World Bank Group | LinkedIn

DATA INSIGHTS:

JOBS, SKILLS AND MIGRATION TRENDS METHODOLOGY & VALIDATION RESULTS

EXECUTIVE SUMMARY

Tingting Juni Zhu Alan Fritzler Jan Orlowski

WORLD BANK GROUP

LinkedIn

ACKNOWLEDGEMENTS

Authors:

Tingting Juni Zhu (TTL, Private Sector Specialist) and Jan Orlowski (Economist) at Finance, Competitiveness and Innovation Global Practice, World Bank Group, Washington, DC, and Alan Fritzler (Senior Data Scientist) at LinkedIn Corporation, San Francisco, CA, prepared this methodology and validation report. For questions and comments, please contact Tingting Juni Zhu at tzhu@worldbank.org.

People we thank:

We would like to thank the broader team, Ramin Aliyev, Rajan Bhardvaj, J. Ernesto Lopez Cordova, Dina Elnaggar, Elena Gex, Anh Le, Une Lee, Jiemei Liu, Renzo Massari, Jeffrey Mccoy, Moira Mckerracher, Rodimiro Rodrigo, David Satola, Nika Soon-Shiong, Nina Vucenik, and Douglas Zhihua Zeng from WBG, and Hannah Brown, Pei Ying Chua, Nick Eng, Nicole Isaac, Paul Ko, Mariano Mamertino, Di Mo ,Kevin Morsony, Akshay Verma, Tony Vu, and Jenny Ying from LinkedIn for their helpful comments and support of the project. In addition, task team leaders of the World Bank Group's operations helped configure the research and provide feedback during the team's work, including Carli Blunding-Venter (Sub-Saharan Africa), Cesar A. Cancho (Europe and Central Asia), John Gabriel Goddard (Sub-Saharan Africa), Marco Antonio Hernandez Ore (Europe and Central Asia), and Marcin Miroslaw Piatkowski (East Asia and Pacific). We would also like to acknowledge the comments and feedback from Ana Paula Cusolito, Mary C. Hallward-Driemeier, Victoria Levin, Espen Beer Prydz, and Hernan Winkler. Finally, we would like to thank the seed funders of this project in a big data competition at the World Bank Group exploring nontraditional datasets for public policy-making in developing countries: Prasanna Lal Das, Trevor Monroe, Victor Mulas, and Klaus Tilmes. Without their entrepreneurial spirit, the project would not have become what it is today.

ABBREVIATIONS

BLS	U.S. Bureau of Labor Statistics
ECA	Europe and Central Asia
EAP	East Asia and Pacific
ICT	Information and Communications Technology
ILO	International Labor Organization
ILOSTAT	International Labor Organization Statistics
ISIC	International Standard Industrial Classification
LAC	Latin America and the Caribbean
MENA	Middle East and North Africa
OECD	Organization for Economic Cooperation and Development
PIAAC	Program for the International Assessment of Adult Competencies
PS-TRE	Problem solving in technology-rich environments
NAC	North America
SA	South Asia
SSA	Sub Saharan Africa
WBG	World Bank Group

EXECUTIVE SUMMARY

The World Bank Group-LinkedIn partnership pilots the use of private company data for generating insights on development trends. This partnership is a three-year effort between the WBG and LinkedIn to investigate the extent to which LinkedIn's data can inform policy (figure 0-1). The first phase of the partnership evaluates LinkedIn data covering 100+ countries with at least 100,000 LinkedIn members, distributed across 148 industries and 50,000 skills categories. The second and third phase focus on automating and scaling insights, and expanding joint research.

This methodology report describes the construction and validation of metrics on skills, industry employment, and talent migration in over 100 countries. This report has three objectives: (1) document the characteristics and coverage of LinkedIn data; (2) report the methods used to develop new metrics; and (3) showcase examples of policy questions that can be answered with this non-traditional data (figure 0-2). Because this is the first time that LinkedIn has shared a nontraditional dataset with a third-party organization globally as a public good (strictly unremunerated), it is important that we make these methodology and validation results available so that researchers and policy-makers can build on this initial effort by the WBG and LinkedIn.

The metrics generated from LinkedIn's data differ from traditional government indicators in important ways. As new development opportunities emerge, especially in the digital economy around the globe, WBG is seeking new data sources that can capture the latest development trends. Traditional government surveys often cannot keep up with this demand. Making LinkedIn real-time data available for development use, especially in developing countries, can be useful for policy-makers. For example, LinkedIn data can help answer pressing questions such as "What skills are gained or lost in association with talent migration in my country?" and "What are the most recent sectoral employment trends, and which skills are most relevant to them?" Nonetheless, because of the granularity and sheer amount of user-generated data, the industry and skills classifications that LinkedIn taxonomy uses are not standard and may not always conform to commonly used standards such as the International Standard Industrial Classification (ISIC); European Skills/Competences, Qualifications, and Occupations (ESCO); and the Occupational Information Network (O*NET). Part of the contribution of this methodology report is to match LinkedIn's taxonomies to these international standards to allow for easier matching of LinkedIn data with external datasets for further analysis. These efforts are central to the continued use of LinkedIn data as a valuable complement to traditional data sources.

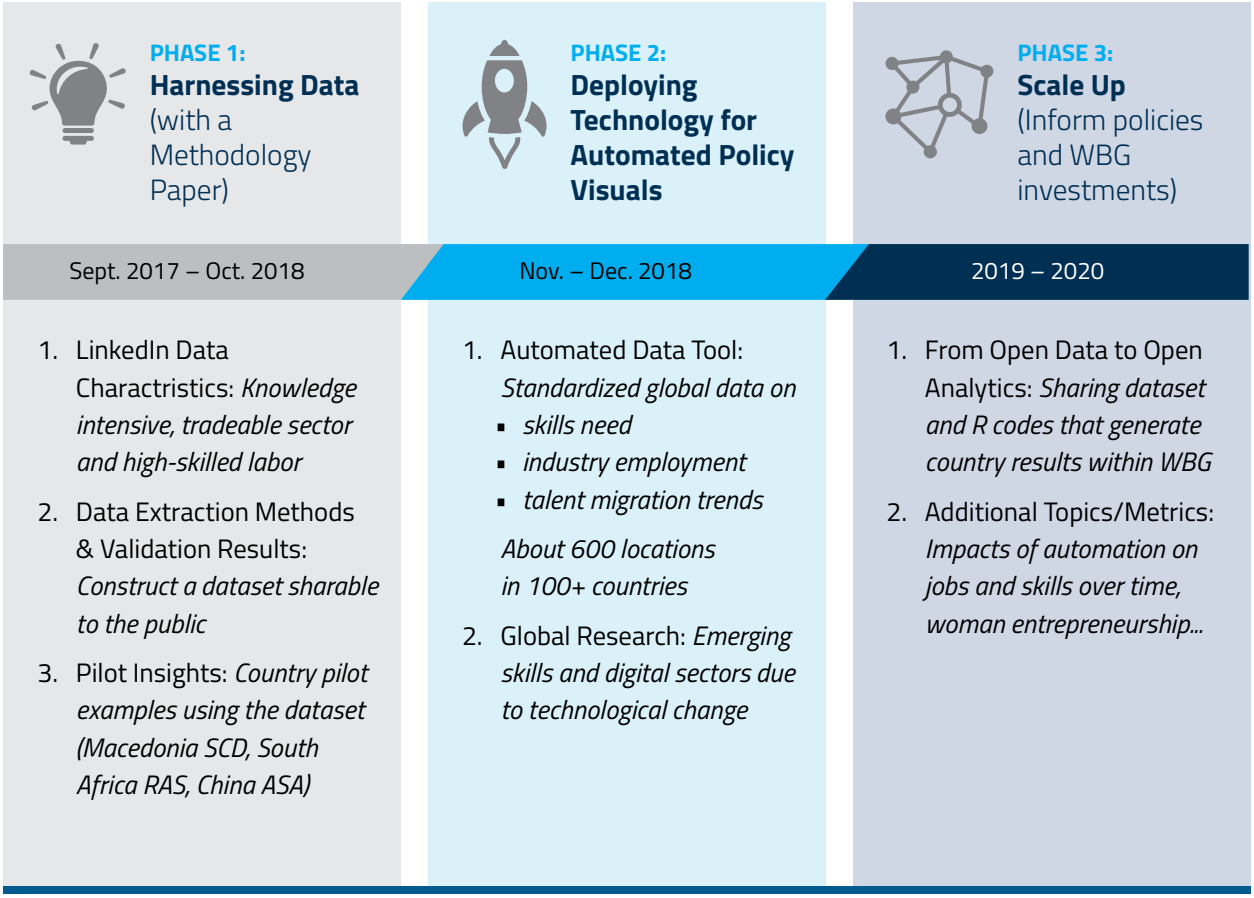
LinkedIn data are best at representing skilled labor in the knowledge-intensive, and tradable sectors. The LinkedIn metrics were compared and validated against 23 internationally standardized data sources on industry, skill, and migration trends.[1] Although LinkedIn may have better coverage in developed than developing countries, there are certain knowledge-intensive and tradable sectors, such as information and communication; professional, scientific, and technical activities; financial and business services; arts and entertainment; manufacturing; and mining and quarrying, that have good LinkedIn coverage globally (figure 0-3).[2] This allows for benchmarking of performance across locations globally in these six sectors.

1 See Table II-3 and Appendix E (Migration), Table II-2 (Skills), and Table III-1 (Industry Employment) for all the external data sources that the team evaluated

2 The strong LinkedIn coverage of the mining and quarrying sector is partially due to companies on LinkedIn incorrectly identifying themselves as oil and energy companies rather than as utilities and hence being misclassified in ISIC sector B instead of D. An example of this is EDF Energy in the United Kingdom. See section II-C-1 (Industry Coverage Globally). Manufacturing has significantly lower coverage, however it is a an important tradable sector for inclusion.

FIGURE 0-1:

World Bank Group (WBG)-LinkedIn Collaboration Schedule



In addition to certain sectoral skewness, young, skilled individuals with at least a bachelor's degree are more likely than those with less education to be on LinkedIn, and women are more likely to be captured in LinkedIn than national statistics. In general, although LinkedIn data are not representative of the entire economy and are self-reported, they can uniquely capture segments of the economy that are among the most innovative, dynamic and high-value add. In addition, because these data are updated more frequently than traditional government statistics, they have the unique ability to capture the latest employment and industry skills needs, which government statistics often miss—especially in the digital and disruptive technology sectors. Industry employment, skills, and talent migration metrics comprise the first phase of this partnership (table 0-1).

FIGURE 0-2:

Three Objectives of this Methodology Report

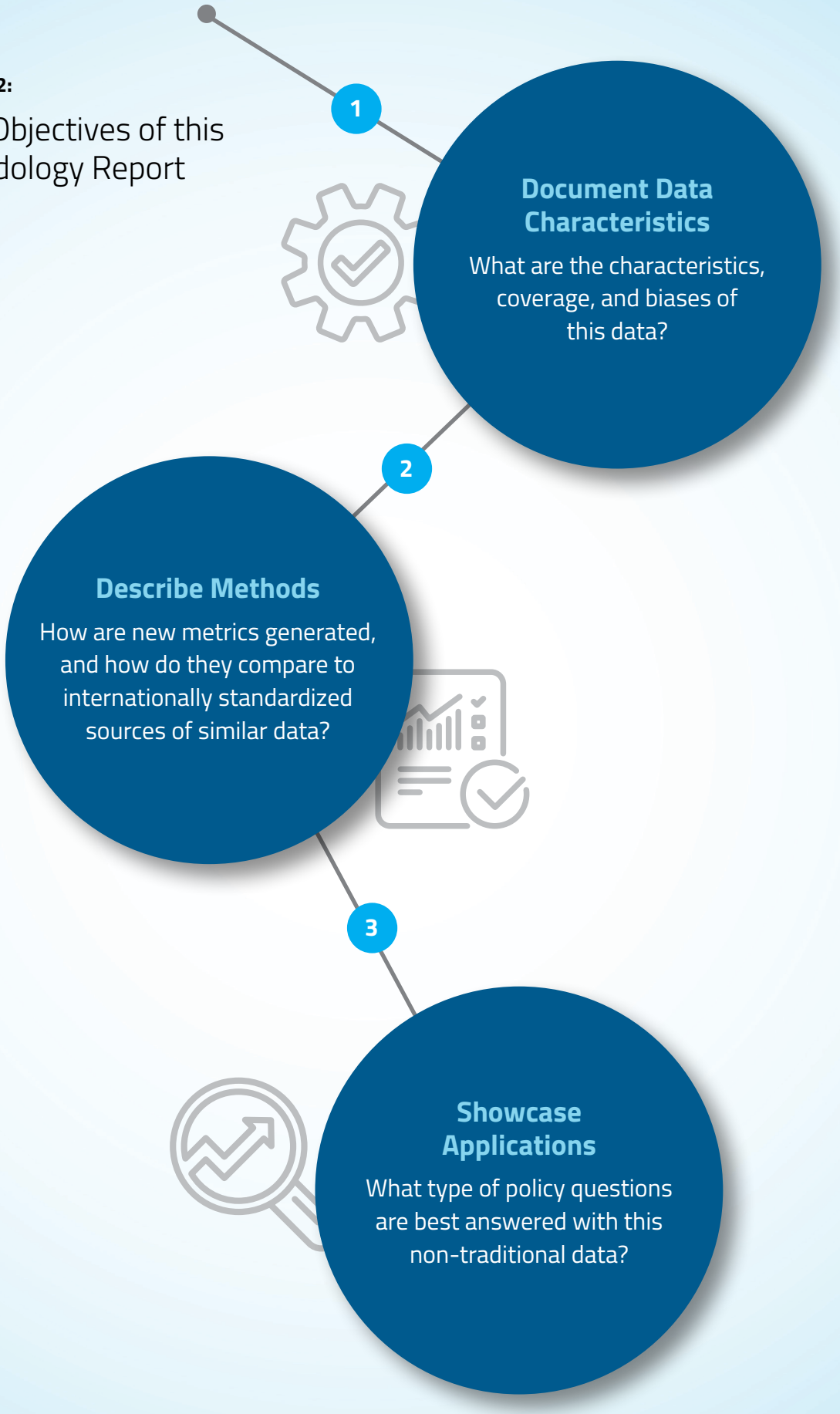
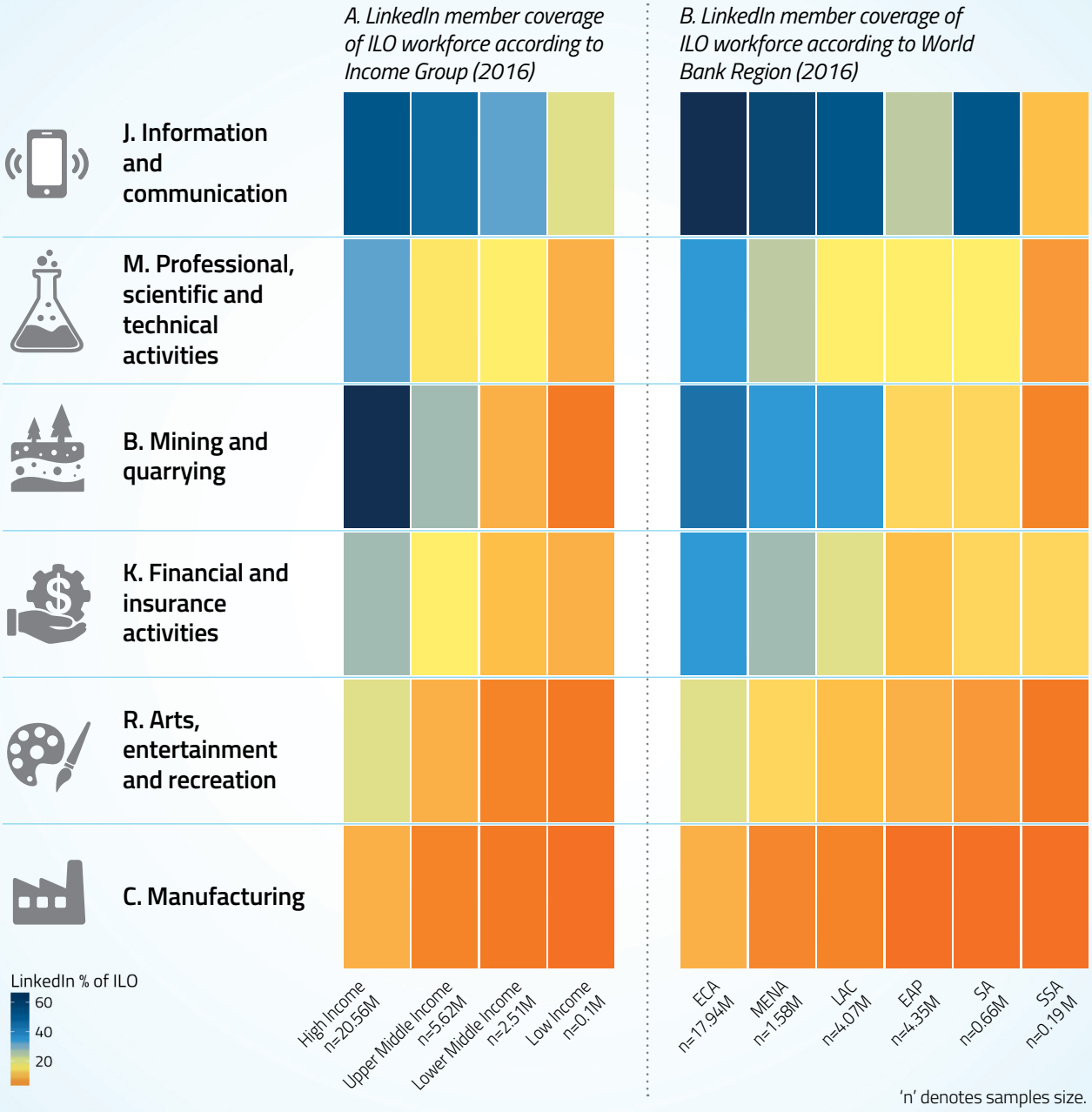


FIGURE 0-3:

LinkedIn Industry Coverage According to (A) Income Group and (B) World Bank Region



Note: See Section III.C for more information on LinkedIn industry representativeness. Because of lower penetration rates of some sectors, the first phase of the World Bank Group-LinkedIn collaboration will share data only from the six knowledge-intensive and tradable sectors to ensure data quality and minimize risks of misinterpretation of the LinkedIn data due to small sample size; the remaining sectors not shown are: : L. Real estate activities; D. Electricity; gas, steam and air conditioning supply; N. Administrative and support service activities; P. Education; O. Public administration and defense; compulsory social security; S. Other service activities; Q. Human health and social work activities; H. Transportation and storage; G. Wholesale and retail trade; repair of motor vehicles and motorcycles; F. Construction; I. Accommodation and food service activities: A. Agriculture; forestry and fishing.

Source: Authors' calculation using LinkedIn and International Labor Organization (ILO) data in 92 countries

TABLE 0-1:

Summary of Metrics: Methodology and Validation Results

METRIC NAME	METHOD TO DERIVE THE METRICS[a]	CONFIDENCE LEVEL (REASONS)
1) Industry Employment		
Industry employment location quotient (LQ) Captures the employment size of an industry in a particular locale, relative to the same industry in other locales.	For a given country *c*, industry *i*, and time *t*, $Industry\ LQ_{c,i,t} = \frac{industry\ size_{c,i,t}}{average\ industry\ size\ of\ all\ countries\ in\ country\ c's\ income\ group_{i,t}}$ where with industry size measured as a relative term: $Industry\ size_{c,i,t} = \frac{member\ count_{c,i,t}}{member\ count_{c,t}}$	High (good global coverage, good validation results)
Industry employment growth[b] Captures the transitions among industries over time by LinkedIn members as a proxy for industry employment growth. Based on the industries declared by the companies in a member's work history.	Growth is given as rate of employment-level change (e.g., 2015-2017) for country *c* and industry *i*, $Industry\ growth_{c,i} = \frac{member\ count_{c,i,t+1} - member\ count_{c,i,t}}{member\ count_{c,i,t}} * 100$	Medium (good global coverage, good validation results but external data source covers only specific countries)
2) Skills		
Industry skills needs Captures the most-distinctive, most-represented skills of LinkedIn members working in a particular industry. Based on the skills section of the LinkedIn profile.	For each country, the weight ($w_{i,s}$) denotes how distinctive and representative each skill s is in industry *i* as: $w_{i,s} = m_{i,s} * \ln\left(\frac{N}{n_s}\right)$ with $m_{i,s}$ indicating the number of members in industry *i* having skill *s*, *N* the total number of industries, and n_s the total number of industries having skill *s*. The first term gives greater weight to skills that have high membership penetration, and the second term gives less weight to "common" skills that appear in all industries (e.g., Microsoft Office). In this sense, the most important skills for each industry are those that have high member penetration but are also unique.	Medium (good global coverage for knowledge-intensive and tradable sectors, good validation results but external data source covers only specific countries)

continues

TABLE 0-1: continued

Skill penetration Measures the time trend of a skill across all occupations within an industry. Based on skill addition rates, and the number of times a particular skill appears in the top 30 skills added across all of the occupations within an industry.	There are four steps to compute skill penetration: 1. Use the industry skills needs framework above to calculate the weight for each skill *s* for each occupation *o* in industry *i*: $$w_{i,o,s} = m_{i,o,s} * \ln\left(\frac{N}{n_s}\right)$$ 2. Construct a list of the 30 top represented skills for each occupation *o* in industry *i*, based on the values of $w_{i,o,s}$: $$\lfloor (s_1, w_1), (s_2, w_2) \cdots, (s_{30}, w_{30}) \rfloor$$ 3. Calculate the skill group penetration rate at the occupation-industry level $p_{i,o,S}$ by counting the number of skills *s* belonging to each skill group S and dividing by 30: $$p_{i,o,S} = \frac{\sum_{s=1}^{30} s \in S}{30}$$ 4. Get the average skill group S penetration rate $p_{i,S}$ across all occupations *o* for the industry *i*: $$\bar{p}_{i,S} = \frac{\sum_{o=1}^{n} p_{i,o,S}}{n_i}$$	Medium (good global coverage for knowledge-intensive and tradable sectors, good validation results but external data source covers only specific countries)
3) Talent migration		
Inter- and intra-country talent migration Based on user-reported location. When a user's updated job location is different from their former location, LinkedIn recognizes this as a physical migration.	Given as net migration, with country *a* the country of interest, and country *b* the source of inflows or destination of outflows, at time *t*, $$Net\ migration_{a,b,t} = \frac{net\ flows_{a,b,t}}{member\ count_{a,t}} * 10{,}000$$ (net flows = arrivals – departures)	High (good global coverage for knowledge-intensive and tradable sectors, good validation results)
Migration – industries gained and lost Based on the industry associated with a member's company at the time of migration.	Given as net migration, with country *a* the country of interest and country *b* the source of inflows or destination of outflows, both considered for a given industry *i* at time *t*, $$Net\ industry\ migration_{a,b,i,,t} = \frac{net\ industry\ flows_{a,b,i,,t}}{member\ industry\ count_{a,i,t}}$$ (This formula is used to calculate the top gaining and losing industries associated with talent migration flows.)	Low (good global migration data for knowledge-intensive and tradable sectors, but migration industry movements have no comparable global external data for validation)

continues

TABLE 0-1: continued

Migration – skills gained and lost Based on the skills associated with a member's profile at the time of migration.	Given as net migration, with country *a* the country of interest and country *b* the source of inflows or destination of outflows, both considered for a given skill *s*, at time *t*, $$Net\ skill\ migration_{a_s,b_s,t} = \frac{net\ skill\ flows_{a_s,b_s,t}}{member\ skill\ count_{a_s,t}}$$ (This formula is used to calculate the top gaining and losing skills associated with talent migration flows.)	Low (good global migration data for knowledge-intensive and tradable sectors, but skills migration has no comparable global external data for validation)

Notes: Confidence level is evaluated against two criteria: 1) *global coverage* (High: good for global, Medium: good only for certain sectors, Low: limited coverage at the moment but expected to improve over time as LinkedIn membership grows and diversifies, and hence worth including in the dataset and dashboard) and 2) *validation results against other independent data sources* (High: highly positively correlated with various government or international organization data sources, Medium: highly positively correlated with one other source that has data on a specific region or country only, Low: the project team was unable to find a comparable dataset for validation). This last point also demonstrates the value of LinkedIn data in that they expand the information available on the topic and can be complementary to traditional survey or administrative data and low confidence level is not a reflection of the quality of the metric.

a All metrics at the city level were calculated in the same manner as at the country level, except for Industry location quotient, because we did not have city-level income for calculation; instead we used country average for the denominator—how a city compares with its own country average.

b Because of rapid LinkedIn membership growth around the globe, the team constructed the balanced panel data to isolate LinkedIn membership growth from industry employment growth, so the growth rate captured here is an employment transition rate for experienced employees who report jobs on the LinkedIn platform across years. For details, see Section IV-A-2.

To protect user privacy and permit comparability of metrics, LinkedIn metrics are normalized. Because user behavior is different in different countries (e.g., overreporting of work experience; not updating profile if unemployed; LinkedIn membership growing exponentially in developing countries and hence the data potentially capturing LinkedIn business growth instead of industry headcount growth), in addition to validating against other data sources, we used statistical methods to normalize and standardize metrics to ensure they can be compared fairly across countries and industries. For example, we normalized most metrics according to LinkedIn membership size in each country so that countries with more workers on LinkedIn did not artificially rank higher.

Based on feedback from three World Bank Group pilot projects in South Africa, Macedonia, and China, sample policy questions that LinkedIn metrics can answer are listed in table 0-2. In addition to determining descriptive trends, another useful application of the LinkedIn metrics is to triangulate across the three categories of metrics. For example, to nurture certain growing industries, one can further explore what skills are needed or whether there is a risk of talent outflow. Furthermore, to conduct analytical and empirical research, the datasets are structured so they can be easily merged with external data sources. For instance, because all the LinkedIn data on industries made available through this partnership are equivalent to the two- to three-digit ISIC level, and the project team has mapped these LinkedIn industry classifications against ISIC 4 standards, merging industry employment and skills needs data with data from economic censuses, such as wage and productivity data, can help in understanding private sector growth and the productivity and human capital components that drive that growth.

TABLE 0-2:

Sample Policy Questions Using LinkedIn Metrics

METRIC NAME	SAMPLE POLICY QUESTIONS
1) Industry employment	
Industry employment location quotient	Which industries are more concentrated in my country or city than in an average country in the same income group?
Industry employment growth	What are the most recent employment growth trends in my country or city, especially in knowledge-intensive and tradable sectors?
2) Skills	
Industry skills needs	For the industries I am interested in, what are the latest, most important skills?
Skill penetration	Are particular skills (e.g. Artificial Intelligence) being applied across industries ? How is this changing over time?
3) Talent migration	
Inter- and intra-country talent migration	Am I (net) losing talent? With which countries do I compete for talent?
Migration – industries gained and lost	To which industries are these talents moving?
Migration – skills gained and lost	What skills are gained or lost in association with talent migration?

To further demonstrate how the above metrics can be used to inform policies for World Bank projects, we provide some sample visuals in this report. One is the top growing and declining sectors globally in 100+ countries (figure 0-4). Emerging sectors, such as renewables and environment and Internet have registered rapid employment growth in the past three years, whereas newspaper and outsourcing are in decline in countries in all income groups. This type of insight can be generalized across World Bank regions or specified to a particular country as well (see Section V: Sample Visual Outputs and Country Applications).

FIGURE 0-4:

Growth from Industry Transitions according to Income Group

Annual Average 2015-2017

Avg. Growth Rate 3Yr Avg

-4% 4%

ISIC Section Index	ISIC Section Name	Industry Name	High Income	Upper Middle Income	Lower Middle Income	Low Income
B	Mining and quarrying	Mining & Metals	0.3%	0.7%	-0.5%	0.1%
		Oil & Energy	-0.7%	-0.5%	-0.7%	-1.1%
C	Manufacturing	Aviation and Aerospace	1.5%	1.2%	0.1%	-1.4%
		Renewables and Environment	1.4%	1.5%	1.4%	2.0%
		Pharmaceuticals	1.2%	0.6%	0.2%	-0.5%
		Automotive	1.0%	0.0%	0.0%	-0.5%
		Industrial Automation	1.0%	-0.6%	-0.6%	
		Packaging and Containers	1.0%	1.0%	0.4%	
		Glass Ceramics and Concrete	0.8%	0.1%	-1.4%	
		Chemicals	0.8%	1.1%	0.0%	0.5%
		Plastics	0.7%	0.6%	0.5%	
		Machinery	0.6%	-0.4%	0.0%	-0.4%
		Paper & Forest Products	0.4%	-0.2%	0.8%	
		Shipbuilding	0.4%	-0.3%	-1.9%	
		Food Production	0.1%	0.1%	0.2%	0.2%
		Electrical and Electronic Manufacturing	0.0%	-0.2%	0.0%	-0.2%
		Textiles	0.0%	-0.2%	-0.2%	
		Railroad Manufacture	0.0%	-2.5%	3.1%	
		Printing	-0.1%	-0.2%	-0.1%	
J	Information and communication	Computer and Network Security	3.9%	0.9%	-0.7%	0.3%
		Internet	3.1%	1.4%	1.0%	-0.1%
		Computer Software	1.8%	0.9%	-0.4%	-0.4%
		Computer Games	1.3%	-1.3%	-1.7%	
		Wireless	1.2%	-1.0%	-1.3%	
		Information Technology and Services	0.5%	-0.2%	-0.5%	-0.5%
		Writing and Editing	0.5%	0.2%	0.6%	
		Computer Networking	0.2%	-0.2%	-0.9%	
		Online Media	0.2%	0.8%	0.1%	
		Motion Pictures and Film	-0.1%	-0.5%	0.7%	
		Semiconductors	-0.3%	-1.1%	-2.9%	
		Computer Hardware	-0.5%	-1.4%	-0.6%	
		Media Production	-0.5%	-0.9%	-0.3%	-0.3%
		Broadcast Media	-0.5%	-0.6%	-0.8%	0.0%
		Telecommunications	-1.0%	-1.2%	-0.8%	-0.5%
		Publishing	-1.2%	-1.7%	-0.7%	-0.3%
		Newspapers	-2.4%	-2.6%	-1.8%	
K	Financial and insurance activities	Venture Capital and Private Equity	4.0%	2.0%	3.2%	
		Investment Management	2.7%	0.8%	1.8%	1.4%
		Capital Markets	1.6%	1.3%	0.2%	
		Financial Services	1.2%	0.4%	0.8%	0.7%
		Insurance	0.9%	0.1%	0.3%	-0.3%
		Banking	0.3%	-0.3%	-0.1%	0.3%
		Investment Banking	-0.1%	-0.1%	0.3%	
M	Professional scientific and technical activities	Biotechnology	2.4%	1.2%	0.0%	
		Alternative Dispute Resolution	1.9%			
		Executive Office	1.6%	1.0%	-0.1%	
		Management Consulting	0.9%	0.2%	0.3%	0.3%
		Information Services	0.8%	-0.2%	0.5%	
		Veterinary	0.7%	-0.2%	-1.1%	
		Translation and Localization	0.7%	-0.4%	0.2%	
		Professional Training & Coaching	0.7%	0.1%	-0.3%	-0.1%
		Environmental Services	0.6%	-0.1%	0.4%	0.1%
		Design	0.6%	-0.5%	-0.1%	-0.9%
		Nanotechnology	0.3%	-1.5%		
		Photography	0.3%	0.4%	0.2%	
		Marketing and Advertising	0.3%	-0.3%	0.0%	0.5%
		Architecture & Planning	0.3%	-0.7%	-0.1%	0.7%
		Legal Services	0.3%	-0.3%	0.1%	0.2%
		Graphic Design	0.2%	-0.3%	0.2%	-0.3%
		Mechanical or Industrial Engineering	0.2%	-0.1%	-0.6%	-0.2%
		Law Practice	-0.2%	-0.3%	-0.2%	-0.3%
		Events Services	-0.8%	-0.6%	-0.6%	
		Accounting	-1.0%	-2.0%	-1.7%	-1.1%
		Public Relations and Communications	-1.0%	-1.5%	0.2%	-0.8%
		Research	-1.1%	-0.9%	-1.2%	-1.2%
		Market Research	-2.4%	-2.1%	-1.2%	-1.7%
		Outsourcing/Offshoring	-3.6%	-1.6%	-2.1%	-1.3%
R	Arts, entertainment and recreation	Gambling & Casinos	2.1%	-0.6%	1.9%	
		Animation	1.8%	1.3%	0.9%	
		Health Wellness and Fitness	1.1%	0.3%	0.3%	0.5%
		Arts and Crafts	0.4%	0.1%	0.4%	
		Fine Art	0.3%	-0.6%	0.2%	
		Sports	0.3%	0.0%	0.5%	
		Libraries	0.0%	0.9%	-1.0%	
		Entertainment	0.0%	-1.3%	0.4%	-0.3%
		Music	0.0%	-0.1%	-0.1%	
		Museums and Institutions	-0.7%	-1.1%	0.8%	
		Performing Arts	-0.8%	-0.2%	-1.1%	

-5% 0% 5%

Note: Industries where N<5 countries are removed

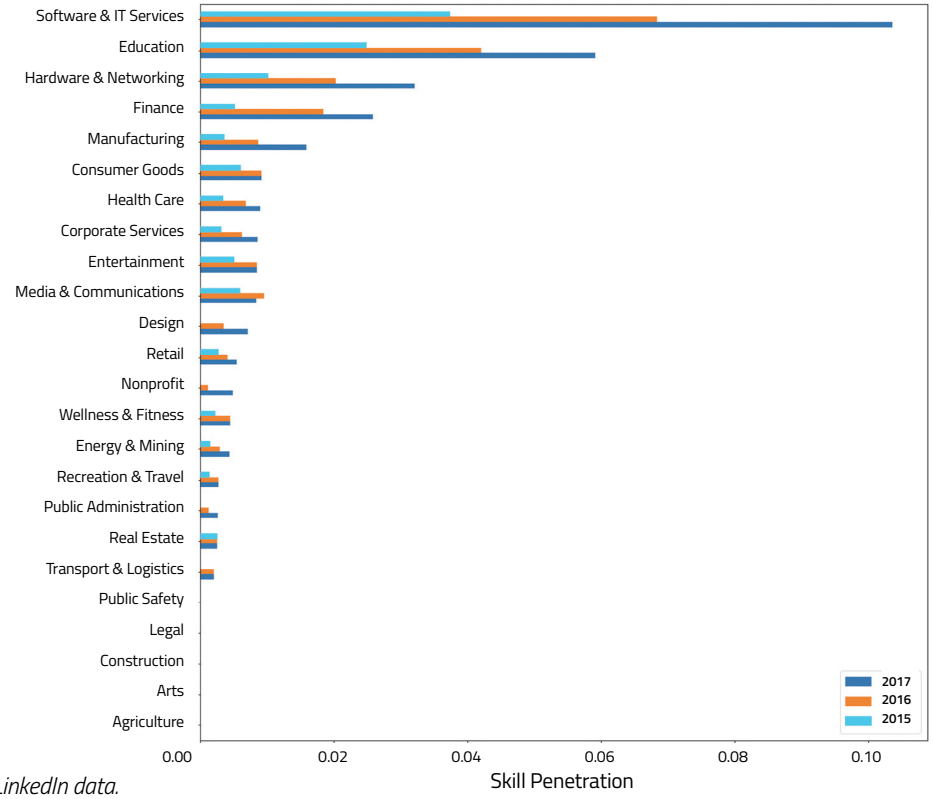
Source: Authors' calculation using LinkedIn data.

Another value that the LinkedIn metrics add is in the emerging skills and industries that official statistics often do not capture. LinkedIn's skill metrics allow the World Bank Group to measure how new technologies—such as artificial intelligence—are spreading across industries and changing labor markets around the globe. For example, artificial intelligence skills are among the fastest-growing skills on LinkedIn, with a 190% increase from 2015 to 2017 across all industries (figure 0-5).

The current round of technological advancement (aka Industry 4.0) seems more pervasive than the previous rounds and is being transmitted to developing countries more quickly. Around the globe, disruptive technology skills have appeared in many developing countries in the past three years, although typically "human" skills (e.g., those related to sociobehavioral characteristics, interpersonal communication, and cognitive skills) are also on the rise (figure 0-6).

FIGURE 0-5:

Global Artificial Intelligence Skill Penetration 2015-2017



Source: Authors' calculation using LinkedIn data.

FIGURE 0-6:

Skills with the Largest Increase in Penetration Across Industries 2015-2017

1. **Leadership**
2. **Development Tools**
3. **Oral Communication**
4. **Web Development**
5. **Business Management**
6. **Digital Literacy**
7. **People Management**
8. **Data Science**
9. **Graphic Design**
10. **People Management**

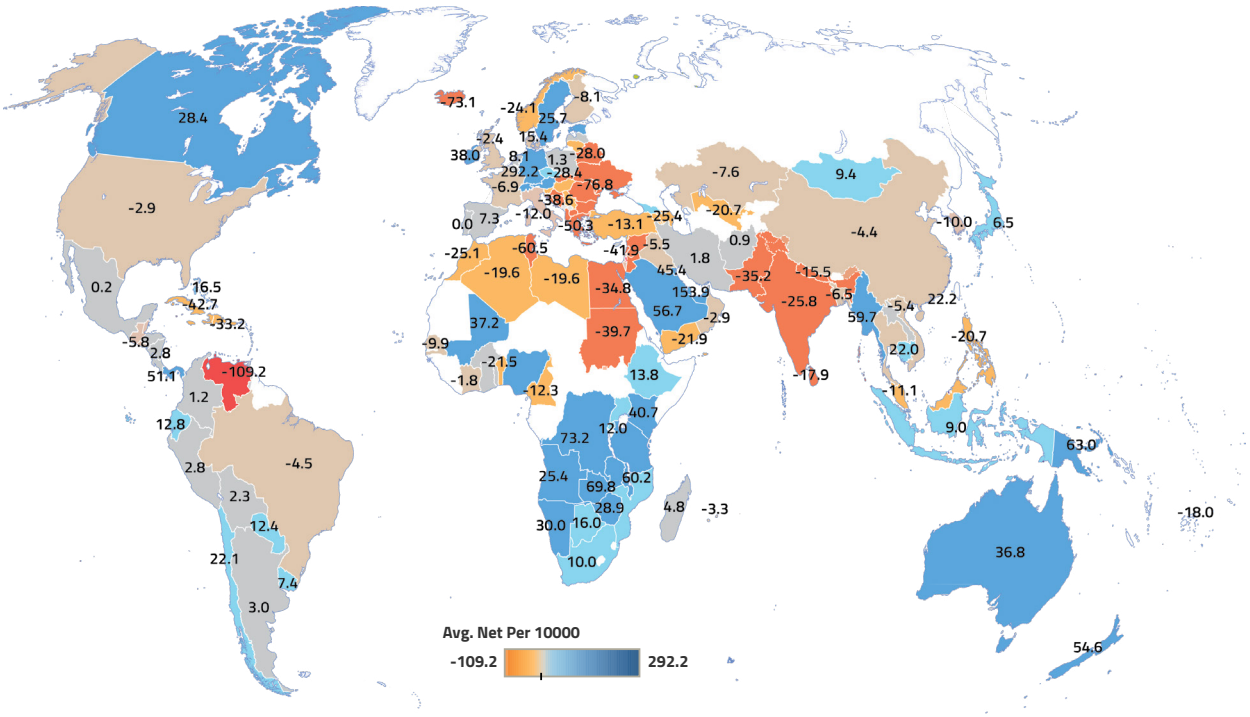
Source: Authors' calculation using LinkedIn data.

Near-real-time global talent migration trends can also be captured through LinkedIn data to allow developing country policy-makers to assess the health of their countries' talent pipelines. The Middle East and North Africa, Latin America and the Caribbean, and South Asia have seen the greatest talent loss in recent years, whereas Organization for Economic Cooperation and Development (OECD) countries such as Australia, New Zealand, and Canada are attracting the most talent (figure 0-7).

All the visuals will be automated and updated annually[3] until June 2020 under this three-year WBG-LinkedIn partnership on linkedindata.worldbank.org. The underlying dataset, as well as other resources that are helpful for policy-makers around the world, will also be updated and made available for free at the same URL as a public good.[4] Subject to demand and user feedback, more metrics may be added later.

FIGURE 0-7:

Global Talent Migration 2015-2017



Source: Authors' calculation using LinkedIn data.

3 There will be a minimum of an annual refresh by LinkedIn. The online visuals can be updated more frequently if there is strong user demand

4 The aggregated datasets and visuals are available to all for the public good under the Creative Commons Attribution 3.0 IGO license with attribution to both LinkedIn Corporation and the World Bank Group. The World Bank Group and LinkedIn Corporation (including its affiliates) do not take responsibility and are not liable for any damage caused through use of data and insights through this website, including any indirect, special, incidental or consequential damages.

WORLD BANK GROUP

LinkedIn

November 2018