Off-the-Job Learning in Cities*

Atsushi Yamagishi†
Princeton

July 24, 2023

Abstract

I document novel stylized facts on learning and its relationship with population density. I use Japanese survey data that provide distinctively rich first-hand information about the frequency, purpose, subject, and method of off-the-job learning. First, people learn more frequently in denser cities. Second, people in denser cities are more likely to learn to gain new employment or cultivate themselves. Third, what people learn by which method is related to the local demand for skills and the local supply of learning opportunities. I discuss the implications of these findings for theory and evidence of urban agglomeration economies.

Keywords: Off-the-job learning, Agglomeration economies, Urban earnings premium, Skills, Human capital accumulation

JEL classification: J31, J62, R11, R12, R20

*I am grateful to Steve Redding and Leah Boustan for their guidance. I am also thankful to Ikuto Aiba, Ellora Derenoncourt, Shota Fujishima, Keisuke Kawata, Tomoya Mori, Mariko Nakagawa, Kentaro Nakajima, Hayato Nishi, Yasuhiro Sato, Kohei Takeda, Xuanli Zhu, and participants of seminars and conferences for their helpful comments. I acknowledge permission to access the microdata of the Japanese time-use survey (*shakai seikatsu kihon chosa*), granted by the Ministry of Internal Affairs and Communications. I thank Yasuhiro Sato for his help in obtaining said permission. I acknowledge financial support from the Funai Foundation for Information Technology. All errors are mine.

†Department of Economics, Princeton University. E-mail: ayamagishi@princeton.edu.

1 Introduction

Learning has been identified as a key urban agglomeration force that justifies higher wages in cities. Indeed, Duranton and Puga (2004) has listed learning, matching, and sharing as the three key microfoundations of urban agglomeration economies. The idea that people learn faster in cities has featured in various theoretical models (e.g., Glaeser 1999; Davis and Dingel 2019; De la Roca, Ottaviano and Puga 2023). Empirically, faster wage growth in cities is often interpreted as empirical evidence consistent with faster learning in cities (e.g., Glaeser and Maré 2001; Baum-Snow and Pavan 2012; De la Roca and Puga 2017; Eckert, Hejlesen and Walsh 2022). However, indirect inference of the importance of learning from wage dynamics remains somewhat inconclusive because alternative mechanisms, such as richer job availability and better labor market matching (e.g., Dauth, Findeisen, Moretti and Suedekum 2022; Eckert et al. 2022; Papageorgiou 2022), can generate a similar pattern. Furthermore, wage information does not reveal additional details about learning, such as frequency, purpose, subject, and learning method. Direct observation of learning behavior would provide more detailed and transparent evidence regarding learning in cities. However, this has been challenging because learning behavior is not typically observed in data.

This paper presents a set of novel stylized facts on the relationship between learning and population density, using a distinctive Japanese survey that collects detailed first-hand information on off-the-job learning.[1] The richness of the data allows me to answer the following questions: (i) Do people learn more frequently in denser cities? (ii) Does learning purpose differ in denser cities? (iii) What skills are people in denser cities more likely to learn and using which methods? Due to the lack of panel data and a natural experiment, I keep the analysis descriptive and do not claim causality in the main part of this paper.[2]

The answers to the above three questions are as follows, each of which has implications for urban agglomeration economies. First, people spend more time on off-the-job learning in denser cities: A 1% increase in population density is associated with a 0.1% increase in the number of days spent learning. This suggests that the faster wage growth in cities might be partially attributed to more efforts of the residents. From the productivity viewpoint, more time spent on learning can be a new source of the widely-documented productivity advantages of denser cities (Rosenthal and Strange 2004).

Second, people in denser cities are more likely to learn to obtain a new job or cultivate themselves. Learning for new jobs is consistent with the evidence that more job opportunities are available in cities (e.g., Papageorgiou 2022) because the presence of more attractive job offerings in larger cities would provide a stronger incentive for workers to acquire the necessary skills to

[1] A distinct strand of literature in labor economics (e.g., Lynch 1992; Gaulke 2021) has analyzed off-the-job training, but its relationship with urbanity has not been investigated. I also investigate new aspects of learning, such as frequency and purpose.

[2] That said, Section 4 assesses how well my estimates approximate causality using several available approaches. Results suggest that my descriptive estimates might be close to causality.

take them.[3] This also implies that matching with better-paying jobs in larger cities (Eckert et al. 2022) might be partially driven by workers' voluntary efforts to improve matching quality. This highlights the potential complementarity between matching and learning in explaining urban agglomeration economies, two of the three key agglomeration forces in Duranton and Puga (2004). I also find that people are less likely to learn skills to use them at their current job, further highlighting the importance of new job opportunities. Learning for cultivation implies more cultural activities in denser cities, which is consistent with the agglomeration of cultural activities or creative industries (e.g., Mitchell 2019; Tao, Ho, Luo and Sheng 2019; Borowiecki and Dahl 2021). This may relate to cities' endogenous amenities in that agglomeration may improve neighborhood quality through cultural amenities and residents' cultural sophistication (Shapiro 2006; Diamond 2016).

Finally, what people learn using which methods is associated with local demand for skills and local supply of learning opportunities. Regarding what people learn, while people in denser cities are generally more likely to learn all skills, an exception is elderly care, which is more likely to be learned in *less* dense cities. This is consistent with the rapid population aging in rural Japan because the high demand for elderly care skills could dominate and overturn the general tendency of more frequent learning in denser cities. Thus, people seem to respond to local demand conditions for skills. Regarding learning methods, though inconclusive, people in denser cities are more likely to use education-market-based learning methods that require physical attendance, such as privately-provided classes and vocational schools. This is consistent with the view that cities provide more learning opportunities since private providers of learning opportunities can enjoy agglomeration economies, which seem strong in the service sector (e.g., Glaeser, Kolko and Saiz 2001; Morikawa 2011; Leonardi and Moretti forthcoming).

An important qualification is that my data do not contain information on on-the-job learning.[4] Although I am unaware of a study that directly observes on-the-job learning and investigates its relationship with city size, Charlot and Duranton (2004, 2006) employs distinctive cross-section data from France and analyzes workplace communication, which is an important medium of on-the-job learning through knowledge spillovers, and how it varies with city size. While my approach is analogous in using survey data to directly analyze how learning depends on urbanity, this paper newly introduces survey data on off-the-job learning with rich information about frequency, purpose, subject and learning method. Such detailed stylized facts on off-the-job learning are important in themselves, and might also be suggestive for better understanding on-the-job learning since there has been no data about on-the-job learning that are equally rich and detailed. In particular, since off-the-job learning analyzed in this research requires time investment (see Section 2), my results might be especially informative for on-the-job learning that requires active learning costs such as time. For instance, if a higher return for skills in

[3]For evidence that different jobs require different skillsets, see Autor, Levy and Murnane (2003) and Ikenaga and Kambayashi (2016).

[4]Since I investigate learning as an act to improve one's skills, another type of learning that is not analyzed in this research is learning about uncertainty in one's own skills or job matching quality (e.g., Papageorgiou 2014; 2022).

denser cities drives people to more frequently engage in costly off-the-job learning, the same incentive would also promote costly on-the-job learning.[5]

This paper is organized as follows. Section 2 describes my data and empirical strategy. Section 3 presents my main descriptive results. Section 4 assesses the causality of my descriptive evidence. Section 5 presents additional results. Section 6 concludes.

2 Data and empirical strategy

Data. I use the Japanese time-use survey (JTUS, *shakai seikatsu kihon chosa*) to obtain data about learning behavior. The JTUS is designed as a time-use survey, analogous to the American Time-Use Survey (ATUS), administered by the Statistical Bureau of Japan, the Ministry of Internal Affairs and Communications.[6] The JTUS provides repeated cross-sectional data that have been available every five years since 1976, and each round has a sample size of approximately 200,000 individuals. I focus mainly on the 2016 data, but I also use 1986 and 2006 JTUS data for robustness checks. The JTUS is designed to be nationally representative and maintains a very high response rate, 95% for the JTUS 2016, because a response is required by law.[7]

While the JTUS has a similar time diary question as the ATUS, I focus on the question about off-the-job learning in the past year, as shown in Figure 1, for which the ATUS does not have a counterpart.[8] The question instructs "[p]lease indicate the item you aimed to enhance your knowledge or level of culture, or to use for your current work..., *excluding* those activities conducted as a job or schoolwork" and "[e]xclude those activities directly related to regular courses in school, or employee training courses."[9] These instructions imply that on-the-job learning and learning that occurs passively (e.g., just being surrounded by smart people unconsciously enhances one's skills) are excluded. Therefore, in this paper, I focus on off-the-job learning that is undertaken actively with the intention to improve one's knowledge or cultural sophistication.

The question pertains to eight skills: English language, other foreign language, computers etc., commerce or business, elderly care, home economics or housework, humanities/social or natural sciences, and art and culture. For each skill, three subsequent questions are posed. First, the questionnaire asks, in nine categories, how many days in a given year the respondent spent learning a particular skill. The total number of days spent learning is obtained by summing

[5]Sandvik, Saouma, Seegert and Stanton (2020) suggests that actively incurring learning costs, rather than learning passively, is important for gaining more from knowledge spillovers at the workplace.

[6]The use of time-use survey has been rare in urban economics literature. Recent notable exceptions include Murphy (2018) and Su (2022).

[7]The response rate is taken from https://www.soumu.go.jp/main_content/000617655.pdf (in Japanese, last accessed on August 27, 2022).

[8]I also use the JTUS time diary question for supplementary analyses (see Section 5 and footnote 10). See Kuroda (2010) and Lee, Kawaguchi and Hamermesh (2012) for more details on the time diary question of the JTUS.

[9]Italics added by the author. Some wording is the author's and differs from the official English translation of the original Japanese questionnaire in Figure 1 to better capture the original sentences.

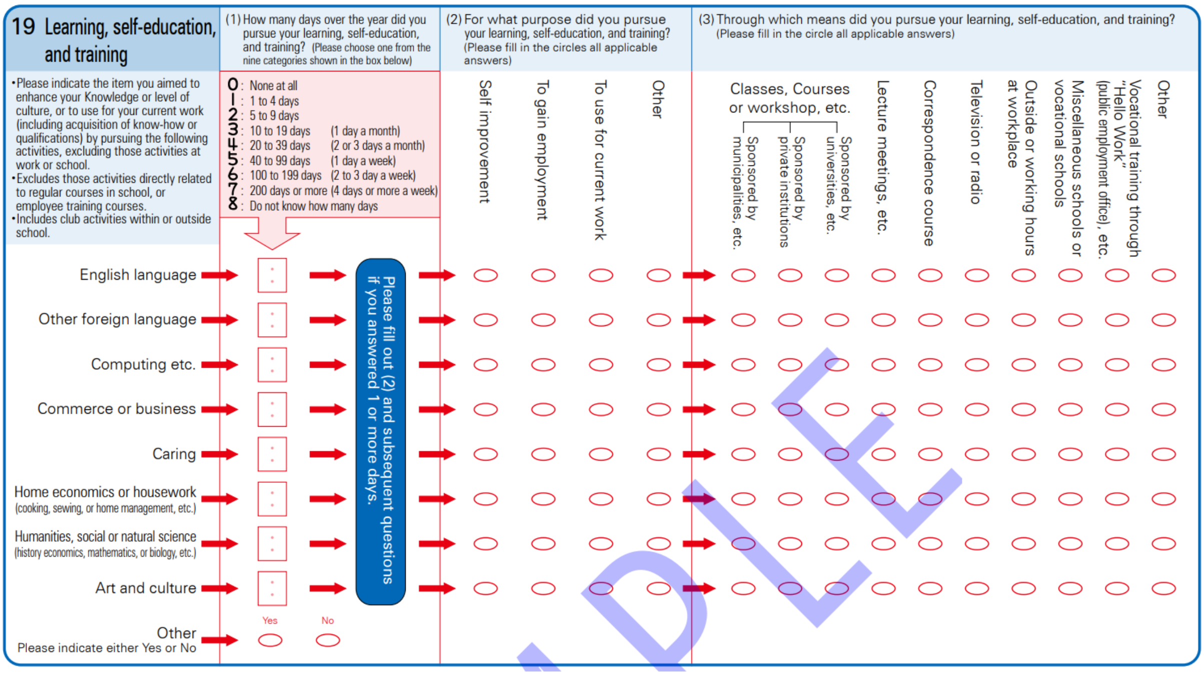


Figure 1: Questions on the off-the-job learning (JTUS, 2016)

Note: The figure is an extract from the English translation of the questionnaire of the JTUS 2016, available at https://www.stat.go.jp/english/data/shakai/index.html (last accessed on August 17, 2022). The original questionnaire in Japanese can be found at https://www.stat.go.jp/data/shakai/2016/index.html (In Japanese. Last accessed on August 17, 2022).

up the median value of each category across all skills.[10] Second, the questionnaire asks the purpose of learning the skill: for cultivating oneself, for gaining employment, for use on one's current job, or for other purposes (multiple choices are allowed).[11] Finally, the questionnaire asks how the respondent learned the skill(s) and lists the following options: classes provided by the public sector, classes provided by the private sector, classes provided by a university, lecture meetings, correspondence courses, television or radio, workplaces outside of working hours, vocational schools, training provided by the public employment office ("Hello Work"), and the other means (multiple choices are allowed). In addition to the eight aforementioned skills, the survey also asks whether the respondent has learned anything that is not included among the eight skills. Overall, the questions allow me to investigate how often people learn off-the-job, what they learn, for what purpose they engage in learning, and how they learn.

The survey also collects a standard set of socioeconomic characteristics for each respondent. Motivated by the urban wage premium literature (e.g., Glaeser and Maré 2001; Charlot and Duranton 2004; De la Roca and Puga 2017), I control for sex, age and its square, marital status, and educational attainment throughout this research. I focus on people between the ages of 25

[10] This might cause a double-counting issue if an individual learns skill A and skill B on the same day. To address this, I also apply an alternative measure of learning frequency taken from the time-diary question of the JTUS, which is minutes spent on off-the-job learning in a given day. I find that people learn more in denser cities, both on the intensive and extensive margins.

[11] The official English translation shown in Figure 1 states "self improvement," but I believe that "for one's own cultivation" better captures the original Japanese sentence (*"jibun no kyouyou wo takameru tame"*).

and 59 years to focus on the working-age population.

I obtain municipal population data from the 2015 Population Census, which is the closest census year to 2016.[12] I also take the habitable area of each municipality from the Municipalities Area Statistics of Japan. I define the population density of each municipality by dividing the total population by the habitable area. As of 2016, there were 1,741 municipalities in Japan, and 1,350 municipalities are covered in my final JTUS 2016 dataset.[13]

To graphically illustrate the relationship between learning outcomes and population density in the raw data, Figure 2 plots key learning outcomes analyzed in Section 3 against municipal population density. It shows that they may have apparent correlations not driven by a few outliers. The scatterplots for other learning outcomes are shown in Figure A.1. Tables A.1 and A.2 present summary statistics.

Empirical strategy. I analyze how various outcome variables on off-the-job learning are related to population density. The outcome of individual i living in municipality j is denoted by y_{ij}. I estimate the following linear model using ordinary-least-squares (OLS):

$$y_{ij} = \beta \ln PopDens_j + \gamma X_i + \epsilon_{ij}. \tag{1}$$

where $\ln PopDens_j$ is the log population density of municipality j, X_i represents the characteristics of individual i that I control for, and ϵ_{ij} is the error term.

In equation (1), we cannot use as y_{ij} the log of the number of days individual i in municipality j engages in learning. In this case, I also estimate the following constant-elasticity regression equation:

$$y_{ij} = \exp(\beta \ln PopDens_j + \gamma X_i) + \epsilon_{ij}, \tag{2}$$

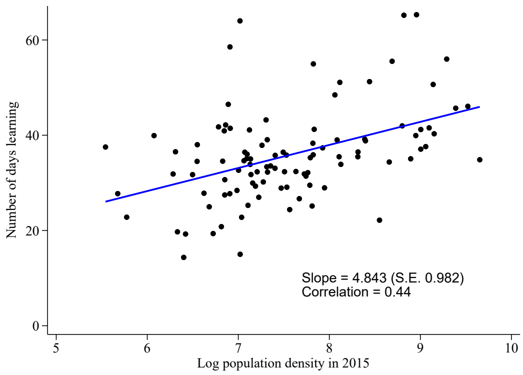
where y_{ij} is the number of days learning. β, the coefficient of interest, is interpreted as the elasticity of the outcome y_{ij} with respect to the population density. Importantly, y_{ij} can be zero, unlike in the standard linear model (1) with a logarithmic outcome variable. Allowing y_{ij} to take zero may be important in this context because only around a third of my sample engaged in off-the-job learning. I follow Santos Silva and Tenreyro (2006) and estimate (2) using the Poisson pseudo-maximum-likelihood (PPML) estimator.[14]

I interpret β in equations (1) and (2) as descriptive. This is because, unlike the literature on the urban wage premium (e.g., Glaeser and Maré 2001; De la Roca and Puga 2017), I cannot control

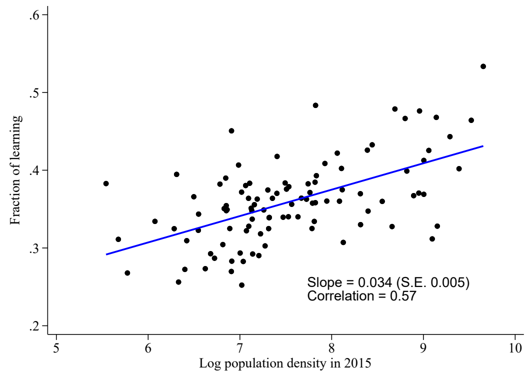
[12] The data for population density are downloaded from https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/mieruka/db_top/index.html (In Japanese. Last accessed on June 30, 2022).

[13] When analyzing the JTUS data prior to 2016 for robustness checks, I use the municipal boundaries as of 2015, while I use the municipal code converter, provided by Kondo (2022), to deal with municipal mergers.

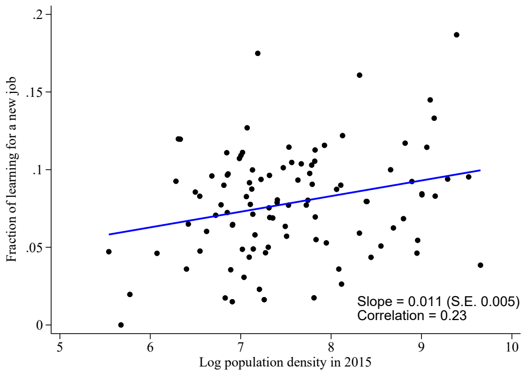
[14] As an alternative method, I apply the inverse hyporbolic sine transformation $\ln(y_{ij} + \sqrt{y_{ij}^2 + 1})$, where y_{ij} is the number of days that the individual engaged in learning and this can be zero (Bellemare and Wichman 2020). By taking this as the outcome variable and estimating the model (1) by the OLS, the estimated elasticity with respect to population density is 0.109, which differs little from 0.102 that I find by the PPML estimation in Table 1.



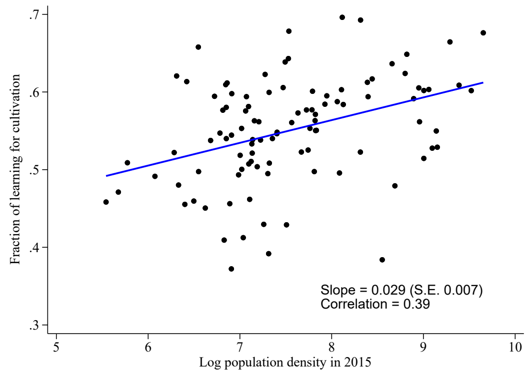
(a) Number of days learning



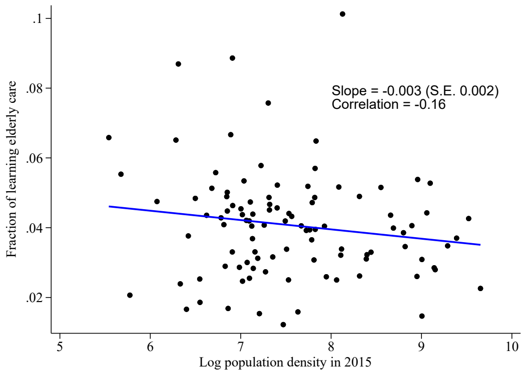
(b) Fraction of learning something



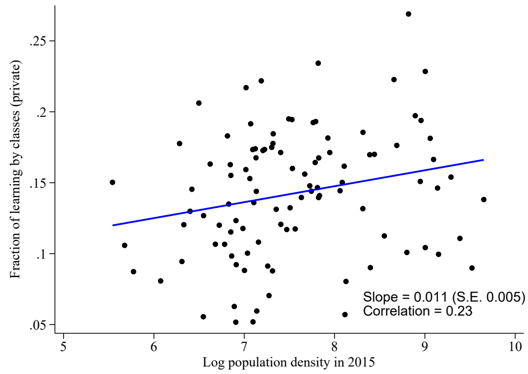
(c) Fraction of learning for a new job



(d) Fraction of learning for cultivation



(e) Fraction of learning elderly care



(f) Fraction of learning by classes (private)

Figure 2: Selected learning outcomes and population density

Note: The figures plot the municipal average of selected learning outcomes against its log population density in 2015 (black dots). The blue line is the regression line. The fractions for a purpose or a method of learning is calculated among those who learned something. To credibly calculate municipal averages and to comply with the data privacy concern by the data provider, I focus on municipalities with more than 150 individual observations (101 municipalities) in these figures. Using alternative threshold numbers of minimum observations produces similar figures. Similar plots for the remaining learning outcomes are found in Figure A.1.

for individual fixed effects to address sorting based on unobserved individual characteristics (e.g., intrinsic motivation) because the JTUS provides cross-sectional data. Moreover, there is no natural experiment that randomly affects population density in this context. Given these limitations, I focus on providing descriptive evidence as my main results (Section 3). The issue of causality is relegated to Section 4,

Throughout this research, I ensure national representativeness by weighting observations according to the sampling probability provided in the dataset. I cluster standard errors at the municipal level.

3 Main descriptive results

3.1 Learning frequency

I first analyze how much more frequently people engage in off-the-job learning in denser cities. The estimation result in Column 1 of Table 1 shows that in specification (2), a 1% increase in population density is associated with a 0.1% increase in the number of days spent learning.[15] This positive effect is driven by both the intensive and extensive margins. Column 2 focuses on the intensive margin by using the log number of days in the linear regression (1), and Column 3 focuses on the extensive margin by using the dummy of whether an individual engages in any learning. I find positive and statistically significant estimates in both columns. Studies have documented the faster wage growth in cities (e.g., Glaeser and Maré 2001; Baum-Snow and Pavan 2012; De la Roca and Puga 2017; Eckert et al. 2022). More frequent learning in denser cities is consistent with this finding.

Longer hours dedicated to learning may provide a new microfoundation for urban agglomeration economies in productivity. Duranton and Puga (2004) points out that learning, matching, and sharing constitute the three important classes of mechanisms contributing to higher urban productivity. Studies have emphasized learning efficiency in cities for a given amount of learning time, especially through knowledge spillovers (e.g., Marshall 1890; Duranton and Puga 2004; Charlot and Duranton 2004; Davis and Dingel 2019). On the other hand, my result indicates a novel simple microfoundation of higher productivity in denser cities based on learning: People in denser cities invest more time in improving their productivity.[16] Taken together, it is likely that both the quality and the quantity of learning are greater in denser cities, both of which contribute to their higher productivity.

Note that welfare implications of agglomeration economies depend on whether learning quality or quantity is improved in cities. Generally, the welfare gain depends on whether it is a pure positive externality or it comes at the sacrifice of other resources. Here, the higher productivity induced by more frequent learning is at the cost of people's time endowment.

[15] Since the standard deviation of log population density is 1.209 (Table A.2), a one standard deviation increase in population density is associated with around 12% increase in learning frequency.

[16] This is consistent with the theoretical prediction of Davis and Dingel (2019).

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.102^a	0.053^a	0.021^a	
	(0.017)	(0.016)	(0.003)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.047	0.014	
N	78727	23513	78727	
R^2	NA	0.014	0.069	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.008^b	-0.009^b	0.026^a	0.004
	(0.004)	(0.004)	(0.004)	(0.005)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.009	-0.013	0.023	0.006
N	24905	24905	24475	24905
R^2	0.016	0.070	0.032	0.025

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 1: Learning frequency and purposes

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	Number of days learning		
	(1)	(2)	(3)
ln population density	0.123^a	0.149^a	0.096^a
	(0.021)	(0.027)	(0.004)
ln population density × College	-0.048		
	(0.030)		
ln population density × Female		-0.090^b	
		(0.035)	
ln population density × Young			0.034
			(0.047)
N	78727	78727	78727

Cluster-robust standard errors in parentheses
^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 2: Learning frequency (heterogeneity)

Note: Adding interaction terms to the constant-elasticity specification (2), I analyze the heterogeneous impact of population density by individual characteristics. The college dummy takes one for those with college education or more. The college dummy takes one for those with college education or more. The female dummy takes one for women. The young dummy takes one for those under the age of 30 or under. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

Therefore, misattributing such productivity gains through more learning to other sources of agglomeration economies, such as more efficient learning through knowledge spillovers, could lead to substantially different welfare implications (c.f., Miyauchi 2022).

I also investigate the heterogeneous impact of density across different socioeconomic groups, which is reported in Table 2. I find that men have a stronger incentive to learn in denser cities where skills are highly valued, which seems plausible in Japan because they have higher labor participation rate than women. Although statistically insignificant ($p \simeq 0.11$), I also find that population density and learning frequency are more strongly associated among people with low educational attainment. This might, again, be consistently explained by labor market incentives. For example, the marginal return of improving skills may be higher for low-skilled workers because it could greatly expand the available jobs whose skill requirements they can meet (Autor et al. 2003; Ikenaga and Kambayashi 2016), and this effect is stronger in denser cities as more jobs are available (Papageorgiou 2022).[17] Consistent with this, albeit statistically insignificant, I find in Table A.3 that men and people with lower educational attainment are more likely to learn to gain new employment as population density increases.

3.2 Purposes of learning

I now analyze the purpose of off-the-job learning. As an outcome, I use a dummy indicating whether a person engages in learning for a given purpose. I consider three purposes, namely, (i)

[17]Alternatively, jobs in denser cities may require more advanced skills even within the same occupation, and workers learn more to meet them.

to obtain a new job, (ii) to utilize the skill at their current job, and (iii) for personal cultivation, although, for completeness, I also report the results on learning for other reasons. To focus on the association between learning purposes and population density while isolating the effect of learning frequency, the sample in this analysis contains only those who engaged in some learning. This leads to a smaller sample size than in Section 3.1.

Columns 4–7 of Table 1 report the results. Column 4 shows that people in denser cities are more likely to learn in order to get a new job. This is consistent with the idea that people are more likely to find new job opportunities (Di Addario 2011; Papageorgiou 2022) in denser cities and thus have a greater incentive to satisfy the skills requirements of the new job opportunities (c.f., Autor et al. 2003; Ikenaga and Kambayashi 2016). Indeed, Eckert et al. (2022) recently finds that gradual matching with more skill-intensive occupations and industries can explain a large part of the faster wage growth in a denser city. My finding suggests that such a transition might be partially made possible by workers' voluntary learning to improve their skills. This highlights the potential complementarity between matching and learning in explaining urban agglomeration economies, two of the three key urban agglomeration forces identified by Duranton and Puga (2004).[18] Column 5 shows that, in stark contrast to Column 4, people are *less* likely to learn skills in order to utilize them at their current job. This further reinforces the view that the rich availability of jobs in the market incentivizes people to acquire the necessary skills.

In addition to labor market concerns, in Column 6, I find that people in denser cities also learn to cultivate themselves.[19] Moreover, in Table A.4, I do not find significant heterogeneity by education, gender, and age. More learning for cultivation in denser cities is consistent with the agglomeration of cultural activities or creative industries (e.g., Mitchell 2019; Tao et al. 2019; Borowiecki and Dahl 2021). This might also be related to studies on endogenous amenities (e.g., Shapiro 2006; Diamond 2016). My result highlights that urban agglomeration may enhance cultural activities, which may lead to more cultural amenities or residents' cultural sophistication.[20] Incorporating this channel into the theories of endogenous amenities might be important.

3.3 What and how people learn

I first investigate what skills people learn. For each skill, I create a dummy variable indicating whether a person spent some time in learning it. I also analyze how people learn by constructing a dummy variable for each learning method indicating whether a person used it.

Table 3 presents the results on what people learn. Almost all skills are more likely to be learned in denser cities, which is intuitive given that people learn more frequently and there are more jobs available in denser cities. However, there is an exception that is more likely to

[18] See Table A.3 for an analysis of the heterogeneous impact of population density on learning for new employment.

[19] A possible reason is that the availability of cultural facilities, such as museums, incentivizes people to acquire the necessary knowledge (e.g., art history) to be able to enjoy cultural activities (Krupka 2009).

[20] Consistent with this, Boualam (2014) suggests that although statistically insignificant, culture might improve consumption amenities.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.020[a]	0.007[a]	0.010[a]	0.010[a]	-0.004[a]
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.015	0.006	0.007	0.008	-0.004
N	78727	78727	78727	78727	78727
R^2	0.080	0.016	0.027	0.039	0.013
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.007[a]	0.004[b]	0.013[a]	0.000	
	(0.002)	(0.002)	(0.002)	(0.002)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.007	0.001	0.010	-0.001	
N	78727	78727	78727	78727	
R^2	0.036	0.050	0.037	0.001	

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 3: Learning by subject

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Classes (public) (1)	Classes (private) (2)	Classes (universities etc) (3)	Lecture meetings (4)	Home-study courses (5)
ln population density	-0.009[a]	0.006[b]	-0.001	-0.009[a]	-0.003
	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	-0.009	0.005	-0.002	-0.010	-0.004
N	24304	24905	24905	24905	24905
R^2	0.019	0.021	0.015	0.019	0.007
	TV and radio (6)	Workplace (outside working hours) (7)	Vocational schools etc (8)	Training by employment service center (9)	Other (10)
ln population density	0.008[b]	-0.007[c]	0.005[a]	-0.001[b]	0.010[b]
	(0.004)	(0.004)	(0.002)	(0.001)	(0.004)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.008	-0.008	0.005	-0.001	0.009
N	24905	24905	24905	24905	24905
R^2	0.026	0.034	0.009	0.004	0.007

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 4: Learning by method

Note: I use the linear probability model that regresses the dummy for learning in a particular method on the municipal population density. The dummy of learning by classes (public) has slightly fewer observations than the other dummy variables on methods because some observations are coded as “unknown" in the original data. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

be learned in less dense cities: elderly care skills. I interpret this as evidence that what people learn is responsive to the skill demand in the local labor market. The demand for elderly care is increasing in rural Japan given the aging population, which makes it attractive relative to jobs in other sectors (Hanaoka 2015).[21] Presumably the strong local demand for the elderly care overturned the general tendency that any skill is more likely to be learned in denser cities.[22] This result highlights that what people learn is responsive to local labor demand conditions.

Table 4 presents the results on the methods of learning.[23] Although not conclusive, there seems to be a tendency that in denser cities, people tend to more often utilize education-market-based methods requiring physical presence (i.e., private classes and vocational schools etc). In contrast, people in less dense cities tend to utilize non-education-market-based methods less often, such as publicly provided methods or training at the workplace outside of working hours.[24] This accords with the view that cities provide more learning opportunities as private providers of learning opportunities can enjoy agglomeration economies, which seem strong in the service sector (Glaeser et al. 2001; Morikawa 2011; Leonardi and Moretti forthcoming).[25]

Overall, consistent with a simple theoretical framework in which a worker decides what to learn under some learning costs, the local skill demand and the supply of learning opportunities matter. This may imply that the higher frequency of learning in Section 3.1 is also caused by both the high skill demand and the rich supply of learning opportunities in denser cities.

4 Are my descriptive estimates also interpretable as causal?

Although the previous section focused on my main goal of descriptively providing stylized facts about the association between learning and population density, I also assess the causal interpretation of my estimates through several additional analyses. Overall, while not conclusive, the following analyses suggest that the estimated β in Section 3 may approximate a causal relationship.

Endogeneity of population. Though not a problem for a descriptive analysis, identification of causal impact of population density on learning is impossible if some unobserved factors simultaneously affect population and learning. To address this, I follow Ciccone and Hall (1996) and use as an IV a long lag of the population density using the 1920 population census, the first

[21] Indeed, the population share of care workers is higher in rural prefectures (https://www.sssc.or.jp/touroku/tourokusya.html, in Japanese. Last accessed on June 2, 2022). A possible reason is that due to a price regulation of elderly care services, the wage rate of the elderly care sector relative to other industries is higher in rural areas than in urban areas (Hanaoka 2015).

[22] In Table A.5, I find that the association of density and learning elderly care skills is stronger for women. This is sensible because the elderly care sector is female-dominant industry is more important for women.

[23] For the same reasons as the analysis of learning purposes, I focus on those who engaged in some learning.

[24] More reliance on publicly-provided opportunities implies that the public sector may supply them to complement the lack of private learning opportunities.

[25] In addition to supply concerns, the utilized methods also depend on the subject of learning. For example, elderly care is often learned through lecture meetings while foreign languages are learned through TV and radio, which might explain statistically significant estimates in Table 4.

modern population census in Japan. I find that, although somewhat noisier, the IV results are similar to the baseline OLS results. This is in line with studies on the urban wage premium that the endogeneity of city size has a limited impact on results (e.g., Ciccone and Hall 1996; De la Roca and Puga 2017). Details are in Appendix B.

Sorting and non-movers. Since my data are cross-sectional, the correlation between learning outcomes and population density might be driven by sorting based on the unobserved heterogeneity of residents, such as a high return from learning and an intrinsic motivation. The urban wage premium literature has addressed this issue by including individual fixed effects (e.g., Glaeser and Maré 2001; De la Roca and Puga 2017).[26] Unfortunately, I cannot use the same strategy because the JTUS has no panel structure.[27]

Instead, I follow Charlot and Duranton (2004) and focus on non-movers to assess how endogenous sorting might affect my results. Intuitively, if the sorting by unobserved factors drives the observed correlation between learning outcomes and density, then focusing on non-movers would eliminate the correlation. Since the 2016 JTUS data do not contain questions about the previous residence, I investigate the 1986 JTUS data containing this information. The 1986 JTUS asks frequency, subject, and method of learning in a different format than the 2016 JTUS. Unfortunately, purpose of learning was not asked in the 1986 JTUS.

I investigate the heterogeneous association of population density with these three types of learning outcomes by analyzing the subsamples of nonmovers defined in two ways: (i) the person stayed in the same residence for more than four years or (ii) the prefecture in which a person went through compulsory education was the same as the current prefecture. I find little evidence that the relationship between learning and population density is different among the sample of non-movers, suggesting that sorting is not a main driver of my results. See Appendix C for details.

Oster's bound. In Tables 1, 3, and 4, I report Oster's (2019) beta. Oster's beta is a conservative estimate of β for causality, taking into account the presence of unobserved selection bias into denser cities. Following Oster (2019), I assume that observed and unobserved factors are equally related to population density. I also assume that the R-squared when controlling for both observed and unobserved variables equals 1.3 times R^2 in each regression. I find that no statistically-significant result changes qualitatively, suggesting that my qualitative results might be causally interpreted.

[26] The fixed effects might control for individual fixed earnings capacity (e.g., Glaeser and Maré 2001; De la Roca and Puga 2017) as well as cognitive learning capacity (Bacolod, Blum, Rangel and Strange 2023) and self-confidence (De la Roca et al. 2023).

[27] Note, however, that including fixed effects could exacerbate the bias because identification relies only on migrants, whose migration choices are made endogeneously (De la Roca and Puga 2017).

5 Additional analyses

Urban earnings premium and learning. I examine the relationship between earnings and population density. I first replicate the earnings premium in my JTUS data. Although it has already been shown that larger cities have higher income in Japan (e.g., Tabuchi and Yoshida 2000), my analysis would offer some new insights by using more recent individual-level microdata. I regress the log annual earnings on the log population density and the same set of control variables as Section 3. I find that a 1% increase in density is associated with a 0.039% increase in income. The magnitude of this estimate is close to 0.04% in the meta-analysis of the urban wage premium by Ahlfeldt and Pietrostefani (2019). See Appendix D for more details.

Higher earnings in denser cities are also consistent with the view that more frequent learning in cities documented in this paper actually improves productivity. To gain some suggestive evidence on this view, I follow Charlot and Duranton (2004) and compute how much of the effect of population density on individual earnings can be explained by the learning channel. Specifically, I regress on log individual earnings the log population density, the inverse hyperbolic transformation of the number of days learning, and the control variables used in Table D.1.[28] Let γ_1 and γ_2 be the coefficient of the log population density and the number of days learning, respectively. I next regress on the (inverse hyporbolic sine transformation of the) number of days learning on log population density and the same control variables, and I denote by γ_3 the coefficient of the log population density. Then, $\gamma_2\gamma_3/(\gamma_1 + \gamma_2\gamma_3)$ gives the fraction of the effects of population density that is mediated through the number of days learning. I estimate that $\gamma_1 = 0.038$ (S.E. 0.006), $\gamma_2 = 0.008$ (S.E. 0.003), and $\gamma_3 = 0.107$ (S.E. 0.015). I get $\gamma_2\gamma_3/(\gamma_1 + \gamma_2\gamma_3) \simeq 0.022$, suggesting that 2.2% of the population density effect is explained by the number of days learning. Note, however, that this may be an underestimation of the importance of learning on earnings as my cross-sectional data do not allow me to analyze the possibility that learning increases future earnings, not current ones.[29]

Learning with whom. To investigate whether learning in denser cities is more likely to be with other people, which might be important in knowledge spillovers, I investigate the JTUS time diary question. I regress on population density the four distinct dummies indicating whether a person engages in learning (i) alone, (ii) with family members, (iii) with colleagues or classmates, and (iv) others. In Table A.6, I find that people are more likely to engage in some learning alone or with family members, but the likelihood of learning with colleagues or other people exhibits little association with population density. Thus, in off-the-job learning, greater knowledge

[28]Instead of the inverse hyporbolic sine transformation, the transformation of adding one and applying the log transformation hardly changes the conclusion.

[29]In particular, this paper focuses on active off-the-job learning that resembles an investment that may bring benefits in the future (e.g., finding a new job or getting promoted) at the cost of current time and effort costs. Moreover, such learning may be particularly active when a negative earnings shock hits, like Ashenfelter's (1978) dip in job training programs. These mask the positive effect of learning on income by creating the negative correlation between current learning frequency and income. Unfortunately, my cross-section data do not allow me to further investigate these possibilities.

spillovers through meeting new people in denser cities do not seem to exist. Note, however, that this does not reject the importance of interactions and knowledge spillovers in cities because they can matter in other dimensions of learning (e.g., Charlot and Duranton 2004; Sandvik et al. 2020).

Metropolitan-area level population density. In Section 3, I used population density at the municipal level as my main variable of interest. As my population data is based on residence, it captures the population density at weekday nights or on holidays. However, with commuting, the relevant density of economic activities might be the density of the metropolitan area rather than the municipality. To address this, I assign to each person the population density of the corresponding urban employment area (UEA, Kanemoto and Tokuoka 2002). The UEA is constructed by aggregating several municipalities to comprise a single commuting zone. I use the UEA definition based on the 2015 Population Census and 100 UEAs are covered in my dataset.[30] While I omit the results to avoid repetitions, I find that using this alternative population density hardly affects my results.[31]

Replicating results in an alternative survey year. To confirm that these results are applicable beyond the 2016 JTUS, I investigate the 2006 JTUS data that asked the same question as the 2016 survey. While I omit the results to avoid repetitions, I have confirmed that essentially all results I presented in Section 3 are replicated.

6 Conclusion

This paper presents new stylized facts about learning and population density. I directly look at off-the-job learning behavior using Japanese survey data that contain very detailed information on the frequency, purpose, content, and method of learning. Although learning has been identified as a key source of urban agglomeration economies that yield higher wages, most empirical studies on the urban wage premium have looked at wage dynamics but not directly analyzed learning behavior (e.g., Glaeser and Maré 2001; Baum-Snow and Pavan 2012; De la Roca and Puga 2017). In contrast, this paper directly looks at the learning behavior by exploiting a distinctive survey on off-the-job learning. This allows me to analyze frequency, purpose, subject, and method of learning, which are hard to infer from wage data.

I document the following stylized facts. First, people in denser cities engage in off-the-job learning more frequently: a 1% increase in population density is associated with 0.1% increase

[30] The UEA definition is available at https://www.csis.u-tokyo.ac.jp/UEA/index_e.htm (last accessed on August 18, 2022).

[31] Note that as density changes, the population size of the city is also likely to change. Unfortunately, it is empirically challenging to separately identify the effects of density and population size as they tend to move together (Ahlfeldt and Pietrostefani 2019). For example, I cannot use the variation in density within the city (e.g., Murphy 2018) as the residential address of the JTUS data is at the level of municipality. That said, I have confirmed that using the population level (i.e., city size) instead of population density hardly changes my conclusion.

in learning frequency. This suggests that the faster wage growth in cities might be partially attributed to more efforts by the city residents. This might also be a new microfoundation of urban agglomeration economics: productivity in denser cities are higher because people voluntarily improve their skills.

Second, people in denser cities are more likely to learn to get a new job or cultivate themselves. This is consistent with the view that there are more job opportunities in denser cities (Di Addario 2011; Papageorgiou 2022) and people have stronger incentives to fulfill the skill requirements by the jobs (c.f., Autor et al. 2003; Ikenaga and Kambayashi 2016). Learning to cultivate oneself implies more cultural activities in denser cities and is consistent with agglomeration of cultural activities or creative industries (e.g., Mitchell 2019; Tao et al. 2019; Borowiecki and Dahl 2021). This might also be related to studies on endogenous amenities (e.g., Shapiro 2006; Diamond 2016) because cultural activities might improve neighborhood quality through cultural amenities or cultural sophistication of residents.

Third, how and what people learn depends on local economic conditions. In particular, while almost all skills are more likely to be learned in denser cities, elderly care skills are more likely to be learned in less dense areas. This is consistent with the high demand for elderly care in the Japanese rural local labor market, where population aging is severe. I also find some evidence that people in denser cities are more likely to utilize education-market-based learning methods that require physical attendance. Presumably, the strong agglomeration economies in the service sector makes the private provision of learning opportunities possible.

I believe that these new stylized facts guide further theoretical and empirical investigations on learning and urban agglomeration economies. That said, limitations on my findings should be considered, which future work can hopefully overcome. First, more investigation is needed to go beyond descriptive evidence to causal statements. Although my analyses in Section 4 suggest that my estimates might approximate causality, obtaining richer data with panel structure or exploiting a natural experiment would further bolster the causal statement. Second, I did not have data on on-the-job learning. People may improve their skills at workplace, such as through communication with others (e.g., Charlot and Duranton 2004, 2006; Sandvik et al. 2020). While my results might also be suggestive of on-the-job learning, direct investigation of on-the-job learning is an important task. Finally, it would be interesting to better disentangle the driving forces behind learning behaviors. Although my discussion suggests that labor market considerations in denser cities can explain many of my key results, richer data with more detailed information on the motivation behind each learning behavior would further highlight the underlying mechanisms.

References

Ahlfeldt, Gabriel M and Elisabetta Pietrostefani, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111*, 93–107.

Ashenfelter, Orley, "Estimating the effect of training programs on earnings," *The Review of Economics and Statistics*, 1978, pp. 47–57.

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly Journal of Economics*, 2003, *118* (4), 1279–1333.

Bacolod, Marigee, Bernardo S. Blum, Marcos A. Rangel, and William Strange, "Learners in cities: Agglomeration and the spatial division of cognition," *Regional Science and Urban Economics*, 2023, *98*, 103838.

Baum-Snow, Nathaniel and Ronni Pavan, "Understanding the city size wage gap," *The Review of Economic Studies*, 2012, *79* (1), 88–127.

Bellemare, Marc F and Casey J Wichman, "Elasticities and the inverse hyperbolic sine transformation," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 50–61.

Borowiecki, Karol Jan and Christian Møller Dahl, "What makes an artist? The evolution and clustering of creative activity in the US since 1850," *Regional Science and Urban Economics*, 2021, *86*, 103614.

Boualam, Brahim, "Does culture affect local productivity and urban amenities?," *Regional Science and Urban Economics*, 2014, *46*, 12–17.

Charlot, Sylvie and Gilles Duranton, "Communication externalities in cities," *Journal of Urban Economics*, 2004, *56* (3), 581–613.

_ and _ , "Cities and workplace communication: some quantitative French evidence," *Urban Studies*, 2006, *43* (8), 1365–1394.

Ciccone, Antonio and Robert E Hall, "Productivity and the Density of Economic Activity," *American Economic Review*, 1996, *86* (1), 54–70.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum, "Matching in Cities," *Journal of the European Economic Association*, 2022, *20* (4), 1478–1521.

Davis, Donald R and Jonathan I Dingel, "A spatial knowledge economy," *American Economic Review*, 2019, *109* (1), 153–70.

De la Roca, Jorge and Diego Puga, "Learning by working in big cities," *The Review of Economic Studies*, 2017, *84* (1), 106–142.

_ , Gianmarco IP Ottaviano, and Diego Puga, "City of dreams," *Journal of the European Economic Association*, 2023, *21*, 690–726.

Di Addario, Sabrina, "Job search in thick markets," *Journal of Urban Economics*, 2011, *69* (3), 303–318.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Duranton, Gilles and Diego Puga, "Micro-foundations of urban agglomeration economies," in "Handbook of Regional and Urban Economics," Vol. 4, Elsevier, 2004, pp. 2063–2117.

Eckert, Fabian, Mads Hejlesen, and Conor Walsh, "The return to big-city experience: Evidence from refugees in Denmark," *Journal of Urban Economics*, 2022, *130*, 103454.

Gaulke, Amanda P, "Post-schooling off-the-job training and its benefits," *Labour Economics*, 2021, *70*, 102007.

Glaeser, Edward L, "Learning in cities," *Journal of Urban Economics*, 1999, *46* (2), 254–277.

— **and David C Maré**, "Cities and skills," *Journal of Labor Economics*, 2001, *19* (2), 316–342.

— **, Jed Kolko, and Albert Saiz**, "Consumer city," *Journal of Economic Geography*, 2001, *1* (1), 27–50.

Hanaoka, Chie, "Kaigo roudou-ryoku busoku ha naze shouzite irunoka (Direct Care Worker Shortages Using Japanese Data : A Review of Previous Studies)," *Nihon Roudou Kenkyu Zasshi*, 2015, *57* (5), 16–25. (In Japanese).

Ikenaga, Toshie and Ryo Kambayashi, "Task Polarization in the Japanese Labor Market: Evidence of a Long-Term Trend," *Industrial Relations*, 2016, *55* (2), 267–293.

Kanemoto, Yoshitsugu and Kazuyuki Tokuoka, "Nihon no toshi-ken settei kijun (Proposal for the standards of metropolitan areas of Japan)," *Journal of Applied Regional Science*, 2002, *7*, 1–15. (In Japanese. An English abstract is available).

Kondo, Keisuke, "Shicho-son gappei wo kouryo shita shikuchoson panel deta no sakusei (Municipality-level Panel Data and Municipal Mergers in Japan)," 2022. RIETI Technical Paper Series 19-T-001. (In Japanese. An English abstract is available).

Krupka, Douglas J, "Location-specific human capital, location choice and amenity demand," *Journal of Regional Science*, 2009, *49* (5), 833–854.

Kuroda, Sachiko, "Do Japanese work shorter hours than before? Measuring trends in market work and leisure using 1976–2006 Japanese time-use survey," *Journal of the Japanese and International Economies*, 2010, *24* (4), 481–502.

Lee, Jungmin, Daiji Kawaguchi, and Daniel S Hamermesh, "Aggregate impacts of a gift of time," *American Economic Review*, 2012, *102* (3), 612–16.

Leonardi, Marco and Enrico Moretti, "The agglomeration of urban amenities: Evidence from Milan restaurants," *American Economic Review: Insights*, forthcoming.

Lynch, Lisa M, "Private-sector training and the earnings of young workers," *American Economic Review*, 1992, *82* (1), 299–312.

Marshall, Alfred, *Principles of Economics*, MacMillan, 1890.

Mitchell, Sara, "London calling? Agglomeration economies in literature since 1700," *Journal of Urban Economics*, 2019, *112*, 16–32.

Miyauchi, Yuhei, "Matching and agglomeration: Theory and evidence from Japanese firm-to-firm trade," *mimeo*, 2022.

Morikawa, Masayuki, "Economies of density and productivity in service industries: An analysis of personal service industries based on establishment-level data," *The Review of Economics and Statistics*, 2011, *93* (1), 179–192.

Murphy, Daniel, "Home production, expenditure, and economic geography," *Regional Science and Urban Economics*, 2018, *70*, 112–126.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Papageorgiou, Theodore, "Learning your comparative advantages," *Review of Economic Studies*, 2014, *81* (3), 1263–1295.

— , "Occupational Matching and Cities," *American Economic Journal: Macroeconomics*, 2022, *14* (3), 82–132.

Rosenthal, Stuart S and William C Strange, "Evidence on the nature and sources of agglomeration economies," in "Handbook of Regional and Urban Economics," Vol. 4, Elsevier, 2004, pp. 2119–2171.

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, and Christopher T Stanton, "Workplace knowledge flows," *The Quarterly Journal of Economics*, 2020, *135* (3), 1635–1680.

Santos Silva, J.M.C. and Silvana Tenreyro, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Shapiro, Jesse M, "Smart cities: quality of life, productivity, and the growth effects of human capital," *The Review of Economics and Statistics*, 2006, *88* (2), 324–335.

Su, Yichen, "Measuring the value of urban consumption amenities: A time-use approach," *Journal of Urban Economics*, 2022, *132*, 103495.

Tabuchi, Takatoshi and Atsushi Yoshida, "Separating urban agglomeration economies in consumption and production," *Journal of Urban Economics*, 2000, *48* (1), 70–84.

Tao, Jin, Chun-Yu Ho, Shougui Luo, and Yue Sheng, "Agglomeration economies in creative industries," *Regional Science and Urban Economics*, 2019, *77*, 141–154.

Appendix to "Off-the-Job Learning in Cities"
(Not for Publication)

A Omitted figures and tables

	mean	sd	count
Number of days learning	39.34	113.63	79879
Dummy of learning something	0.38	0.49	79879
Dummy of learning for a new job	0.084	0.278	25170
Dummy of learning for a current job	0.510	0.500	25170
Dummy of learning for cultivation	0.560	0.496	24729
Dummy of learning for other objectives	0.260	0.439	25170
Dummy of learning English	0.110	0.313	79879
Dummy of learning other foreign language	0.035	0.183	79879
Dummy of learning computers etc	0.132	0.338	79879
Dummy of learning commerce and business	0.104	0.306	79879
Dummy of learning elderly care	0.039	0.193	79879
Dummy of learning housekeeping	0.106	0.307	79879
Dummy of learning humanities and sciences	0.074	0.262	79879
Dummy of learning arts and culture	0.100	0.300	79879
Dummy of learning others	0.079	0.270	79879
Dummy of learning by classes (public)	0.046	0.209	24553
Dummy of learning by classes (private)	0.146	0.353	25170
Dummy of learning by classes (universities etc)	0.019	0.137	25170
Dummy of learning by lecture meetings	0.098	0.297	25170
Dummy of learning by home-study courses	0.074	0.262	25170
Dummy of learning by TV and radio	0.193	0.394	25170
Dummy of learning by workplace outside working hours	0.235	0.424	25170
Dummy of learning by vocational schools etc	0.023	0.149	25170
Dummy of learning by training by employment service center	0.011	0.105	25170
Dummy of learning by others	0.588	0.489	25170

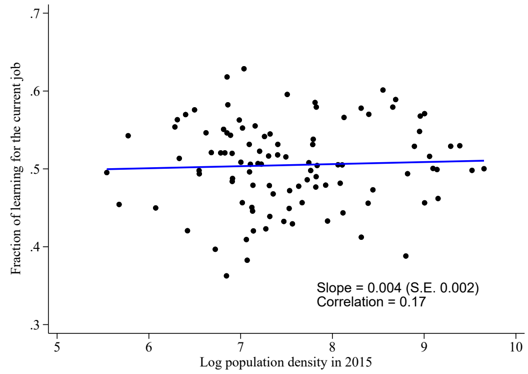
Table A.1: Summary statistics on learning variables

Note: All variables are weighted by the sampling weights. The number of observations for a purpose or a method of learning is smaller because they are defined only for those learning something. The dummy of learning for cultivation has slightly fewer observations than the other dummy variables on purposes because some observations are coded as “unknown" in the original data. Similarly, the dummy of learning by classes (public) has slightly fewer observations than the other dummy variables on methods because some observations are coded as “unknown" in the original data.

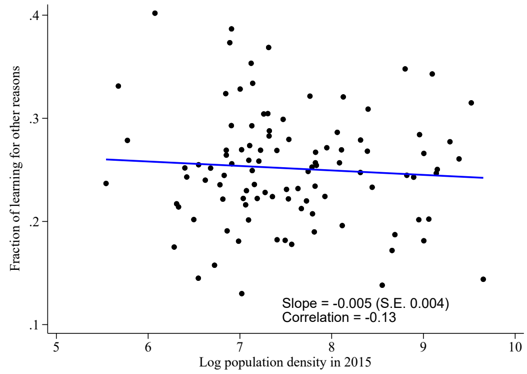
	mean	sd	count
Age	42.57	9.58	79879
Dummy for female	0.496	0.500	79879
Dummy for being married	0.670	0.221	79692
Dummy for college or more	0.330	0.221	78972
Individual annual income (10,000 Japanese yen)	382.22	283.63	67979
ln 2015 municipal population density	6.630	1.209	1350 (the number of municipalities)

Table A.2: Summary statistics on individual characteristics

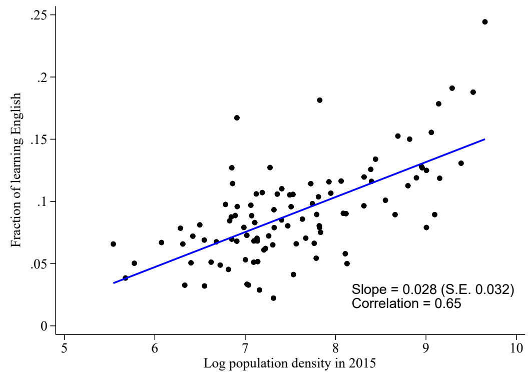
Note: All individual characteristics are weighted by the sampling weights. In the data, marital status take three values (unmarried, married, widowed) and education attainment takes nine values (elementary school, junior high school, high school, vocational school (1–2 years), vocational school (2–4 years), vocational school (more than 4 years), associate degree, bachelors degree, and master or doctoral degree. I summarize these two variables into binary dummies to concisely present the summary statistics. The number of observations varies across characteristics due to missing values. The summary statistics of the log population density is calculated, without weights, at the municipal level because this is the level of variation.



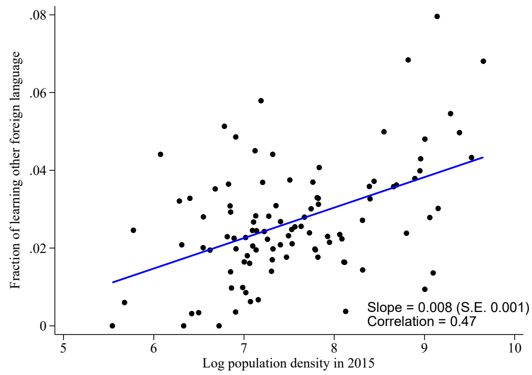
(a) Fraction of learning for the current job



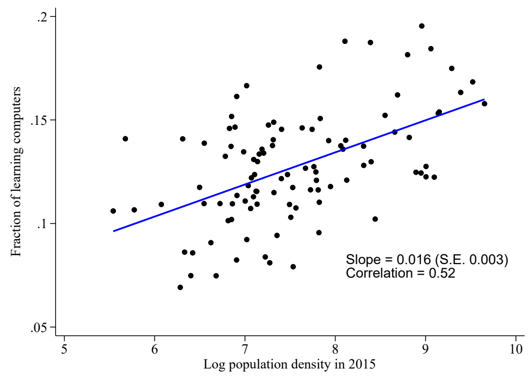
(b) Fraction of learning for other reasons



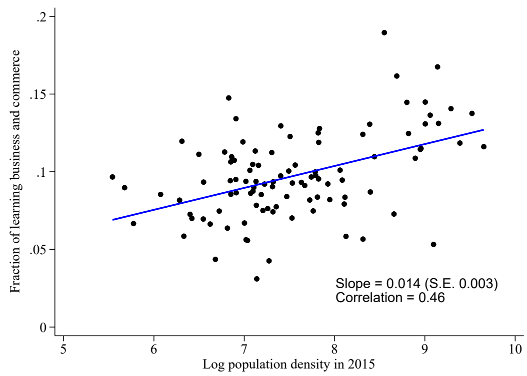
(c) Fraction of learning English



(d) Fraction of learning other foreign language



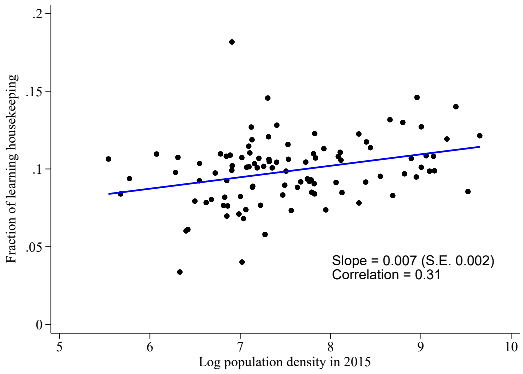
(e) Fraction of learning computers etc



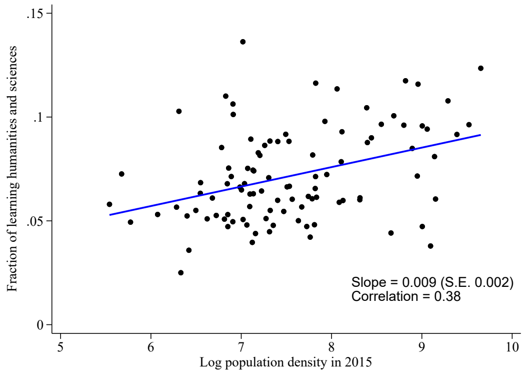
(f) Fraction of learning business and commerce

Figure A.1: Learning outcomes and population density

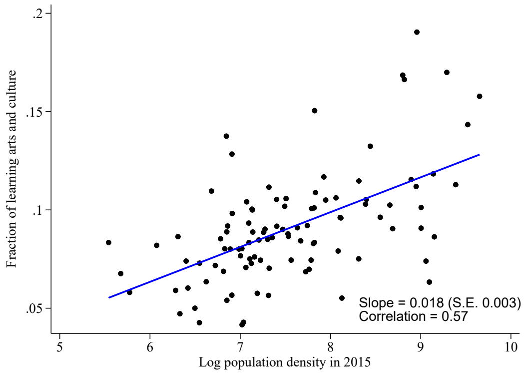
Note: The figures plot the municipal average of learning outcomes against its log population density in 2015 (black dots). The blue line is the regression line. The fractions for a purpose or a method of learning are calculated among those who learned something. To credibly calculate municipal averages and to comply with the data privacy concern by the data provider, I focus on municipalities with more than 150 individual observations (101 municipalities) in these figures. Using alternative threshold numbers of minimum observations produces similar figures.



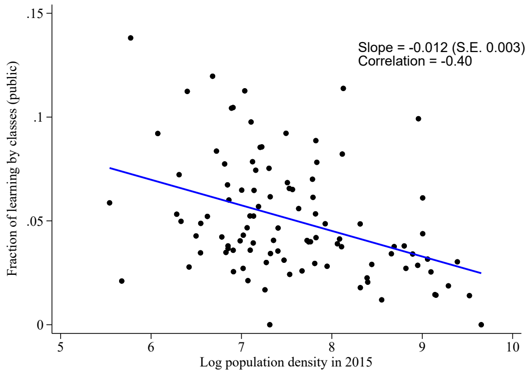
(g) Fraction of learning housekeeping



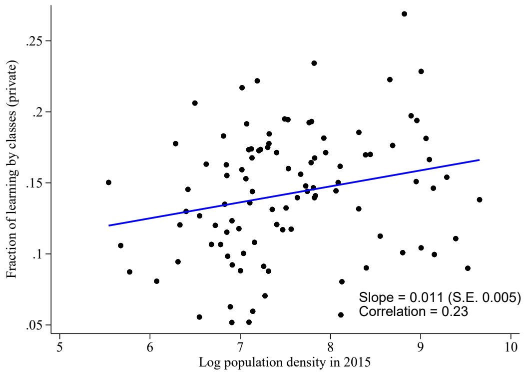
(h) Fraction of learning humanities and social sciences



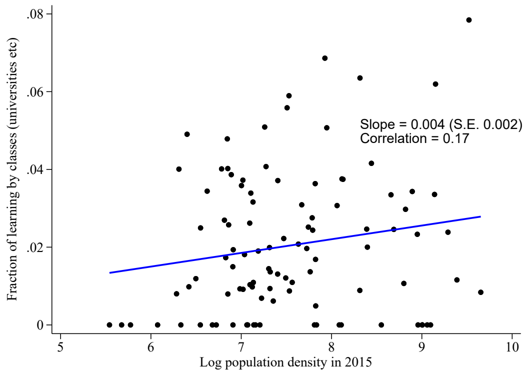
(i) Fraction of learning arts and sciences



(j) Fraction of learning by classes (public)



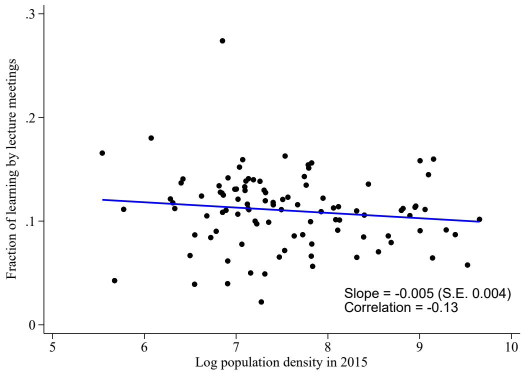
(k) Fraction of learning by classes (private)



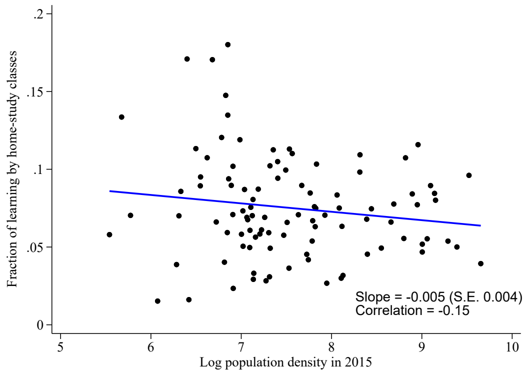
(l) Fraction of learning by classes (universities etc)

Figure A.1: Learning outcomes and population density (cont.)

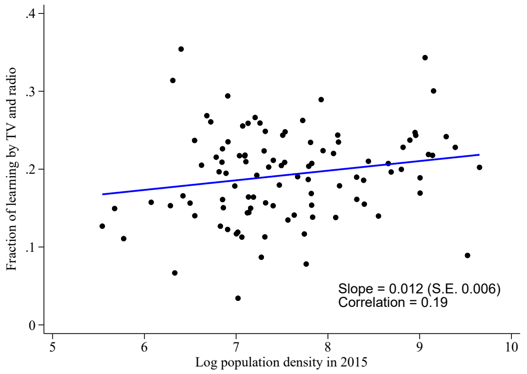
Note: The figures plot the municipal average of learning outcomes against its log population density in 2015 (black dots). The blue line is the regression line. The fractions for a purpose or a method of learning are calculated among those who learned something. To credibly calculate municipal averages and to comply with the data privacy concern by the data provider, I focus on municipalities with more than 150 individual observations (101 municipalities) in these figures. Using alternative threshold numbers of minimum observations produces similar figures.



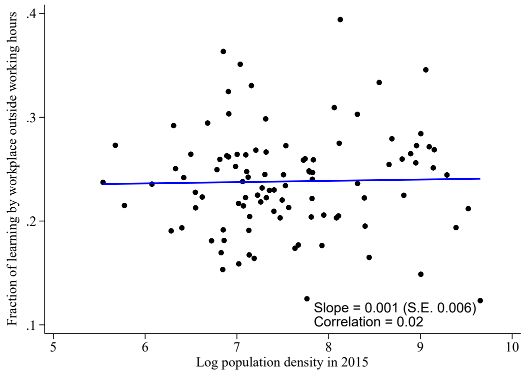
(m) Fraction of learning by lecture meetings



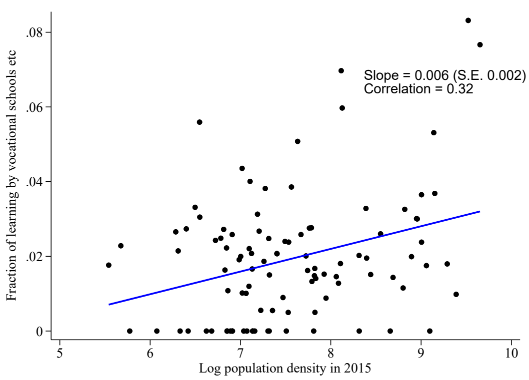
(n) Fraction of learning by home-study classes



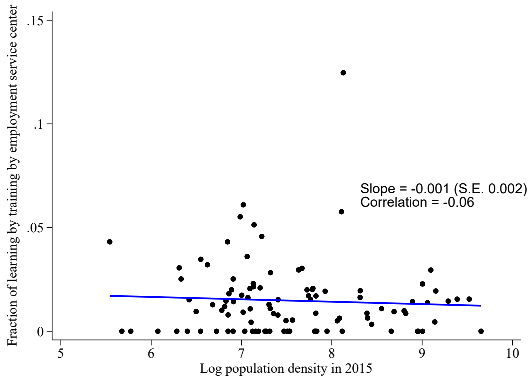
(o) Fraction of learning by TV and radio



(p) Fraction of learning by workplace outside working hours



(q) Fraction of learning by vocational schools etc



(r) Fraction of learning by training at employment service

Figure A.1: Learning outcomes and population density (cont.)

Note: The figures plot the municipal average of learning outcomes against its log population density in 2015 (black dots). The blue line is the regression line. The fractions for a purpose or a method of learning are calculated among those who learned something. To credibly calculate municipal averages and to comply with the data privacy concern by the data provider, I focus on municipalities with more than 150 individual observations (101 municipalities) in these figures. Using alternative threshold numbers of minimum observations produces similar figures.

	Learning for cultivation		
	(1)	(2)	(3)
ln population density	0.033^{a}	0.018^{b}	0.026^{a}
	(0.005)	(0.007)	(0.005)
ln population density	-0.018		
× College	(0.010)		
ln population density		0.015	
× Female		(0.010)	
ln population density			-0.000
× Young			(0.014)
N	24475	24475	24475

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.4: Learning for cultivation (heterogeneity)

Note: Using interaction terms, I analyze the heterogeneous impact of population density by individual characteristics. The college dummy takes one for those with college education or more. The college dummy takes one for those with college education or more. The female dummy takes one for women. The young dummy takes one for those under the age of 30 or under. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

	Learning for a new job		
	(1)	(2)	(3)
ln population density	0.133^{b}	0.135^{b}	0.010^{b}
	(0.006)	(0.006)	(0.004)
ln population density	-0.012		
× College	(0.008)		
ln population density		-0.010^{c}	
× Female		(0.005)	
ln population density			-0.007
× Young			(0.009)
N	24905	24905	24905

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.3: Learning for a new job (heterogeneity)

Note: Using interaction terms, I analyze the heterogeneous impact of population density by individual characteristics. The college dummy takes one for those with college education or more. The college dummy takes one for those with college education or more. The female dummy takes one for women. The young dummy takes one for those under the age of 30 or under. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

	Learning elderly care skills		
	(1)	(2)	(3)
ln population density	-0.004[a]	-0.002[c]	-0.003[a]
	(0.001)	(0.001)	(0.001)
ln population density	-0.002		
× College	(0.002)		
ln population density		-0.004[b]	
× Female		(0.002)	
ln population density			-0.004
× Young			(0.002)
N	78727	78727	78727

Cluster-robust standard errors in parentheses
[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table A.5: Learning elderly care skills (heterogeneity)

Note: Using interaction terms, I analyze the heterogeneous impact of population density by individual characteristics. The college dummy takes one for those with college education or more. The college dummy takes one for those with college education or more. The female dummy takes one for women. The young dummy takes one for those under the age of 30 or under. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

	Alone	Family members	Colleagues or classmates	Others
	(1)	(2)	(3)	(4)
ln population density	0.002[b]	0.002[b]	0.0002	0.0003
	(0.001)	(0.001)	(0.0004)	(0.0003)
N	154105	154105	154105	154105
R^2	0.010	0.006	0.003	0.003

Cluster-robust standard errors in parentheses
[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table A.6: Learning with whom

Note: In Column 1, I regress a dummy that takes one if the person engages in some learning alone. Similarly, Column 2 uses a dummy that takes one if the person engages in some learning with family member, Column 3 uses a dummy that takes one if the person engages in some learning with colleagues or classmates, and Column 4 uses a dummy that takes one if the person engages in some learning with other people. The sample size is larger than the main text because one is required to report his or her behavior for two consecutive days. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I also control for the survey day of week since this analysis uses the time diary question about a particular day. I cluster standard errors at the municipal level.

B Endogeneity of population

While not a problem for a descriptive analysis, identification of causal impact of population density on learning is problematic if some unobserved factors simultaneously affect population and learning. To address these issues, I follow Ciccone and Hall (1996) and use as an IV a long lag of the population density using the 1920 population census, the first modern population census in Japan.

I use the GIS version of the 1930 population census, which also records population from the 1920 census, by Yuji Murayama (http://giswin.geo.tsukuba.ac.jp/teacher/murayama/data.html, in Japanese, last accessed on September 8, 2022). I combine this with the shape file of municipalities as of 2016 and proportionally assign the historical municipal population to current municipalities based on the overlapping area size. Then I divide by the physical area of each municipality to address another concern that the habitable area could be determined endogenously (Hayakawa, Koster, Tabuchi and Thisse 2021). Note that Okinawa prefecture is not covered by this dataset and thus dropped in the IV analysis. This matters little in practice because my OLS results change little by dropping Okinawa prefecture. I use the GMM Poisson regression for estimating equation (2) and 2SLS regression for equation (1).

Tables B.1–B.3 present the IV results, which correspond to Tables 1–4 in the main text. I find that the IV results are fairly similar to the baseline OLS results. The similarity of the IV results implies the limited importance of the endogeneity of population size, as in Ciccone and Hall (1996) and De la Roca and Puga (2017).

	Number of days learning (1)	ln number of days learning (2)	Engaged in learning (3)	
ln population density	0.130^{a}	0.080^{a}	0.025^{a}	
	(0.027)	(0.025)	(0.005)	
Estimation method	GMM	2SLS	2SLS	
N	77236	23084	77236	
R^2	NA	0.014	0.069	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.011	-0.006	0.024^{a}	0.021^{b}
	(0.008)	(0.008)	(0.007)	(0.009)
Estimation method	2SLS	2SLS	2SLS	2SLS
N	24438	24438	24013	24438
R^2	0.016	0.070	0.032	0.025

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table B.1: Learning frequency and purposes (IV results)

Note: In Column 1, I estimate by the GMM constant-elasticity model (2) regressing number of days spent for learning on the municipal population density. In Column 2, I estimate by two-stage-least-square (2SLS) the linear regression model taking log population density as the outcome variable. In the remaining columns, I estimate the linear probability model. In all columns, I use as an IV for the population density in 2015 the population density from 1920 population census. The first-stage F-statistic is 43.85 for Column 1 and 3, 32.97 for Column 2, 34.22 for Column 6 and 33.53 for Columns 4, 5, and 7. Column 6 has slightly fewer observations than Columns 4, 5, and 7 because some observations are coded as “unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.023[a]	0.010[a]	0.013[a]	0.011[a]	-0.004[a]
	(0.004)	(0.002)	(0.003)	(0.004)	(0.001)
Estimation method	2SLS	2SLS	2SLS	2SLS	2SLS
N	77236	77236	77236	77236	77236
R^2	0.080	0.016	0.027	0.039	0.013
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.012[a]	0.004	0.019[a]	0.002	
	(0.004)	(0.003)	(0.004)	(0.003)	
Estimation method	2SLS	2SLS	2SLS	2SLS	
N	77236	77236	77236	77326	
R^2	0.036	0.050	0.037	0.007	

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table B.2: Learning by subject (IV results)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. I use as an IV for the population density in 2015 the population density from 1920 population census. The first-stage F-statistic is 43.85. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

	Classes (public) (1)	Classes (private) (2)	Classes (universities etc) (3)	Lecture meetings (4)	Home-study courses (5)
ln population density	-0.011[a]	0.013[b]	-0.003[c]	-0.007	-0.000
	(0.002)	(0.006)	(0.002)	(0.005)	(0.004)
Estimation method	2SLS	2SLS	2SLS	2SLS	2SLS
N	23847	24438	24438	24438	24438
R^2	0.019	0.021	0.014	0.019	0.007
	TV and radio (6)	Workplace (outside working hours) (7)	Vocational schools etc (8)	Training by employment service center (9)	Other (10)
ln population density	0.012[b]	-0.001	0.004	-0.001	0.005
	(0.006)	(0.006)	(0.003)	(0.001)	(0.009)
Estimation method	2SLS	2SLS	2SLS	2SLS	2SLS
N	24438	24438	24438	24438	24438
R^2	0.021	0.034	0.009	0.004	0.001

Cluster-robust standard errors in parentheses

[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table B.3: Learning by method (IV results)

Note: I use the linear probability model that regresses the dummy for learning in a particular method on the municipal population density. I use as an IV for the population density in 2015 the population density from 1920 population census. The first-stage F-statistic is 34.47 for Column 1 and 33.53 for the other columns. The first column has slightly fewer observations because some observations are coded as “unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

C Sorting and non-movers

Since my data is cross-sectional, the correlation between learning outcomes and population density might be driven by sorting based on the unobserved heterogeneity of residents, such as a high return from learning and an intrinsic motivation. The urban wage premium literature has addressed this issue by including individual fixed effects (e.g., Glaeser and Maré 2001; De la Roca and Puga 2017). Unfortunately, I cannot use the same strategy in this paper because the JTUS has no panel structure.[C.1]

Instead, I follow Charlot and Duranton (2004) and focus on non-movers to assess how endogenous sorting might affect my results. Intuitively, if the sorting by unobserved factors drives the observed correlation between learning outcomes and density, then focusing on non-movers would eliminate the correlation. Since the 2016 JTUS data do not contain questions about the previous residence, I investigate the 1986 JTUS data containing such information. Although in a different format than the 2016 JTUS, the 1986 JTUS asks frequency, subject, and method of learning in a different format than the 2016 JTUS.[C.2] Unfortunately, purpose of learning was not asked in the 1986 survey. I analyze the association between population density and learning outcomes by investigating the full sample and the subsample of (i) people living in the same residence for more than four years and (ii) people staying in the prefecture in which they went through compulsory education, which ends at the age of 15 or 16 in Japan. I use the same set of controls as the 2016 data.

Table C.1 reports the results. I find a similar effect size of population density for the full sample and the subsample of non-movers. In particular, for estimates that are statistically significant in the full sample, focusing on the non-mover samples never changes the qualitative result. This limited importance of sorting for my results is consistent with evidence that the migration decision is not necessarily determined by the expected returns, but by other factors such as self-confidence (De la Roca, Ottaviano and Puga forthcoming). Overall, while not conclusive, this result suggests that my results on learning frequency, subjects, and methods are unlikely to be driven by residential sorting based on unobserved characteristics.

[C.1] Note, however, that including individual fixed effects could exacerbate the bias because the identification relies only on migrants, whose migration choices are made endogeneously (De la Roca and Puga 2017).

[C.2] The 2016 JTUS first asks learning frequency for each *skill* as seen in Figure 1. In contrast, the 1986 JTUS first asks learning frequency for each *method*. Total number of days spent for learning is now defined by summing them up. For each method, it then asks the subjects that were learned by this method. The listed methods and skills also somewhat differ from the 2016 JTUS. See https://d-infra.ier.hit-u.ac.jp/Japanese/statistical-yb/b007.html for the questionnaire of the 1986 JTUS data (in Japanese, last accessed on October 2, 2022).

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	Foreign languages (4)	Business and commerce (5)
Full sample	0.091[a]	0.155[a]	-0.001	0.010[a]	0.007[a]
	(0.014)	(0.010)	(0.003)	(0.001)	(0.001)
Staying in the prefecture	0.089[a]	0.161[a]	-0.003	0.008[a]	0.006[a]
of compulsory education	(0.016)	(0.012)	(0.001)	(0.001)	(0.001)
Not moving	0.117[a]	0.173[a]	0.003	0.008[a]	0.007[a]
in four years	(0.016)	(0.011)	(0.004)	(0.001)	(0.001)
	Engineering (1)	Medicine (2)	Cooking and beauty (3)	Housekeeping (4)	Childbearing (5)
Full sample	0.000	-0.002[a]	-0.001[c]	-0.000	-0.004[a]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Staying in the prefecture	0.001	-0.003[a]	-0.002[a]	-0.000	-0.004[a]
of compulsory education	(0.001)	(0.001)	(0.001)	(0.005)	(0.001)
Not moving	0.001	-0.002[a]	-0.002[a]	0.000	-0.004[a]
in four years	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	Education and social welfare (1)	Social science (2)	Natural and science (3)	Arts and culture (4)	Current affairs (5)
Full sample	-0.011[a]	-0.000	-0.003[a]	0.004[a]	-0.002[a]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Staying in the prefecture	-0.011[a]	-0.000	-0.004[a]	0.005[a]	-0.002[c]
of compulsory education	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Not moving	-0.009[a]	0.001	-0.004[a]	0.005[a]	-0.002[b]
in four years	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	Miscellaneous school (1)	Vocational school (2)	Classes (public) (3)	Classes (private) (4)	Lecture meetings (5)
Full sample	0.002[a]	-0.0004[b]	-0.010[a]	0.006[a]	-0.011[a]
	(0.0003)	(0.0002)	(0.001)	(0.001)	(0.001)
Staying in the prefecture	0.001[a]	-0.0003	-0.009[a]	0.005[a]	-0.011[a]
of compulsory education	(0.0003)	(0.0003)	(0.001)	(0.005)	(0.002)
Not moving	0.001[a]	-0.0004[c]	-0.010[a]	0.007[a]	-0.010[a]
in four years	(0.0003)	(0.0002)	(0.001)	(0.001)	(0.002)
	Correspondence study (1)	TV and radio (2)	Workplace (outside (working hours) (3)	Group study (4)	Study alone (5)
Full sample	0.001[b]	0.007[a]	-0.001	-0.004[a]	0.007[a]
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.002)
Staying in the prefecture	0.000	0.006[a]	-0.003[a]	-0.004[a]	0.005[a]
of compulsory education	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Not moving	0.0013[a]	0.008[a]	-0.000	-0.003[a]	0.007[a]
in four years	(0.0005)	(0.002)	(0.001)	(0.001)	(0.002)

Cluster-robust standard errors in parentheses

[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table C.1: Learning and population density among non-movers (1986 JTUS)

Note: Using the 1986 JTUS data, I estimate the association between population density and learning outcomes using the linear regression, except for Column 1 in which I use the constant-elasticity model (2). Column 1 is estimated by the PPML while the remaining columns are estimated by the OLS. Coefficients of the log population density is reported, separately for the full sample ($N = 149,872$), the sample of those staying in the prefecture of compulsory education ($N = 114,595$), and the sample of those staying in the same residence over the past four years ($N = 113,307$). In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. Standard errors are clustered at the municipal level.

D Replicating the urban earnings premium

I examine the relationship between earnings and population density to replicate the urban earnings premium in my JTUS data. Although it has already been shown that larger cities have higher income in Japan (e.g., Tabuchi and Yoshida 2000), my analysis would offer some new insights by using more recent individual-level microdata.[D.1]

I create a continuous measure of the individual income from the JTUS. For those with positive earnings, the JTUS 2016 asks the individual annual earnings in 15 categories. I assign the median value of each category to create a continuous measure of the annual earnings.[D.2] I then regress the log individual annual earnings on the log population density and the same set of control variables as Section 3 (i.e., sex, age and its square, marital status, and education attainment). Column 1 of Table D.1 shows that 1% increase in density is associated with 0.039% increase in income. The magnitude of this estimate is consistent with other estimates and is close to 0.04% in the meta-analysis of the urban wage premium by Ahlfeldt and Pietrostefani (2019). Higher earnings in denser cities are also consistent with the view that more frequent learning in cities documented in this paper actually improves productivity.

Note that this OLS regression excludes those with no income. To address this issue, I also estimate the constant-elasticity regression model (2). Column 2 of Table D.1 shows a somewhat larger urban earnings premium of 0.058%, but broadly consistent with the prior studies (e.g., De la Roca and Puga 2017; Ahlfeldt and Pietrostefani 2019).

Strictly speaking, earnings and wages may differ due to differences in working hours. In particular, earnings are mechanically higher if workers in denser cities work more, as documented by Rosenthal and Strange (2008) for professional workers in the US. However, using the time-use diary question of the JTUS, I find that denser cities do not have longer working hours or, if any, have somewhat shorter working hours, which is consistent with the result of Rosenthal and Strange (2008) for non-professional workers in the US.[D.3] This implies that in my context, the urban earnings premium would be similar to or somewhat smaller than the urban wage premium.

[D.1] Tabuchi and Yoshida (2000) conducts the city-level analysis and does not use individual-level microdata. In general, evidence on urban earnings premium is scant in the Japanese context (see Higashi 2022 for a recent survey). Another subtle difference of my analysis from Tabuchi and Yoshida (2000) is that I measure urbanity by population density, while Tabuchi and Yoshida (2000) use the size of the city. See also Ahlfeldt and Pietrostefani (2019) on this issue.

[D.2] 15 categories are (1) 0–0.5 million yen, (2) 0.5–1 million yen, (3) 1–1.5 million yen, (4) 1.5–2 million yen, (5) 2–2.5 million yen, (6) 2.5–3 million yen, (7) 3–4 million yen, (8) 4–5 million yen, (9) 5–6 million yen, (10) 6–7 million yen, (11) 7–8 million yen, (12) 8–9 million yen, (13) 9–10 million yen, (14) 10–15 million yen, (15) more than 15 million yen. Category 15 does not have an upper bound, so I assign 15 million yen to everyone in this category.

[D.3] For example, the constant-elasticity regression model (2), regressing working hours on the log population density, yields the coefficient of -0.024 (S.E., 0.004).

	ln individual income	
	(1)	(2)
ln population density	0.039^{a}	0.058^{a}
	(0.006)	(0.006)
Estimation method	OLS	PPML
N	66545	67057

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table D.1: Learning frequency among movers and non-movers

Note: Column 1 presents the linear regression result of log individual income on log population density. Column 2 adopts the constant-elasticity model in equation (2) to include those with zero income. The estimation is by PPML recommended by Santos Silva and Tenreyro (2006). In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

References for the Appendix

Ahlfeldt, Gabriel M and Elisabetta Pietrostefani, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111*, 93–107.

Charlot, Sylvie and Gilles Duranton, "Communication externalities in cities," *Journal of Urban Economics*, 2004, *56* (3), 581–613.

Ciccone, Antonio and Robert E Hall, "Productivity and the Density of Economic Activity," *American Economic Review*, 1996, *86* (1), 54–70.

De la Roca, Jorge and Diego Puga, "Learning by working in big cities," *The Review of Economic Studies*, 2017, *84* (1), 106–142.

_ , Gianmarco IP Ottaviano, and Diego Puga, "City of dreams," *Journal of the European Economic Association*, forthcoming.

Glaeser, Edward L and David C Maré, "Cities and skills," *Journal of Labor Economics*, 2001, *19* (2), 316–342.

Hayakawa, Kazunobu, Hans Koster, Takatoshi Tabuchi, and Jacques-François Thisse, "High-speed rail and the spatial distribution of economic activity: Evidence from Japan's Shinkansen," *mimeo*, 2021.

Higashi, Yudai, "Toshika ga roudousha ni ataeru eikyou: Roudou shijou ni okeru shuseki no keizai to fukeizai (On the impact of urbanization on workers: Agglomeration economies and diseconomies in a labor market)," *Nihon roudou kenkyu zassi*, 2022, *746.* (In Japanese).

Rosenthal, Stuart S and William C Strange, "Agglomeration and hours worked," *The Review of Economics and Statistics*, 2008, *90* (1), 105–118.

Santos Silva, J.M.C. and Silvana Tenreyro, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Tabuchi, Takatoshi and Atsushi Yoshida, "Separating urban agglomeration economies in consumption and production," *Journal of Urban Economics*, 2000, *48* (1), 70–84.