjective, participants' responses might not adhere to a uniform standard. Rigorous procedures might require objective verification of participants' understanding, possibly through automated processing (e.g., natural language processing) of participants' descriptions of various occupations.

Nevertheless, retaining certain imperfections could be practical. The knowledge limitations observed among participants are likely common among the general workforce. Unless targeting a niche group with highly specialized preferences, a preference model reflecting common imperfections could be more realistic and applicable.

Conclusion

We proposed an arrangement of entity-based, data-driven occupational preferences based on Big Five traits such as Hexagonal O–E–A occupational personality traits. The proposed model may suggest that Openness, or the degree of intellectual activity; Extraversion, or the degree of interpersonal activity; and Agreeableness, or the degree of following others, explain occupational preferences. This explanation allows for understanding working people from a broader social-psychological perspective.

Electronic Supplementary Materials

The following electronic supplementary material is available with this article at https://doi.org/10.1027/1614-0001/a000423.

ESM 1. This document provides the exploratory factor analysis matrix, the exploratory factor correlation, the confirmatory factor correlation, the correlation results of Big Five traits and OPTI factors, and Japanese occupational titles in the OPTI and their counterparts in the US.

References

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycholexical study. *Psychological Monographs, 47*(1), i–171. https://doi.org/10.1037/h0093360

Ang, S., Van Dyne, L., & Koh, C. (2006). Personality correlates of the four-factor model of cultural intelligence. *Group & Organization Management, 31*(1), 100–123. https://doi.org/10.1177/1059601105275267

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*(1), 74–94. https://doi.org/10.1177/009207038801600107

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*(3), 261–280. https://doi.org/10.1177/014662168801200305

Bollen, K. A. & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Sage.

Broughton, R., Trapnell, P. D., & Boyes, M. C. (1991). Classifying personality types with occupational prototypes. *Journal of Research in Personality, 25*(3), 302–321. https://doi.org/10.1016/0092-6566(91)90022-I

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting incremental fit indices for nonnormality. *Multivariate Behavioral Research, 49*(5), 460–470. https://doi.org/10.1080/00273171.2014.933697

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99–136). American Psychological Association.

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology Press.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

De Hoogh, A. H. B., Den Hartog, D. N., & Koopman, P. L. (2005). Linking the Big Five-Factors of personality to charismatic and transactional leadership; Perceived dynamic work environment as a moderator. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 26*(7), 839–865. https://doi.org/10.1002/job.344

Deng, C.-P., Armstrong, P. I., & Rounds, J. (2007). The fit of Holland's RIASEC model to US occupations. *Journal of Vocational Behavior, 71*(1), 1–22. https://doi.org/10.1016/j.jvb.2007.04.002

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2005). Sources of openness/intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality, 73*(4), 825–858. https://doi.org/10.1111/j.1467-6494.2005.00330.x

Ehrhart, K. H., & Makransky, G. (2007). Testing vocational interests and personality as predictors of person–vocation and person–job fit. *Journal of Career Assessment, 15*(2), 206–226. https://doi.org/10.1177/1069072706298105

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286–299. https://doi.org/10.1037/1040-3590.7.3.286

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error.

Journal of Marketing Research, 18(1), 39–50. https://doi.org/10.1177/002224378101800010

Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2011). The Big Five personality traits in the political arena. *Annual Review of Political Science, 14*, 265–287. https://doi.org/10.1146/annurev-polisci-051010-111659

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Gurtman, M. B. (2009). Exploring personality with the interpersonal circumplex. *Social and Personality Psychology Compass, 3*(4), 601–619. https://doi.org/10.1111/j.1751-9004.2009.00172.x

Gutiérrez, J. L. G., Jiménez, B. M., Hernández, E. G., & Puente, C. P. (2005). Personality and subjective well-being: Big Five correlates and demographic variables. *Personality and Individual Differences, 38*(7), 1561–1569. https://doi.org/10.1016/j.paid.2004.09.015

Hoff, K. A., Song, Q. C., Wee, C. J., Phan, W. M. J., & Rounds, J. (2020). Interest fit and job satisfaction: A systematic review and meta-analysis. *Journal of Vocational Behavior, 123*, Article 103503. https://doi.org/10.1016/j.jvb.2020.103503

Hofstee, W. K., De Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63*(1), 146–163.

Hogan, R. (1983). A socioanalytic theory of personality. In M. Page (Ed.), *Nebraska symposium on motivation* (pp. 55–89). University of Nebraska Press.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity, 44*(1), 153–166. https://doi.org/10.1007/s11135-008-9190-y

Holland, J. L. (1985). *Vocational Preference Inventory (VPI): Professional manual*. Psychological Assessment Resources.

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

Holland, J. L., Fritzsche, B. A., & Powell, A. B. (1997). *SDS: Self-directed search: Technical manual*. Psychological Assessment Resources.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hurtado Rúa, S. M., Stead, G. B., & Poklar, A. E. (2019). Five-factor personality traits and RIASEC interest types: A multivariate meta-analysis. *Journal of Career Assessment, 27*(3), 527–543. https://doi.org/10.1177/1069072718780447

Iwai, R., Kumada, T., Kawahara, D., & Kurohashi, S. (2018, March 1–3). Automatic conceptual translation approach: Construction and validation of a Japanese Five-Factor personality questionnaire based on statistical machine translation [Poster presentation]. *The 19th Annual Meeting of Society for Personality and Social Psychology*, Atlanta, GA, Unites States.

Iwai, R., Kawahara, D., Kumada, T., & Kurohashi, S. (2019). Applying machine translation to psychology: Automatic translation of personality adjectives. *Proceedings of Machine Translation Summit XVII: Translator, Project and User Tracks,* 23–29. https://aclanthology.org/W19-6704

The Japan Institute for Labour Policy and Training. (2011). *Daiyonkai kaitei koseirodoshohen shokugyobunrui*. [Fourth Revised Classification of Occupations, edited by the Ministry of Health, Labor and Welfare]. https://www.jil.go.jp/institute/seika/shokugyo/04th/index.html

Jensen-Campbell, L. A., & Malcolm, K. T. (2007). The importance of conscientiousness in adolescent interpersonal relationships. *Personality and Social Psychology Bulletin, 33*(3), 368–383. https://doi.org/10.1177/0146167206296104

King, L. A., Walker, L. M., & Broyles, S. J. (1996). Creativity and the five-factor model. *Journal of Research in Personality, 30*(2), 189–203. https://doi.org/10.1006/jrpe.1996.0013

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*(1), 1–27. https://doi.org/10.1007/BF02289565

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Sage.

Larson, L. M., Rottinghaus, P. J., & Borgen, F. H. (2002). Meta-analyses of Big Six interests and Big Five personality factors. *Journal of Vocational Behavior, 61*(2), 217–239. https://doi.org/10.1006/jvbe.2001.1854

Long, L., Watanabe, N., & Tracey, T. J. (2006). Structure of interests in Japan: Application of the personal globe inventory occupational scales. *Measurement and Evaluation in Counseling and Development, 38*(4), 222–236. https://doi.org/10.1080/07481756.2006.11909784

Lorenzo-Seva, U., & Ferrando, P. J. (2021). Not positive definite correlation matrices in exploratory item factor analysis: Causes, consequences and a proposed solution. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(1), 138–147. https://doi.org/10.1080/10705511.2020.1735393

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement, 31*(3), 657–674. https://doi.org/10.1177/001316447103100307

Matsumoto, S., Sato, M., Kamakura, T., Nishizawa, H., & Matsumoto, J. (2012). *Shokumukozo ni kansuru kenkyu*. (Rodo-seisaku kenkyuhokokusho No. 146). [Research on job structure (Labor Policy Research Report No. 146)]. https://www.jil.go.jp/institute/reports/2012/0146.html

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81–90. https://doi.org/10.1037/0022-3514.52.1.81

McCrae, R. R., & Costa, P. T. (1989). The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. *Journal of Personality and Social Psychology, 56*(4), 586–595. https://doi.org/10.1037/0022-3514.56.4.586

McCrae, R. R., & Costa Jr, P. T. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (pp. 159–181). The Guilford Press.

Mount, M. K., Barrick, M. R., Scullen, S. M., & Rounds, J. (2005). Higher-order dimensions of the Big Five personality traits and the big six vocational interest types. *Personnel Psychology, 58*(2), 447–478. https://doi.org/10.1111/j.1744-6570.2005.00468.x

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7*(4), 384–403. https://doi.org/10.1177/1745691612449021

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research, 12*, 2825–2830. https://jmlr.org/papers/1161v12/pedrego-sa11a.html

R Core Team. (2016). *R: A language and environment for statistical computing*. Foundation for Statistical Computing. https://www.R-project.org

Revelle, W. R. (2017). *psych: Procedures for personality and psychological research*. Northwestern University. http://CRAN.R-project.org/package=psych

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315

Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin, 28*(6), 789–801. https://doi.org/10.1177/0146167202289008

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Rounds, J., & Tracey, T. J. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *Journal of Counseling Psychology, 43*(3), 310–329. https://doi.org/10.1037/0022-0167.43.3.310

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Steiner, M. D., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software, 5*(53), Article 2521. https://doi.org/10.21105/joss.02521

Streiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *The Canadian Journal of Psychiatry, 39*(3), 135–140. https://doi.org/10.1177/070674379403900303

Strong, E. K. (1935). Predictive value of the vocational interest test. *Journal of Educational Psychology, 26*(5), 331–349. https://doi.org/10.1037/h0062498

Tracey, T. J. G., & Rounds, J. (1995). The arbitrary nature of Holland's RIASEC types: A concentric-circles structure. *Journal of Counseling Psychology, 42*(4), 431–439. https://doi.org/10.1037/0022-0167.42.4.431

Tracey, T. J. G., & Rounds, J. (1996). The spherical representation of vocational interests. *Journal of Vocational Behavior, 48*(1), 3–41. https://doi.org/10.1006/jvbe.1996.0002

Tracey, T. J., Watanabe, N., & Schneider, P. L. (1997). Structural invariance of vocational interests across Japanese and American cultures. *Journal of Counseling Psychology, 44*(4), 346–354. https://doi.org/10.1037/0022-0167.44.4.346

Tracey, T. J. G. (2002). Personal Globe Inventory: Measurement of the spherical model of interests and competence beliefs. *Journal of Vocational Behavior, 60*(1), 113–172. https://doi.org/10.1006/jvbe.2001.1817

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*(3), 321–327. https://doi.org/10.1007/BF02293557

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*(6), 1049–1062. https://doi.org/10.1037/0022-3514.67.6.1049

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*(3), 395–412. https://doi.org/10.1037/0022-3514.37.3.395

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*(3), 1–13. https://doi.org/10.33151/ajp.8.3.93

Yamashita, J. (2023). *A state-trait approach for bridging the gap between basic and applied occupational psychological constructs* [Doctoral dissertation, Kyoto University]. KURENAI. https://doi.org/10.14989/doctor.k24821

Yamashita, J., Iwai, R., Oishi, H., & Kumada, T. (2023). *Personality Traits Systematically Explain the Semantic Arrangement of Occupational Preferences* [Data Set]. https://doi.org/10.17605/osf.io/r4dgw

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*(1), 21–29. https://doi.org/10.22237/jmasm/1177992180

History
Received October 13, 2023
Revision received March 17, 2024
Accepted April 29, 2024
Published online August 5, 2024

Acknowledgments
This paper is a revised version of a chapter included in the first author's doctoral dissertation (Yamashita, 2023).

Conflict of Interest:
Jumpei Yamashita and Haruo Oishi are paid employees of Nippon Telegraph and Telephone Corporation (NTT). NTT has entered into a joint research agreement with Takatsune Kumada and has provided funding for this research. The authors affirm that there are no patents or marketed products related to the content of this manuscript.

Publication Ethics
Informed consent was obtained from all participants involved in the study. Procedures involving human participants adhered to the ethical guidelines set forth by the Ethics Committee of the Graduate School of Informatics, Kyoto University (KUIS–EAR–2019–005), ensuring compliance with established ethical standards.

Authorship
Jumpei Yamashita: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft; Ritsuko Iwai: conceptualization, investigation, supervision, writing – review & editing; Haruo Oishi: funding acquisition, project administration; Takatsune Kumada: conceptualization, methodology, resources, supervision, writing – review & editing. All authors approved the final version of the article.

Open Data
The data underlying the main findings reported in this paper are openly available in Open Science Framework (OSF) at https://doi.org/10.17605/osf.io/r4dgw (Yamashita et al., 2023).

Funding
The research grant for this study was provided to Takatsune Kumada by the Nippon Telegraph and Telephone Corporation (NTT). Additionally, NTT supported the work of Jumpei Yamashita and Haruo Oishi by paying their salaries.

ORCID
Jumpei Yamashita
https://orcid.org/0000-0001-9411-0523

Jumpei Yamashita
NTT Access Network Service Systems Laboratories
Nippon Telegraph and Telephone Corporation
3-9-11 Midori-cho
180-8585, Musashino-shi, Tokyo
Japan
junpei.yamashita@ntt.com