

Clustering Occupations

Timothy F. Slaper, Ph.D.: Director of Economic Analysis, Indiana Business Research Center, Kelley School of Business, Indiana University

Which is more important: what we do or what we make?

Which does one hear more about: skills gaps or industry gaps? Know this, if one were to type in “industry gap” in Google, one of the top matches is “industry skills gap.” Given that occupations embody knowledge and skills, it may well be that “what we do” is more important.

A region’s occupational mix may be at least as important as a region’s industrial mix in driving economic performance. Indeed, several years ago, this publication showed that the reason Indiana lagged the nation in terms of personal income was that the state’s occupation mix did not reflect the nation’s mix.¹ Many occupations were over-represented in the Hoosier state while others were under-represented.

Many economic development practitioners (EDPs), as well as policymakers and analysts, are familiar with industry constructs and analysis (the Standard Industrial Code was around from the late 1930s until being replaced with the North American Industrial Classification System in the mid-1990s), but are not as familiar with occupational constructs and analysis. Although it was developed in the late 1970s, the Standard Occupational Classification (SOC) system did not really get the attention it deserved until the 1990s.

Why consider occupation clusters? Isn’t the full list of 923 detailed occupations better for an EDP to understand a region? Why cluster occupations when there are established broader aggregates of 23 occupation groups, or families, as defined by the Bureau of Labor Statistics (BLS) and O*NET? Detailed analysis does require the full set of occupations, but distilling 923 occupations into 34 clusters—as presented below—allows one to

view a region’s occupation profile, or human capital, in one view. Moreover, occupation clusters are superior to job families because occupation clusters are in closer alignment with the types of industries those occupations inhabit.

The purpose of creating and using occupation clusters as well as industry clusters is to develop an additional dimension for analyzing and describing a regional economy.

This methodology is different from the methods commonly used to categorize industry clusters. Identification of industry clusters involves tracing value-chain relationships between industries and businesses (that is, businesses that buy and sell things to each other that they need in order to process and produce products). The occupational mix of a region is based on the BLS occupational employment survey (OES) that is used to determine industry staffing patterns. Staffing patterns are a list of the occupations employed within a particular industry.

One would not be far off the mark to say that the regional presence of industries largely indicates the region’s occupational mix. And the reverse is largely true as well. A region’s occupational mix largely implies the type of industries in greatest concentration in the region. That said, there may be cases for which this does not hold. A generic drug manufacturer may have a materially different staffing pattern than a boutique pharmaceutical manufacturer. The Indiana Business Research Center (IBRC) estimates county-specific occupation counts using staffing patterns and adjusts the occupation estimates using region-specific OES results published by BLS.²

The IBRC recently performed occupational cluster analysis to

“
The 34 occupation clusters allow one to view a region’s occupation profile, or human capital, in one view.
”

continue a research effort to develop a web-based database and analytical framework that would enable EDPs, policymakers and researchers to better understand their region or state. Occupation clusters have the advantage of compressing important information about the detailed occupation definitions (which total 923 in the SOC vintage used here) to make analysis more manageable. The goal is to help users:

- Understand their local workforce and educational situation within the broader regional economic development context
- Understand the associated knowledge and skills that will help local and regional stakeholders to bridge the gap between workforce and economic development when constructing a regional economic development strategy
- Diagnose how well-positioned the region and its communities are to participate effectively in a knowledge-based economy
- Determine a region’s strengths and weakness in terms of knowledge and skills

Analyzing industry concentration, we posit, overshadows occupation

concentration for EDPs and policymakers. Why this is the case when occupations embody the knowledge, skills, and training of the individuals who work for businesses and industries is anybody's guess. In contrast to simply using educational attainment to measure a region's human capital, occupation cluster analysis can offer a deeper understanding into the talent of the regional workforce. Given that globalization is increasingly making borders irrelevant in terms of the movement of talent and human capital, occupation cluster analysis—or perhaps one should call it “human capital cluster analysis”—is particularly valuable.

Global integration has diluted many regional competitive advantages. Many factors of production are increasingly low-cost, be it labor, land, transportation, communications or commodities. Technological know-how knows no national borders. Given this leveling of costs across countries, a region's best chance to differentiate itself is with its brainpower: the education, knowledge, skills and know-how of its workforce.

Markusen and Barbour (2003) emphasized that both industries and occupations are important for understanding complex and changing regional economies, and they have suggested that economic development strategists look into occupation targeting in addition to industrial targeting. They note that whereas industry targeting includes a wish list of industries that regions want to have, occupation targeting could benefit a wide array of the industries that are built around similar occupations.

As an example, they cite engineers in the southern California aerospace/defense industry clusters who found employment opportunities (and the opportunity to create innovations) in other southern California industries, such as sportswear and sports

equipment using exotic materials developed for defense and aerospace.

In later work, Barbour and Markusen (2007) noted the limitations of publicly available data for occupation analysis and proposed to develop an occupation structure for state and metropolitan areas by using the national industry-by-occupation matrix. A major finding was that certain occupations in high-tech industries were distributed quite differently in some metropolitan areas, even when the regions shared a similar industry mix.

One can attribute at least four characteristics to an industry cluster:

1. Geographically bounded concentration of similar, related, or complementary businesses
2. Active channels for transactions and communications among these businesses
3. Shared and specialized infrastructure, labor markets or services
4. Common competitive opportunities and threats

Just as an industry cluster is defined as a collection of industries that are similar or interdependent in certain ways, an occupation cluster shares many similar characteristics.

Feser (2003) and Koo (2005) refined the cluster concept to occupations by including knowledge characteristics of the individual occupations and developing knowledge-based occupation clusters. Feser proposed that these clusters could not only describe the local labor pool, but also serve as inputs in explanatory models of regional growth and change.

The IBRC occupation cluster analysis relies heavily on Feser's (2003), Koo's (2005) and Nolan et al's (2011) previous efforts. That said, Nolan's knowledge-based clusters account for only half of the labor force. As a result, the IBRC extended the analysis to include what is here called the “skill-based” occupations that were not the focus of these earlier works. After all, not all

“
In contrast to simply using educational attainment to measure a region's human capital, occupation cluster analysis can offer a deeper understanding into the talent of the regional workforce.

regions are suitable for development strategies that focus on high-tech, knowledge-intensive occupations. Additionally, to be valuable to EDPs, one needs to embrace developmental strategies that seek to capitalize on existing local and regional skills and expertise.

One of the building blocks in this study was to identify and categorize occupations into clusters based on the Occupational Information Network–Standard Occupational Classification (O*NET-SOC) system. Occupation clusters are groups of occupations that share similar knowledge, skills and other characteristics, such as formal education levels, training, wage levels and availability of benefits.³

O*NET also places an occupation in one of five “job zones.” A job zone is a group of occupations that are similar in terms of the education, experience and on-the-job training

TABLE 1: Knowledge-Based Occupation Clusters

Cluster Number	Knowledge-Based Cluster Titles	Number of Occupations in Cluster	Job Zone Average
01	Arts, Entertainment and Broadcasting Specialists and Management	26	4.2
02	Engineering, Architecture and Related Disciplines	49	4.1
03	Finance, Legal, and Real Estate	22	4.2
04	Health Care: Life and Medical Scientists	18	4.7
05	Health Care: Medical Practitioners and Scientists	29	5.0
06	Health Care: Nurses and Specialized Care Delivery	25	4.5
07	Health Care: Therapy, Counseling and Rehabilitation	13	4.7
08	Information Management and Computing	24	4.2
09	Managerial, Sales, Marketing and Human Resources	27	4.2
10	Mathematics, Statistics, Data Analysis and Accounting	13	4.5
11	Natural Sciences and Environmental Management	40	4.4
12	Postsecondary Education and Knowledge Creation	25	5.0
13	Primary, Secondary and Vocational Education, Remediation and Social Services	29	4.1
14	STEM and Applied Science Technicians	41	3.0
15	Transportation, Logistics and Planning	14	4.1

Source: Indiana Business Research Center

that people need to do the job. Job Zone 1 includes occupations that require little preparation (e.g., parking lot attendants, counter clerks or dishwashers). Job Zone 2 occupations usually require at a minimum a high school diploma, plus some vocational training or job-related coursework. At the other end of the spectrum, Job Zone 5 occupations require advanced communication and organizational skills, as well as specialized knowledge. Job Zone 5 occupations include lawyers, aerospace engineers, physicists and surgeons.

Nolan et al (2011) focused on job zones 3, 4 and 5, with the view that these knowledge-based occupations drive innovation. Fair enough for the purposes of their study, but our goal was to expand the focus.

Data on the knowledge level, type of skills and the extent of training for each occupation (KST) were the basis for our clustering procedure. Abilities, as in the common formula “knowledge, skills and abilities” (KSA), in contrast to knowledge and skills, were not found to differentiate between occupations very well. Following Feser, Koo and Nolan et

al, we used the Ward agglomerative hierarchical clustering algorithm to identify and categorize occupations into clusters.

Ward’s clustering algorithm is commonly used to determine cluster patterns in large multivariate data sets. It minimizes variation based on the differences in measurements for KST within a cluster. One potential weakness of Ward’s clustering algorithm is that the clustering process is sensitive to overly influential observations that “pull” the cluster “center” away from other occupations that would have minimized variances within a particular cluster. Rather than removing the so-called “outliers” from the clustering process, the results were reviewed for consistency and reasonableness. In several cases, occupations were reassigned based on the knowledge component or the industry alignment of the occupation. For example, morticians were shifted from the medical professions to the knowledge-based personal services occupations. As a result, there is a small element of subjectivity and evaluation in the construction of the clusters.

The results of the cluster analysis are presented based on whether the cluster is knowledge-based or skill-based. **Table 1** presents the clusters that are dominated by higher levels of specialized knowledge.

The table also shows the average job zone for the cluster. Except for STEM-related technicians (cluster 14), the clusters average more than 4 on a 5 point scale. In the case of STEM-related technicians, these occupations tend to use specialized or specific knowledge domains, even if the level of education for these occupations may not be as extensive as engineers, computer scientists or financiers.

Table 2 presents the skill-based clusters. In this set of clusters, there are several clusters that support knowledge-based clusters. For example, cluster 21 is “Financial, Legal and Inspection Services, Support.” These occupations would tend to work at the same firms as those in cluster 3, “Finance, Legal, and Real Estate.” The same can be said of the “Administration and Office Support” cluster supporting the work of managers (cluster 9), finance (cluster 3) or university professors

TABLE 2: Skill-Based Occupation Clusters

Cluster Number	Skill-Based Cluster Titles	Number of Occupations in Cluster	Job Zone Average
16	Administration and Office Support	27	2.0
17	Artisans, Craftsman, Designers, including Performance	22	2.6
18	Attendants and General Services	19	1.6
19	Construction Trades	37	1.9
20	Facility, Plant and Large Equipment Operators and Technicians	41	2.0
21	Financial, Legal and Inspection Services, Support	33	2.8
22	Food Preparation and Service	16	1.4
23	Health Care: Therapists, Technicians and Aides	37	2.8
24	Machinists and Skilled Operators and Tenders	22	2.6
25	Managers and First-Line Supervisors	24	2.6
26	Mechanics and Repair Technicians	55	2.7
27	Media, Web Development and Programming	16	2.9
28	Personal Services	16	2.8
29	Production Operators and Tenders	43	2.0
30	Production, General	34	1.7
31	Safety, Security and Emergency	33	3.0
32	Sales, Agents, Brokers and Customer Relations, Support	14	2.5
33	Transportation Equipment Operators	23	2.2
34	Transportation, Logistics and Dispatch, Support	16	1.9

Source: Indiana Business Research Center

(cluster 12), depending on the nature of the firm, school or office.

Table 1 and Table 2 also show the number of occupations in a cluster. The distribution is not even, and it is here that the “science” of the clustering algorithm meets the “art” of deciding the number of clusters. One can decide how many clusters to have ahead of time, but if one decided that 25 clusters was the right number, one may find some odd bundling of occupations. This is especially true for the skill-based occupations that are in the lower job zones because “skills” as O*NET defines and collects data for are more evenly distributed across the general population and occupational landscape. If one limited the number of clusters to 25, for example, it is likely that attendants would be grouped with food preparation and general (unskilled) production workers.

Why use the above occupation clusters, in contrast to the full list of

923 occupations or the 23 occupation groups or families? While important for more detailed analysis, grasping a region’s occupation profile or human capital in one view is better done with a more compressed set of categories.

The downside to the 23 job families of the SOC codes used by BLS is that a job family can be as few as eight occupations—as is the case for both

For the detailed occupations in each cluster, see the appendix at www.ibrc.indiana.edu/ibr/2014/summer/appendix.xlsx.

the legal occupations family and the cleaning and maintenance occupations family—or as many as 109 occupations in the production occupations family. The range of job zones for a job family can be relatively wide as well. Several service-type job

“Occupation clusters allow one to see an additional dimension of a region’s human capital—not just occupations, but also how those occupations may be deployed in industry.”

TABLE 3: Descriptive Statistics for Two-Digit Families of Occupations

2-Digit Family SOC Code	Family of Occupations	Number of Occupations in Family	Family Maximum Job Zone	Family Minimum Job Zone	Family Average Job Zone
11	Management	47	5	3	4.0
13	Business and Financial Operations	46	4	2	3.7
15	Computer and Mathematical	29	5	3	4.0
17	Architecture and Engineering	61	5	2	3.7
19	Life, Physical, and Social Science	59	5	3	4.4
21	Community and Social Service	14	5	4	4.5
23	Legal	8	5	3	4.3
25	Education, Training, and Library	58	5	3	4.6
27	Arts, Design, Entertainment, Sports, and Media	43	4	2	3.3
29	Health Care Practitioners and Technical	83	5	2	4.2
31	Health Care Support	17	3	2	2.6
33	Protective Service	28	4	1	2.7
35	Food Preparation and Serving Related	16	3	1	1.5
37	Building and Grounds Cleaning and Maintenance	8	3	1	1.9
39	Personal Care and Service	32	4	1	2.4
41	Sales and Related	22	4	1	2.7
43	Office and Administrative Support	61	4	1	2.3
45	Farming, Fishing, and Forestry	17	4	1	1.9
47	Construction and Extraction	59	3	1	2.0
49	Installation, Maintenance, and Repair	54	3	1	2.6
51	Production	109	3	1	2.2
53	Transportation and Material Moving	52	4	1	2.3

Source: Indiana Business Research Center

families have occupations that range from Zone 1 to Zone 4.

Table 3 presents some descriptive statistics of the job zones and occupation counts for occupation families, while **Table 4** presents these data for the occupation clusters.

Probably the best reason that occupation clusters are superior to job families is that the occupation clusters are in closer alignment with a type of industry (broadly defined). For example, the management occupations family from BLS and O*NET ranges from business executives to power plant managers to logistics managers to food service managers to clinical research coordinators. The latter occupation appears in the “health care: life and medical scientists”

cluster, while logistics managers fall in the “transportation, logistics and planning” cluster.

Thus, in a way, occupation clusters allow one to see an additional dimension of a region’s human capital—not just occupations, but also how those occupations may be deployed in industry. ■

Notes

1. Timothy F. Slaper and Ryan A. Krause, “Occupational Hazard: Why Indiana’s Wages Lag the Nation,” *Indiana Business Review*, Spring 2010, www.ibrc.indiana.edu/ibr/2010/spring/article1.html.
2. See The Regional Labor Mix tool at www.hoosierdata.in.gov/mix/mix_menu.aspx.
3. O*NET-SOC is developed under the sponsorship of the U.S. Department of Labor’s Employment and Training Administration and is the nation’s primary source of occupational information. The O*NET-SOC taxonomy includes hundreds of

occupations, with each occupation including information on 33 different knowledge variables.

References

- Barbour, E., and A. Markusen. 2007. “Regional Occupational and Industrial Structure: Does One Imply the Other?” *International Regional Science Review* 30 (1): 72-90.
- Feser, E. 2003. “What Regions Do Rather than Make: A Proposed Set of Knowledge-Based Occupation Clusters.” *Urban Studies* 40 (10): 1937-1958.
- Koo, J. 2005. “How to Analyze Regional Economy with Occupation Data.” *Economic Development Quarterly* 19: 356-372.
- Markusen, A. and E. Barbour. 2003. “California’s Occupational Advantage.” Working Paper No. 12. Accessed from International Relations and Security Network, www.isn.ethz.ch/isn/Digital-Library/Publications/.

TABLE 4: Descriptive Statistics for Occupation Clusters

Cluster Number	Occupation Cluster	Number of Occupations in Cluster	Cluster Maximum Job Zone	Cluster Minimum Job Zone	Cluster Average Job Zone
01	Arts, Entertainment and Broadcasting Specialists and Management	26	5	4	4.2
02	Engineering, Architecture and Related Disciplines	49	5	4	4.1
03	Finance, Legal, and Real Estate	22	5	4	4.2
04	Health Care: Life and Medical Scientists	18	5	4	4.7
05	Health Care: Medical Practitioners and Scientists	29	5	5	5.0
06	Health Care: Nurses and Specialized Care Delivery	25	5	3	4.5
07	Health Care: Therapy, Counseling and Rehabilitation	13	5	4	4.7
08	Information Management and Computing	24	5	4	4.2
09	Managerial, Sales, Marketing and Human Resources	27	5	4	4.2
10	Mathematics, Statistics, Data Analysis and Accounting	13	5	4	4.5
11	Natural Sciences and Environmental Management	40	5	4	4.4
12	Postsecondary Education and Knowledge Creation	25	5	4	5.0
13	Primary, Secondary and Vocational Education, Remediation and Social Services	29	5	3	4.1
14	STEM and Applied Science Technicians	41	3	2	3.0
15	Transportation, Logistics and Planning	14	5	4	4.1
16	Administration and Office Support	27	3	1	2.0
17	Artisans, Craftsman, Designers, including Performance	22	3	1	2.6
18	Attendants and General Services	19	2	1	1.6
19	Construction Trades	37	2	1	1.9
20	Facility, Plant and Large Equipment Operators and Technicians	41	3	1	2.0
21	Financial, Legal and Inspection Services, Support	33	3	2	2.8
22	Food Preparation and Service	16	3	1	1.4
23	Health Care: Therapists, Technicians and Aides	37	3	2	2.8
24	Machinists and Skilled Operators and Tenders	22	3	2	2.6
25	Managers and First-line Supervisors	24	3	2	2.6
26	Mechanics and Repair Technicians	55	3	1	2.7
27	Media, Web Development and Programming	16	3	2	2.9
28	Personal Services	16	3	2	2.8
29	Production Operators and Tenders	43	2	2	2.0
30	Production, General	34	3	1	1.7
31	Safety, Security and Emergency	33	4	2	3.0
32	Sales, Agents, Brokers and Customer Relations, Support	14	3	2	2.5
33	Transportation Equipment Operators	23	3	1	2.2
34	Transportation, Logistics and Dispatch, Support	16	3	1	1.9

Source: Indiana Business Research Center

- Nolan, C., E. Morrison, I. Kumar, H. Galloway, and S. Cordes. 2011. "Linking Industry and Occupation Clusters in Regional Economic Development." *Economic Development Quarterly*, 25 (1): 26-35.
- Purdue Center for Regional Development. 2007. "Unlocking Rural Competitiveness: The Role of Regional Clusters." www.statsamerica.org/innovation/report_role_of_regional_clusters_2007.html.