

Research notes

Development of the Input Data for the Occupational Information Network of Japan

KAMAKURA Tetzushi

MATSUBARA Ayako

MATSUMOTO Shinsaku

I. Introduction: What is the occupational information network of Japan?

The occupational information network of Japan¹ is a brand-new website (opened on March 19, 2020) developed by the Ministry of Health, Labour and Welfare (MHLW). It provides information on approximately 500 occupations (only available in Japanese) from several viewpoints including jobs, tasks, and skill requirements so it can support job hunting for job seekers and recruiting activities of companies.

Four types of information are currently available: text-based descriptions, cross-occupational numeric estimates, recent labor market information, and visual content (short videos). The source of the first two types of information is the database called “input data” developed by the Japan Institute for Labour Policy and Training (JILPT), an independent administrative institute affiliated with MHLW. This paper reports the outline of JILPT’s data collecting efforts over the past two years.

II. Method: How to collect information for the database?

JILPT has taken different approaches to data collection for the two types of information: text-based descriptions and numeric estimates. About 250 occupations (about half of the total of 500 to be provided) were targeted per year because of the need to prepare the database within two years in time for the release of the website. This section mainly

describes the process of data collection.

1. Data collection for text-based descriptions

The first drafts of the text-based descriptions were basically written by entrusted research companies based on interviews with relevant organizations both in the public and private sectors. The legacy data of descriptions that JILPT had accumulated and revised over the past several decades were utilized to the extent possible in this data collection.

The data collection for targeted occupations was divided into three task levels according to the effort needed to collect recent information: minor revision, major revision, and newly written. At the minor revision level, the legacy descriptions were substantively reused and only some outdated parts were revised. At the major revision level, the legacy descriptions were referred to, but interviews were conducted as well to revise them drastically. At the newly written level, as there were no legacy descriptions to refer to, interviews and relevant authoritative information available on websites through the Internet were largely relied on as sources of information.

When these activities were finished and the first drafts were submitted to JILPT, the quality of the content was discussed among six editors in an editorial committee mainly composed of MHLW and JILPT researchers. It was possible for the drafts to be thoroughly rewritten from the viewpoints of neutrality of the information as provided by a governmental institution, objectivity, accuracy,

and the freshness of information. Second drafts mainly at the newly written level were sent to relevant organizations for their final checks and modifications. Detailed sections on descriptions which are currently available will be described in the next section.

2. Data collection for numeric estimates

Several domains of numeric estimates were fundamentally calculated based on the results of online surveys of incumbents working in targeted occupations. In addition, paper-based surveys were conducted for some occupations for which the sample size of the online survey was less than 20.

(1) Online surveys

Two online surveys, which basically had the same items, were conducted in 2018 and 2019, each targeting about 250 occupations. All respondents were pre-registered members of Internet research firms in Japan.² The research firms sent a recruiting e-mail to each member for the survey, and members voluntarily answered the questionnaire via computer or smartphone. Although each survey period was about one month, it was possible for data collection to close for some occupations if the number of responses reached 60 or above.

Participants answered some demographic questions and selected their own occupation from a list of targets. The items of the six cross-occupational domains and the Tasks domain (occupation-specific items) were shown to them only when they found their own occupation on the list and the chosen occupation was one for which information was still being collected. Detailed domains will be described in the next section.

(2) Paper-based surveys

In 2019, JILPT also conducted paper-based surveys for occupations for which it had not been possible to collect an adequate number of samples (at least 20) online in the 2018 survey. JILPT asked for the cooperation of relevant organizations directly or through MHLW. Several organizations, which were mostly in the public sector, accepted our request and we sent printed questionnaires to them.

The survey period was 2–4 months depending on

the organizations. When the number of responses by occupation reached 20 or above, the samples from the paper-based surveys became the only source for estimation and the 2018 online survey data were discarded. The two sources were mixed for the estimation only when the number of samples did not reach 20 in one source (there was only one such case).

(3) Final data for estimation

The total number of samples both from online and paper-based surveys is 25,974 (after data screening) for 511 occupations. The average is 50.8 per occupation (SD=19.0). Estimates (mean or proportion) were then calculated for each item for 426 occupations and tabulated as the master data for the occupational information network of Japan. Of the occupations, 85 were excluded mainly because it was not possible to reach a sample size of 20.³

Even though the master data has been stored at JILPT, this is not the same as the input data submitted to MHLW for the website. On the website, each page for a given occupation must have a text-based description, with no exception. In cases where these descriptions have not been prepared yet, the corresponding numeric estimates have been eliminated in the input data, although they are free from defects.

III. Results: What kinds of data are initially available?

1. Available text-based descriptions

The structure of the text-based descriptions is just like the Occupational Outlook Handbook (OOH) published online by the U.S. Bureau of Labor Statistics (BLS). This means each description has three sections: “What They Do,” “How to Become One,”⁴ and “Work Environment.” In addition, several pieces of relevant information are appended including other names of the occupation, codes in the Standard Occupational Classification in Japan (compiled and edited by MHLW), related credentials and licenses, and sources of and links to additional information.

2. Available numeric estimates

The domains of the numeric estimates are basically derived from the Occupational Information Network (O*NET) sponsored by the U.S. Department of Labor/Employment and Training Administration (USDOL/ETA). Most cross-occupational descriptors of O*NET have an academic basis mainly in psychology, and they possess appropriate generality to assure that they can be asked for any kinds of occupations (Peterson et al. 2001). These properties make it easier to also apply the same descriptors for other countries.

Even so, we had to perform some “localization” beyond simple translation. As a source for scientific research, it is critically important to maintain the same structure with O*NET as well as to keep it

possible to assure that anyone could compare the two databases with minimal effort. However, the occupational information network of Japan is a website primarily for Japanese students, job seekers, workers, career counselors, and HR professionals. This means we should choose to focus on the aspect of practical utility rather than scientific comparability in the event of conflict. For instance, “Philosophy and Theology,” a descriptor in the Knowledge domain, was replaced by “Philosophy and Religious Studies” because in Japan, jobs which require knowledge of religions often need knowledge of various religions including Buddhism. We also had to consider the sustainability of the website (mainly in terms of the budget) as well as the burden on respondents.

Six domains of cross-occupational information

Sample page for “Semiconductor Processors”: Text-based description and other information

半導体製造

印刷する ★ マイリストに保存

職業別名 : 半導体製品製造設備オペレーター
職業分類 : 電気機械器具組立設備制御・監視員
属する産業 : 製造業 (クリックすると別サイトのしょくばらへ移ります)

● どんな仕事? ● 就業するには? ● 労働条件の特徴 ● しごと能力プロフィール ● 類似する職業 ● 関連リンク

どんな仕事?

半導体の製造工程で設備、装置、機械を操作、監視し半導体を製造する。

半導体は、シリコンなどの元素半導体を材料として作られる電子部品で、トランジスタやダイオードなどの素子単体（ディスクリート半導体部品）とトランジスタなどで構成されるLSI（大規模集積回路）がある。コンピュータ、冷蔵庫、電子レンジ等の家電製品、スマートフォン、タブレット等の一般消費者向け製品から産業用機械まで、小型化、軽量化、省エネルギー化が実現されているのは、半導体がチップやLSIに大量に組み込まれるように設計・開発を進め、効率よく基盤の中に収めることに成功した成果といえる。

生産職は、半導体工場の中で、半導体の量産、組み立て、検査の工程を担っている。生産の自動化が進み、仕事は製造のための設備、装置、機械の操作、監視が中心となっている。設備、装置、機械は非常に精密であり、仕様書、マニュアル等の操作手順に従った正確な操作を行い、担当する装置等が順調に動いているかを常に監視する。軽微なトラブルには対応するが、解決が困難な場合にはエンジニア等に状況を報告し、相談するなど対応を図る。

半導体の製造は、高い製品歩留まりや品質管理を実現するために、クリーンルーム内で各種の装置等を操作する。

総工程数は400～600工程に及ぶが、前工程（ウェーハ処理工程：表面に電気回路を作る）と後工程（組み立て工程）に大きく分けることができる。

具体的な仕事を工程に沿ってみると、前工程では、単結晶棒をスライスしてシリコンウェーハに加工し、成膜（せいまく）工程で熱処理装置を使って、CVD（化学気相蒸着）法によりウェーハ上にシリコン酸化膜（せいかまい）をシリコン窒素の

[動画]



Retrieved from <https://shigoto.mhlw.go.jp/User/Occupation/Detail/268> (May 19, 2020)

are currently available: Interests, Work Values, Skills, Knowledge, Work Context, and Education-and-Training in addition to one occupation-specific information, Tasks domain.

(1) Interests

In this domain, there are six descriptors precisely corresponding to those in O*NET: “Realistic,” “Investigative,” “Artistic,” “Social,” “Enterprising,” and “Conventional.” This RIASEC taxonomy, advocated by Holland (1959), is also well-known among practitioners in Japan. It seemed that there were no cultural concerns with regard to applying it directly.

Participants were asked the question “What kind of interest is suited to your current job?” and instructed to answer in a five-point scale: from “Not suitable” as 1 to “Very suitable” as 5. Simple average values among incumbents were recorded in the input data as estimates for each occupation.

(2) Work Values

In this domain, there are 10 descriptors. Half of them basically correspond to those in O*NET: “Achievement,” “Independence,” “Recognition,” “Relationship,” and “Organizational Support.” The other five descriptors are the results of localization: “Stability,” “Safety and Health,” “Expertise,” “Service and Dedication,” and “Lifestyle.”

According to Sager (1999), the Work Values in O*NET were originally derived from the Theory of Work Adjustment (TWA) (Dawis and Lofquist 1984) and the Minnesota Job Description Questionnaire (MJDQ) (Dawis 1991). Although this background assures generalizability for Japanese incumbents to some extent, some addition and segmentation seemed to be appropriate to improve the practical utility for them.

First, “Working Conditions” in O*NET was divided into “Stability” and “Safety and Health.” It seemed to be more informative for Japan’s labor market to provide these estimates separately because they would vary independently and both could be crucial values for job seekers.

Second, three values were added referring to the Theory of Career Anchors (Schein 1990). In Japan, the concept of “life-work balance” (i.e.,

the “lifestyle” Career Anchor) has recently been attracting attention to correct long working hours. However, neither TWA nor MJDQ includes such a concept. Thus, we decided to add three concepts from eight Career Anchors as work values, including lifestyle, which could be important for Japanese workers.

Participants were asked the question “In what aspect is it easy to get satisfaction in your current job?” and instructed to answer in a five-point scale: from “Not easy to get satisfaction” as 1 to “Very easy to get satisfaction” as 5. Simple average values were recorded in the input data.

(3) Skills

In this domain, there are 39 descriptors, 35 of which basically correspond to those in O*NET, such as “Reading Comprehension,” “Active Learning,” “Complex Problem Solving,” “Management of Financial Resources,” “Coordination,” “Judgment and Decision Making,” “Equipment Maintenance,” and so on. Four descriptors were added as basic content skills in foreign languages: “Reading in a Foreign Language,” “Listening in a Foreign Language,” “Writing in a Foreign Language,” and “Speaking in a Foreign Language.”

For many people living in the US, their first language (i.e., English) has been common in international situations, which would explain why “Foreign Language” has been set only in the Knowledge domain in O*NET. However, in Japan, foreign language skills (e.g., English, Chinese) should be handled as basic content skills “that allow people to acquire information and convey this information to others” (Mumford et al. 1999, 51) in the context of the increasing globalization of business.

Participants were asked “What level of skills is needed for your current job?” and answered using a somewhat atypical eight-point scale. First, the option of “Not relevant for my job” was set as 0. Subsequently, a seven-point level scale was shown below that option only when the skill was relevant, with three “anchors” on 2, 4, and 6 just like the O*NET level scales. Although trained analysts are rating both importance and level in current O*NET

data collection, we have been collecting only level information, because it has been repeatedly reported in previous studies that these two ratings have a high correlation (e.g., Handel 2016; Taylor, Li, Shi, and Borman 2008; Tsacoumis and Van Iddekinge 2006). As a result of discussion in the development committee where overall policies for the data collection were determined, we concluded that level information with concrete anchors is more practical than abstract importance information.

At first, we tried to reuse O*NET anchors (105 in total) with simple translation for the corresponding 35 descriptors. However, most anchors seemed to have problems either in terms of apprehensibility for Japanese users or validity at a located level (too low as level 2 or too high as level 6). Therefore, 101 anchors for corresponding descriptors were newly prepared as well as 12 anchors for four additional ones through discussions among working group members, and just four anchors remain as simple translations from O*NET. Simple average values were recorded in the input data.

(4) Knowledge

In this domain, there are 33 descriptors basically corresponding to those in O*NET. There are some localizations as shown below to assure apprehensibility and practical utility for Japanese users. Detailed reasons for each localization can be seen in the development report of JILPT (Kamakura 2020; only available in Japanese).

- “Food production” was replaced by “Agriculture and Farming.”
- “Sociology and Anthropology” was replaced by “Sociology.”
- “English Language” was replaced by “Japanese Grammar and Vocabulary.”
- “Foreign Language” was replaced by “Foreign Language Grammar and Vocabulary.”
- “Philosophy and Theology” was replaced by “Philosophy and Religious Studies.”

Participants were asked “How important is the knowledge for your current job?” and instructed to answer in a six-point scale composed of “Not Relevant to my Job” as 0 and a five-point importance scale from “Not Important” as 1 to “Very

Important” as 5. Again, although importance and level information are collected in O*NET, we chose one of them to reduce the burden on respondents. For the Knowledge domain, we decided to collect only importance because (1) it was too difficult to prepare level anchors both for skills and knowledge simultaneously in a limited period of time, and (2) it seemed that the need for level information for Knowledge was not as high as for Skills. Simple average values were recorded in the input data.

(5) Work Context

In this domain, there are 23 descriptors basically selected from the 57 in O*NET. This is the only domain in which the number of items has been reduced compared to O*NET. To prioritize items and reduce the burden on respondents, some interviews were conducted targeting organizations specializing in supporting women, senior citizens, and disabled persons.

As a result, from the viewpoint of practical utility, we chose 23 descriptors shown below.

- Nine descriptors from the subgroup of 14 Interpersonal Relationships.
“Contact With Others,” “Face-to-Face Discussions,” “Telephone,” “Letters and Memos,” “Frequency of Conflict Situations,” “Work with Work Group or Team,” “Deal with External Customers,” “Coordinate or Lead Others,” and “Responsibility for Outcomes and Results.”
- Six descriptors from the 30 Physical Work Conditions.
“Indoors, Environmentally Controlled,” “Indoors, Not Environmentally Controlled,” “Outdoors,” “Spend Time Sitting,” “Spend Time Standing,” and “Spend Time Making Repetitive Motions.”
- Eight descriptors from the 13 Structural Job Characteristics.
“Time Pressure,” “Importance of Being Exact or Accurate,” “Importance of Repeating Same Tasks,” “Pace Determined by Speed of Equipment,” “Consequence of Error,” “Freedom to Make Decisions,” “Structured versus Unstructured Work,” and “Work Schedules.”

In this domain, the questions and scales varied depending on the descriptor corresponding to the O*NET original: importance, frequency, degree of responsibility, time proportion in working hours, and so on. Most questions were followed by five-point scale options. As an exception, “Work Schedules” had three category options: “Regular,” “Irregular,”

and “Seasonal.” Although simple average values are basically recorded in the input data for most descriptors, proportions became the estimates for “Work Schedules.”

(6) Education and Training

In this domain, there are four descriptors partially corresponding to the O*NET questionnaire with the same title: “Common Academic Background,” “Training Period before Entering the Occupation (Excluding School Education),” “Related Work Experience Period before Entering the Occupation,” and “On the Job Training Period after Entering the Occupation.” Although the data for the latter three questions have been integrated into the Job Zones domain on O*NET OnLine, we decided to provide them directly mainly because we did not comprehend the detailed process to rate the level of Job Zones.

For “Common Academic Education,” participants were asked “What educational background is common in your current occupation?” and there were eight options allowing multiple answers: from “Less than High School Diploma” to “Doctoral Degree.” For the other three questions, there were nine options about period (single answer) from “Not needed” to “More than ten years.” In addition, the option of “I don’t know” was set for all four questions. Proportions are recorded in the input data.

(7) Tasks

This domain is occupation-specific and the number and content of questions vary by occupation. For instance, there are 11 tasks for “Plumber”: “Read drawings and inspect structures to prepare the necessary materials and tools,” “Use a ruler or level to determine where a pipe will be laid and mark it,” “Attach openings to a pipe and pipe fittings in structures using a hand tool and a power tool,” and so on.

Participants were asked “Which of these tasks do you perform in your current job?” and checked all the tasks they perform. Although O*NET has collected several pieces of information (relevance, frequency and importance) for each task, we decided to collect only the implementation rates mainly to reduce the burden on respondents. Proportions are

recorded in the input data.

IV. Discussion: Several considerations about the database

The input data prepared for the website can be downloaded in the csv or xlsx file type on the Internet. However, there are several precautions when using them, especially for academic purposes.

1. Some descriptive statistics released a half year later

The information included in the currently available data only consists of means (mainly in cross-occupational domains) or proportions (mainly in the Education-and-Training and Tasks domains). Other descriptive statistics values like standard error, 95% confidence interval, and sample size are currently being prepared considering finite population correction. These pieces of information will be available in around October 2020.

2. Limited domains compared to O*NET

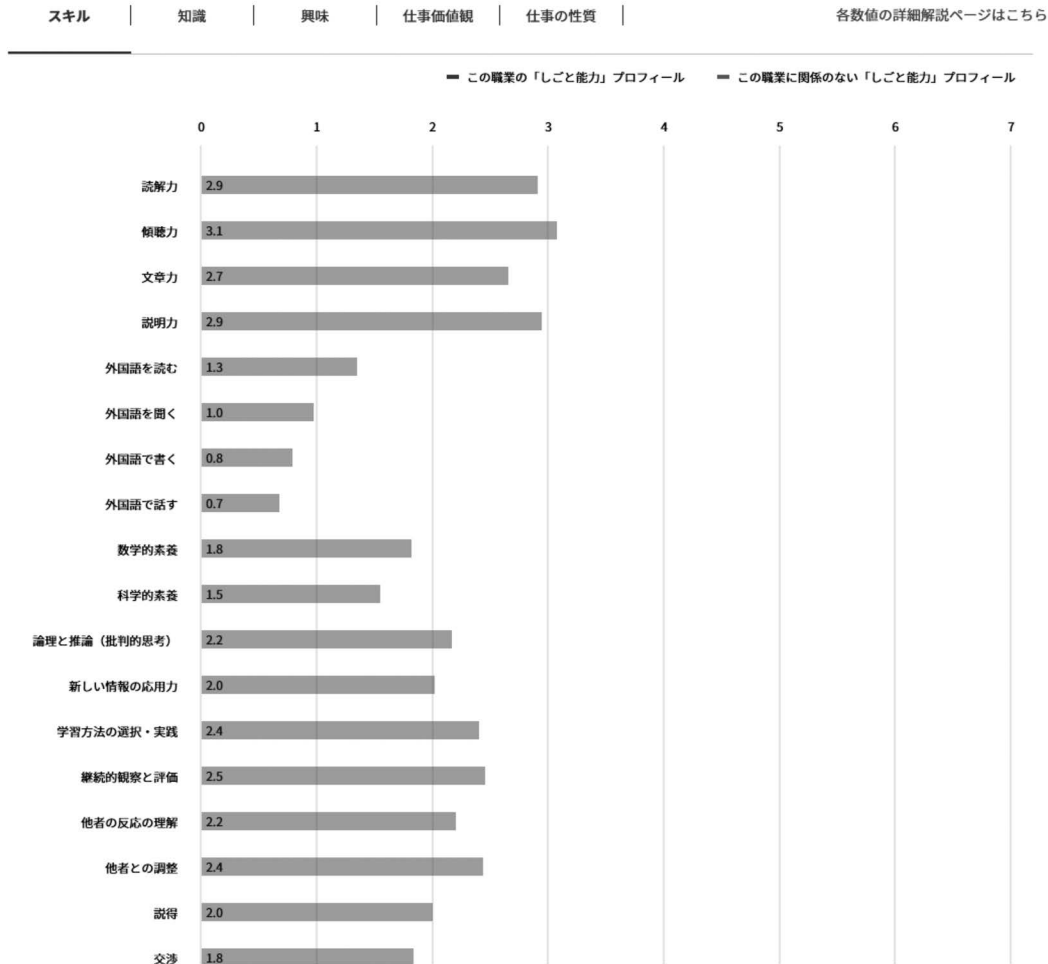
Some informative domains in O*NET have yet not been prepared, including Abilities, three tiers of Work Activities, Work Styles, and Tools & Technology (T2). In the initial development, we had to ask participants to answer all questions for seven targeted domains to develop the database in just two years. It seemed difficult to add any more questions at once.

From the next survey in 2020, however, it has been decided to separate the questionnaire into several parts and conduct surveys for one part a year. We are now planning to add the Generalized Work Activities domain based on these surveys. Although a clear long-term roadmap does not yet exist, the other remaining domains in O*NET might also be developed in the future.

3. Some challenges to conducting cross-national comparisons

Even though the cross-occupational domains basically correspond to O*NET, there are some challenges to conducting cross-national comparisons between the United States and Japan.

Sample page for “Semiconductor Processors”: Skills information しごとと能力プロフィール



Retrieved from <https://shigoto.mhlw.go.jp/User/Occupation/Detail/268> (May 19, 2020)

First, the targeted occupations are widely different. For instance, “Tofu Maker” is the occupation of producing tofu, a traditional Japanese food made of soybeans, and cannot be targeted independently in O*NET. Researchers would have to devise a method to find correspondences between the lists of occupations for each country.

Second, as mentioned above, there are some localizations even in corresponding domains. Interests might be relatively easy to compare. Knowledge (importance only) and Work Context have some localized descriptors, but can still be compared per item. Work Values and Education-

and-Training are difficult to compare because the structures of domains were changed to some extent. Skills (excepting the four additional descriptors) are superficially easy to compare, but researchers have to consider the differences of the “anchors” for the level scale. In such a situation, it could be reasonable to utilize a rank ordering rather than mean ratings, referring to the findings in Taylor et al. (2008).

Finally, the language barrier is an unavoidable challenge, especially for foreign language speakers. We do not have any plan to provide information in another language for now. Even so, it might be the case that some kinds of correspondence tables will

be officially published in the future.

4. Need for methodological sophistication in data collection

There are many concerns about the methodological validity of the surveys for incumbents described in this paper, including population coverage, sampling bias, possible commingling of other occupations, level anchors for Skills, and the burden to answer. Over the last two years, we have not had a chance to fully consider these issues. It is necessary to address these issues in future surveys based on the available evidence.

1. The occupational information network of Japan is a tentative name that comes from the pioneering project O*NET (Occupational Information Network) sponsored by the U.S. Department of Labor/Employment and Training Administration (USDOL/ETA). As described in this paper, we have deeply consulted O*NET to develop cross-occupational numeric estimates. MHLW plans to solicit names for the website in the near future.
2. They offered their basic personal information (e.g., gender, age, prefecture of residence) at the first registration and have continued to answer various kinds of online surveys requested by outside organizations so they can obtain points that can be converted into benefits. There are no obligations and they can decide whether or not to answer these surveys.
3. This minimal number for estimation was decided based on findings about estimate accuracy (e.g., SE and 95% CI) referring to the U.S. Department of Labor (2018). Even so, there are an exception. Detailed information on this topic can be seen in the development report by JILPT (Kamakura 2020; only available in Japanese).
4. The name of the occupation is inserted instead of “One” on the OOH website.
5. Although there are two descriptors relating to “Outdoors” in O*NET, we decided to integrate them into one based on the responses in the interviews.

References

- Dawis, René V. 1991. “Vocational Interests, Values, and Preferences.” In *Handbook of Industrial and Organizational Psychology*, vol.2, edited by Marvin D. Dunnette and Leaetta M. Hough, 833–872. Palo Alto, CA: Consulting Psychologists Press.
- Dawis, René V., and Lloyd H. Lofquist. 1984. *A Psychological*

Theory of Work Adjustment: An Individual-Differences Model and Its Applications. Minneapolis, MN: University of Minnesota Press.

- Handel, Michael J. 2016. “The O*NET Content Model: Strengths and Limitations.” *Journal of Labour Market Research* 49, no.2: 157–176.
- Holland, John. L. 1959. “A Theory of Vocational Choice.” *Journal of Counseling Psychology* 6, no.1: 35–45.
- Kamakura, Tetzushi. 2020. “Shokugyo ni kansuru suchi joho no sakusei” [The development of occupational information: Numerical estimates]. In *Research for the development of input data for the occupational information network of Japan*. JILPT research material series no.227, 40–104. Accessed April 13, 2020. <https://www.jil.go.jp/institute/siryō/2020/227.html>.
- Mumford, Michael D., Norman G. Peterson, and Ruth A. Childs. 1999. “Basic and Cross-Functional Skills.” In *An Occupational Information System for the 21st Century: The Development of O*NET*, edited by Norman G. Peterson, Michael D. Mumford, Walter C. Borman, P. Richard Jeanneret, and Edwin A. Fleishman, 49–69. Washington, DC: American Psychological Association.
- Peterson, Norman G., Michael D. Mumford, Walter C. Borman, P. Richard Jeanneret, Edwin A. Fleishman, Kerry Y. Levin, Michael A. Campion, Melinda S. Mayfield, Frederick P. Morgeson, Kenneth Pearlman, Marilyn K. Gowing, Anita R. Lancaster, Marilyn B. Silver, and Donna M. Dye. 2001. “Understanding Work Using the Occupational Information Network (O*NET): Implications for Practice and Research.” *Personnel Psychology* 54: 451–492.
- Sager, Christopher E. 1999. “Occupational Interests and Values.” In *An Occupational Information System for the 21st Century: The Development of O*NET*, edited by Norman G. Peterson, Michael D. Mumford, Walter C. Borman, P. Richard Jeanneret, and Edward A. Fleishman, 197–211. Washington, DC: American Psychological Association.
- Schein, Edgar. H. 1990. *Career Anchor: Discovering Your Real Values*. San Diego, CA: Pfeiffer.
- Taylor, Paul J., Wen-Dong Li, Kan Shi, and Walter C. Borman. 2008. “The Transportability of Job Information across Countries.” *Personnel Psychology* 61: 69–111.
- Tsacoumis, Suzanne, and Chad. H. Van Iddekinge. 2006. *A Comparison of Incumbent and Analyst Ratings of O*NET Skills*. Accessed February 18, 2020. https://www.onetcenter.org/dl_files/SkillsComp.pdf.
- U.S. Department of Labor. 2018. *O*NET Data Collection Program: Office of Management and Budget Clearance Package Supporting Statement Part B*. Accessed April 13, 2020. https://www.onetcenter.org/dl_files/omb2018/Supporting_StatementB.pdf.

KAMAKURA Tetzushi

Researcher, The Japan Institute for Labour Policy and Training. Research interest: Educational psychology, educational technology, career guidance, occupational information.
<https://www.jil.go.jp/english/profile/kamakura.html>



MATSUBARA Ayako

Research Director, The Japan Institute for Labour Policy and Training. Research interest: Vocational guidance, occupational information.
<https://www.jil.go.jp/english/profile/matsubara.html>



MATSUMOTO Shinsaku

Visiting Researcher, The Japan Institute for Labour Policy and Training. Research interest: vocational aptitude, occupational information, human resource management.

