Article

Urban Analytics and City Science

EPB: Urban Analytics and City Science
2020, Vol. 47(7) 1219–1236

DOI: 10.1177/2399808318824112
journals.sagepub.com/home/epb
SAGE

Measuring pedestrian volume by land use mix: Presenting a new entropy-based index by weighting walking generation units

Ha Na Im
The Seoul Institute, Republic of Korea

Chang Gyu Choi
Hanyang University, Republic of Korea

Abstract

This study proposes an alternative to the conventional entropy-based land use mix index, which is generally used to measure the diversity of land use. Pedestrian volume was selected as the dependent variable as it represents the vitality of districts, which many recent urban studies now consider important. The study investigates an entropy-based weighted land use mix index, which is weighted by different land use types. For the index, different areas are needed to generate a unit of pedestrian volume, whose measure is m^2/person/day. The study demonstrates that this alternative is more effective than the existing conventionally used entropy-based land use mix index for explaining pedestrian volume. The research confirms that the conventionally used entropy-based land use mix index can have a positive or negative impact depending on the land use characteristics of the survey points because the conventionally used entropy-based land use mix index has a non-linear relationship with pedestrian volume. By analysing 9727 surveyed locations of pedestrian volume in Seoul, Korea, the study demonstrates that the weighted land use mix index, rather than the conventionally used entropy-based land use mix index, can improve the explanatory power of the estimation model for the relationship between pedestrian volume and built environments, showing consistent results throughout the empirical analysis. In future built-environment studies, the utility of the weighted land use mix index is expected to improve if studies include how to find the accurate weighting of the land use in estimating the pedestrian volume.

Corresponding author:
Chang Gyu Choi, Graduate School of Urban Studies, Hanyang University, Seoul 04763, Republic of Korea.
Email: cgchoi@hanyang.ac.kr

Keywords
Land use diversity, pedestrian volume, entropy-based land use mix index, walking generation unit, entropy-based weighted land use mix index, Seoul

Introduction

Recent trends in sustainable development demonstrate that land use mix (LUM) is considered critical today in urban planning and urban design. Planning arguments, which include smart growth, new urbanism, and transit-oriented development, emphasize the importance of land use diversity as a central principle. The issue of how to measure LUM is practical as well as important in evaluating whether the concepts are applicable. If researchers and/or planners are unable to estimate the effects of land use diversity in their analysis and planning, the mix of land use arguments may be vulnerable to criticism.

The entropy-based LUM index established by Cervero and Kockelman (1997) has been the most representative index used to measure land use diversity in built environment studies, including urban planning and design, transportation planning, and health studies. However, several previous studies have pointed to the limits of this index (e.g. Cerin et al., 2007; Duncan et al., 2010; Lee and Koo, 2013; Yoon et al., 2014; Yun and Choi, 2013). Some analyses using the index have not presented statistically positive effects of LUM on pedestrian volume and, in certain cases, have even shown a negative relationship.

Im and Choi (2019) found that the LUM index had a critical limitation beyond those already discussed by Kockelman (1996), Brown et al. (2009), Hess et al. (2001), and others. They confirmed that the conventional entropy-based land use mix index (hereafter, CLUM) has an n-shaped relationship with pedestrian volume and that, in the case of the sample from the whole city, the explanatory power of pedestrian volume increased in the empirical model when using both CLUM and CLUM squared rather than only using CLUM. However, they only suggested the use of the CLUM and did not suggest an advanced alternative. Based on the general fact that different land use generates varied walking volume, this study develops and proposes an index that corrects and complements the CLUM.

To propose an alternative index, the study examines the origin and scope of the CLUM and confirms the limits of the index. This study focuses on the walking generation unit (WGU) by building use[1] area as a link between pedestrian volume and land use diversity. Because each building use requires a different area to generate one unit of pedestrian volume, this application can increase the likelihood of estimating pedestrian volume in an area with mixed land use. This study presents an entropy-based weighted land use mix index (WLUM) and applies different weighted values by WGU of each land use. The pedestrian volume estimation models using CLUM and WLUM are developed for the high-density and mixed land use city, Seoul, South Korea. This study verifies that WLUM is a more efficient indicator by examining the direction of the coefficient values of the index and the explanatory power of the empirical models. The suggested WLUM will help researchers more readily understand the effects of land use diversity and can be applied to practical plans.

Literature review

Land use diversity and the CLUM

According to recent urban planning principles, appropriate LUM will reduce car trips, improve people's health, encourage street vitality, and increase property value (e.g. Cervero and Duncan, 2003; Diao and Ferreira, 2010, 2014; Forsyth et al., 2008; Frank et al., 2005; Handy et al., 2006; Sung et al., 2015; Zhang, 2004). While traditional Euclidian zoning and modernism-based development have facilitated land use segregation, many cities have been attempting to promote a LUM. Measuring the mix of land use informs the understanding of the effects of such diversity as well as of establishing urban policies and planning. Cervero and Kockelman (1997) developed an entropy-based land use mix index or LUM (here, CLUM) based on the entropy concept (see equation (1)). This is a typical indicator used for measuring LUM (Ewing and Cervero, 2010; Song et al., 2013)

$$\mathrm{CLUM} = -\sum_{i=1}^{k} \frac{P_i \times \ln(P_i)}{\ln(k)} \quad (1)$$

where k is the number of land use categories (hereafter: building use) and P_i refers to the proportion of developed land area (hereafter: building area) in the ith use type. CLUM values range from 0 (no mixture, single land use) to 1 (all land use categories equally present).[2] The number following CLUM represents the k value of equation (1). For example, CLUM2 is classified into two land use categories ($k = 2$), and CLUM4 is classified into four ($k = 4$). The entropy-based index is useful when measuring three or more mixed land use categories, as not only is it difficult to present a mixture of three or more numbers but also this mixture often lacks proven indicators. Compared to any other index that can represent three or more mixed categories, CLUM is the most efficient index when an integrated measurement method is needed and/or when the number of uses is more than two (Frank et al., 2005; Song et al., 2013). Song et al. (2013) and van Eck and Koomen (2008) demonstrated that the result of the Herfindahl index, which is often used as an alternative to CLUM, had almost the same value as CLUM when CLUM was used as a key indicator of land use diversity.

It is commonly expected that there will be a positive (+) relationship between CLUM and the dependent variable of built environments. Im and Choi (2019: 7–8), however, demonstrated that the positive (+) relationship between CLUM and people's behaviour mainly occurred in residential areas, and the association is limited when CLUM values are around 1. They also empirically demonstrated that the relationship between CLUM and pedestrian volume was negative (−) in commercial areas.

Pedestrian volume, CLUM, and the limitation of the index

Over 50 years ago, Jacobs (1961) suggested land use diversity as a means to promote the vitality of districts. However, it was not until the last decade that research activity increased in the areas of the diversity of land use and the vitality of streets. Recently, Gehl and Svarre (2013) showed how pedestrian volume can be used as a representative index to measure district liveability. In fact, studies that seek to determine the factors that affect pedestrian volume only began to be published in the last decade (e.g. Hajrsouliha and Yin, 2015; Lee and Koo, 2013; Liu and Griswold, 2009; Yun and Choi, 2013). In these studies, CLUM was used as the index in most of the studies that examined the effect of LUM on pedestrian volume. However, these pedestrian volume studies do not discuss the limitations of the index.

Kockelman (1996: 49) stated that 'the index of land use balance, entropy, remains constant when distinct land use types remain in constant relative proportions; yet mixing or integration of land uses can change dramatically', but this topic has not been discussed in depth. Although there are studies that have attended to the issue of CLUM, these focus solely on geometric points and have suggested only a fragmentary alternative. As such, it has been difficult to overcome the fundamental limitations of their alternatives to CLUM (e.g. Brown et al., 2009; Hess et al., 2001). In addition, health studies have concentrated on the number of land use categories, namely k in equation (1). Other health studies have focused on specific land use categories that should be excluded when calculating CLUM because they do not affect people's behaviour (e.g. Cerin et al., 2007; Christian et al., 2011; Duncan et al., 2010; Forsyth et al., 2008).

Most previous studies analysing the characteristics of CLUM have not discussed the fundamental characteristics of the index. However, Im and Choi (2019) showed the non-linear characteristic of CLUM. An example of this is shown in Figure 1, in which point 'a' comprises a mixture of 80% residential and 20% commercial use and point 'b' comprises 80% commercial and 20% residential use, with both equal to 0.72 of the CLUM value. Kockelman (1996), Brown et al. (2009), and Hess et al. (2001) all noted this issue with CLUM but focused only on the proportion of 20:80 or on the fact that points 'a' and 'b' had the same CLUM value.

This issue is even more evident when CLUM2 is plotted in linear form in Figure 1. The CLUM2 value is the dotted line and the quadratic function curve: the form of the 'n' shape for CLUM2 makes a fixed relationship with the dependent variable difficult (Im and Choi, 2019). On the left side of the vertex in Figure 1, CLUM2 and the estimated pedestrian volume have a positive (+) relationship, but on the right side, they have a negative (−) relationship.[3]

Korea Transport Surveys and DB Construction Projects (Korea Transport Institute, 2001) show that different pedestrian volumes are generated through various building use categories: residential is 0.11 person/m^2/day, commercial is 0.85/m^2/day, office is

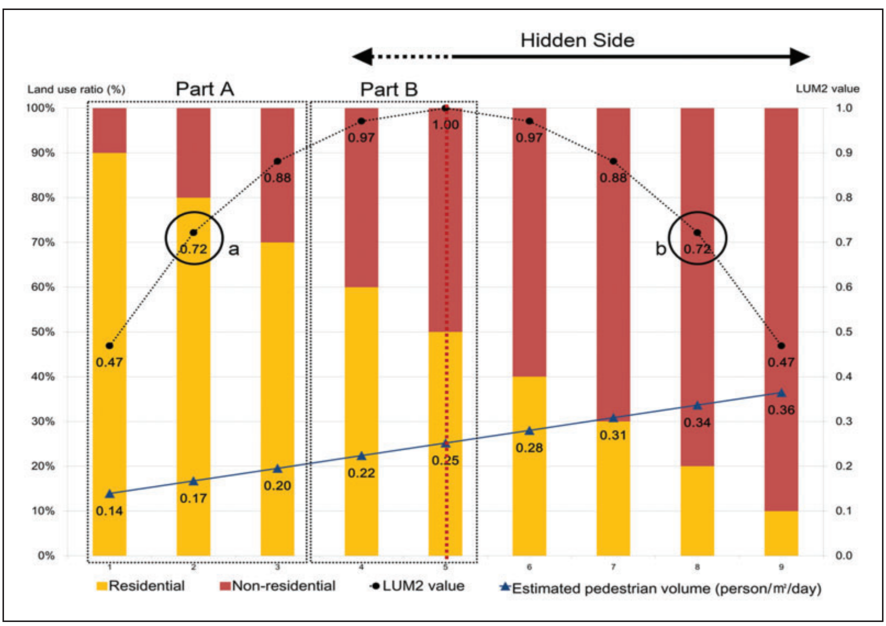

Figure 1. The relationship between CLUM2 index and estimated pedestrian volume. Source: Im and Choi (2019) revised. LUM: land use mix.

$0.35/m^2$/day, and other is $0.33/m^2$/day. Applying these values for CLUM2 in residential and non-residential uses, one can draw a straight line, which is similar to the one that can be drawn for estimated pedestrian volume in Figure 1 (Im and Choi, 2019). This straight line serves as a logical image of the pedestrian volume created when mixing residential and non-residential uses, and it reconfirms a limitation of CLUM.

Im and Choi (2019) show how using CLUM4 and $CLUM4^2$ together in the empirical estimation model for pedestrian volume is logical. CLUM generates a positive pedestrian volume (+) coefficient, and $CLUM4^2$ generates a negative (−) one. However, the researchers overlook the possibility of changes in the meaning of CLUM, reflecting various land use categories with different WGUs. They do not consider the weighting of pedestrian volumes generated in different use scenarios.

Methodology and data

Developing WLUM

The issue of weighting each land use to overcome the limitation of CLUM was discussed by Kockelman (1996: 49), who was the first to suggest the standardization of the equation in the entropy-based model. She recommended that 'one feasible weighting scheme might depend on deviations from the regional balance of proportions across the J use types'. However, there are no extant studies that reflect weighting values. Perhaps this is because previous studies have been based on heuristic beliefs that the CLUM would have a positive (+) relationship with the dependent variable or that biased samples come from largely residential areas (Im and Choi, 2019).

This study proposes a WLUM, which modifies the CLUM, to reflect the weighted value of different land use categories. When different land uses are classified into *n* uses, each classification needs a different area to make a unit of pedestrian volume. Alternatively, each use generates a different pedestrian volume per unit area, and it can be deduced that the WGUs can be created.

There are two ways to calculate the WGU: in one, the denominator is the area of use, and in the other, the denominator is the number of pedestrians. Korea Transport Surveys and DB Construction Projects (Korea Transport Institute, 2001) provide a survey of the weights for the former, and the measured value is person/m^2/day. If we use the weights of the data as is, the weighted proportion of residential:commercial: business:other use is 1:7.68:3.21:3.05. Im and Choi (2019) called the former WGU, but when WLUM is calculated using the former proportions, the values logically show a negative (−) relationship to pedestrian volume and can create unnecessary confusion, because the general public would understand that LUM in a city has a positive relationship to pedestrian volume. However, if the number of pedestrians is the denominator, a positive (+) relationship to pedestrian volume is shown. Based on the concept of weighting, the latter is also right. The pedestrian denominator's unit is m^2/person/day. The weights for residential:commercial:business:other use are then 1:0.13:0.31:0.33. Therefore, this study used the weighted value in which the denominator is the number of pedestrians.

For ease of understanding, different unit areas can be estimated for each building use to represent one pedestrian WGU. If unit values are applied, equation (1) is transformed into equation (2). We can deduce that if each land use is classified into *n*, different values for pedestrians per unit area are accommodated, which will appear in the form of WGU.

Equation (2) refers to WLUM as a dependent variable, and the weighting is based on the amount of pedestrian WGUs for each building use

$$\mathrm{WLUM} = -\sum_{j=1}^{k} \frac{\left\{ \frac{P_j \times W_j}{\sum_{i=1}^{k} P_i \times W_i} \times \ln\left(\frac{P_j \times W_j}{\sum_{i=1}^{k} P_i \times W_i} \right) \right\}}{\ln(k)} \tag{2}$$

where i and j are the number of building uses, and P_i refers to the proportion of developed building area in the ith or jth use type. W_i and W_j refer to the weighted value for the i land use type of pedestrian volume generation. WLUM values range from 0 (no mixture, single land use) to 1 (land use having weighted values, equally mixed).

Equation (2) proposes a biased 'n' shape rather than a general symmetrical 'n' shape for the CLUM by including the weights of the pedestrian WGU amounts for each use in the entropy formula. The relation between WLUM and the amount of walking will be mostly positive (+), although CLUM has a non-linear relationship with the estimated pedestrian volume. If the WGU proposition for residential and non-residential is set to 1:0.248, reflecting the data from Korea Transport Surveys and DB Construction Projects (2001), the WLUM2 value graph can be added to Figure 1 with CLUM2 and estimated pedestrian volume, as shown in Figure 2.

The slope of the WLUM depends on the value of the weighted value (W_i). If the W_i of each land use approaches 1, the shape of WLUM becomes similar to that of CLUM. In other words, if the difference in the weighted values among the land uses increases, the negative (−) relationship with the pedestrian volume, shown on the WLUM graph's right

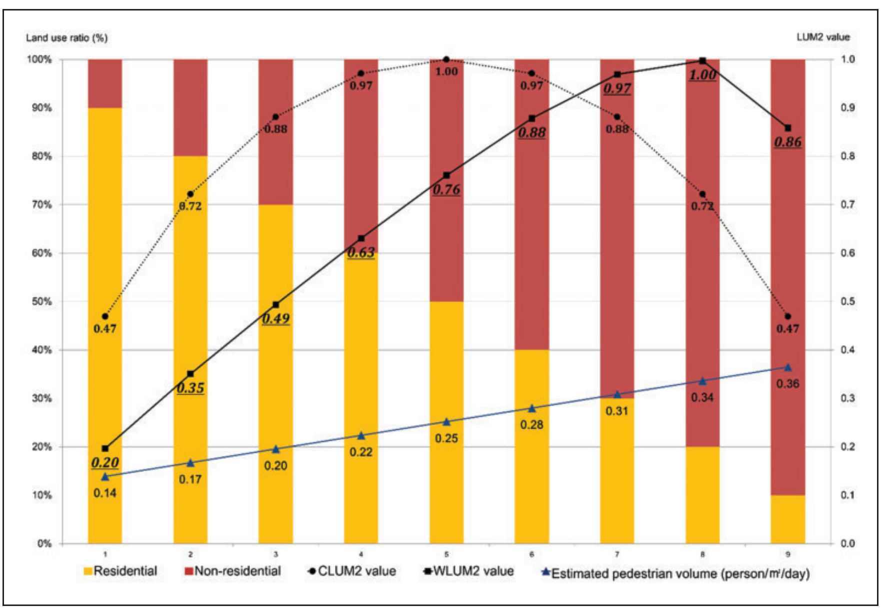

Figure 2. LUM, CLUM2, WLUM2, and estimated pedestrian volume. CLUM: conventional entropy-based land use mix index; LUM: land use mix; WLUM: weighted land use mix index. Note: WLUM4 weight (residential:non-residential = 1:0.248).

side, decreases. Based on the fact that each land use has different WGUs, WLUM can provide a better explanation for pedestrian volume than can the CLUM (see Figure 2).

Seoul, data set, and initial findings

Seoul is a city with a use mix characteristic of high density, which many urban studies have focused on recently (Jang et al., 2017; Sung and Choi, 2017). Its size is 605.2 km^2 and its population is 10,204,057 (Seoul Metropolitan Government, 2017). The public subway and bus networks are well connected and have high utilization rates (Pan, 2013). As the metropolitan government researched the city's pedestrian volume at a wide range of surveyed locations in order to identify how to promote walking and street activity, it offered an appropriate research opportunity to analyse the relationship between pedestrian volume and LUM.

The data for empirical analysis are based on the pedestrian survey conducted by the Seoul Metropolitan Government in 2009 (hereafter, the 2009 pedestrian volume survey). The survey included 10,000 locations in diverse parts of the city. Survey spots were randomly selected from areas where there was a possibility for a certain amount of walking. The pedestrian volume of each survey spot was studied during weekdays and weekends. The physical environment at the street level (street-level design elements such as road width, survey spot slope, the presence of braille blocks, and pedestrian-only streets) within 50 m of every spot was listed.

The dependent variable for the analysis was set as the weekday average pedestrian volume at the survey spots. Information on independent variables affecting the number of pedestrians was obtained from previous studies (e.g. Hajrsouliha and Yin, 2015; Lee and Koo, 2013; Liu and Griswold, 2009; Yun and Choi, 2013). In some previous studies, socioeconomic factors such as the numbers of workers and residents were considered as independent variables for pedestrian volume estimation. This study mainly employed independent variables for the built environment to analyse the urban planning and design objects that are subject to manipulation. Independent variables were divided into LUM (diversity), street-level design elements, accessibility, and regional and land use characteristics.

Street-level design elements in this study were collected from the 2009 pedestrian volume survey. Sidewalk width and numbers of road lanes would be expected to have positive effects on pedestrian volume, as they indicate high levels of traffic or use. Pedestrian-only streets, street furniture, and braille block would be expected to have a positive impact on volume since they suggest levels of street maintenance and safety. Bus stops and subway entrances within 50 m and bus-only lanes would also be expected to have positive effects, because they facilitate the movement of high volumes of pedestrians. The slope of the survey spot would have a negative effect in that it would impede walking.

LUM and characteristics were calculated by using the Seoul Metropolitan Government property tax list in 2009, which has approximately 120 categories listed as 'type of buildings according to taxation-purpose use'. Land use in a 100 m radius[4] of the survey spots was calculated by aggregating the floor area for each building use. Four types were used to compute CLUM4 and WLUM4 for a combination of residential, commercial, business, and other uses. The last 9728 survey points, excluding spots where land use characteristics were missing, were selected for the analysis (see Figure 3).

Accessibility to bus stops and the subway was calculated using the Seoul Metropolitan Government's road network data. Accessibility would negatively affect pedestrian volume because mass transit absorbs pedestrians. As central commercial business districts (CBDs)

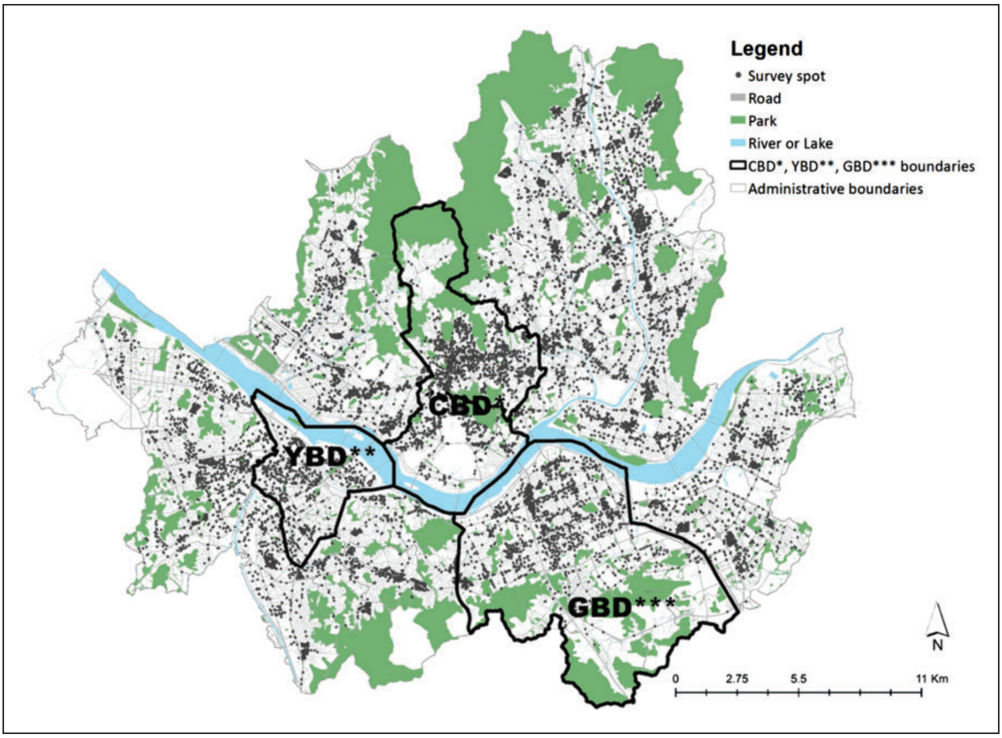

Figure 3. Survey locations of pedestrian volume and district classification in Seoul. Note: *CBD is Central business district in Seoul. **YBD is Yeong-deung-po/Yeo-ui-do business district in Seoul. ***GBD is Gangnam business district in Seoul.

are expected to have more pedestrians than others and because Seoul is assumed to have three CBDs, these were coded. The north of the Han River ('big river' in Korean) are relatively older districts compared to the south of the river, and the former have more pedestrians than the latter. Thus, region type was also included in this study's data set.

This research constructed CLUM4 and WLUM4 within a radius of 100 m from the survey spots and analysed how these two indicators correlated with the measured pedestrian volume. Figure 4 is a graph of pedestrian volume values based on the CLUM4 and WLUM4 values. The LUM values are represented by a line, and the pedestrian volume is represented by a bar graph. Graph (a) in Figure 4 does not clearly show the relationship between walking and CLUM4, although graph (b) in Figure 4 suggests that WLUM4 and pedestrian volume have a weak linear positive (+) relationship. Pearson's correlation coefficient between CLUM4 and pedestrian volume is 0.126, for $CLUM4^2$ it is 0.108, for WLUM4 it is 0.286, and for $WLUM4^2$ it is 0.269. All of these are statistically significant. On the other hand, Figure 4 shows that the two LUM values having low pedestrian volumes in an average weekday are widely distributed in both graphs.

The distribution pattern of points with volumes over 10,000 pedestrians seems to be random regardless of the CLUM4 values in graph (a) of Figure 4. However, the pattern shows a relatively linear relationship with WLUM4 in graph (b). In graph (a) of Figure 4, the circle represents the spot with the highest walking volume, having around 100,000 pedestrians per weekday, representing about 0.53 of the CLUM4 value, whereas this circle in (b) is 0.62 of the WLUM4 value. The spot is located on the main street of a

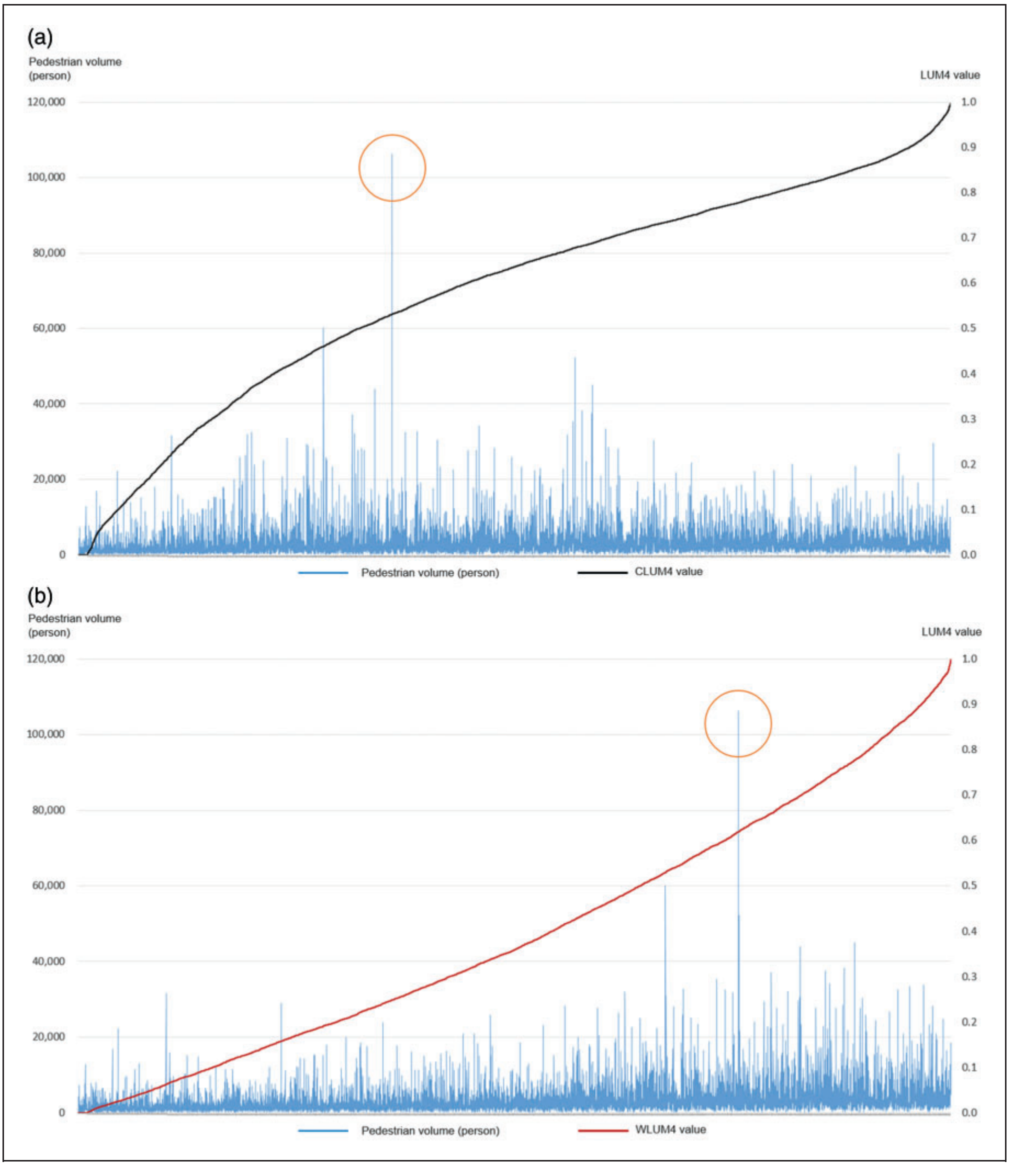

Figure 4. Pedestrian volume, CLUM4 and WLUM4 (measured from 9728 spot). (a) CLUM4 and pedestrian volume and (b) WLUM4 and pedestrian volume. LUM: land use mix. Note: WLUM4 weight (residential: commercial:office:other = 1:0.13:0.31:0.33).

retail and restaurant area called the Gangnam Station district. The commercial rate of land use within a radius of 100 m from the survey spot is 66.6% with office at 30.5%, and there is no residential land use there (see Table 1).

Table 1 shows a comparison of the values of CLUM4 and WLUM4 at the top 10 and bottom 10 spots in terms of pedestrian volume. In the top 10 locations, the proportion of residential use is almost zero. At eight points, WLUM4 is higher than CLUM4, and the commercial use ratio is very high. Two spots are observed for which the latter is higher than the former. In these cases, office use is more than 50%, indicating a strong business district.

Table 1. Top 10 and bottom 10 pedestrian volume spots, CLUM4, and WLUM4.

		Pedestrian volumeon weekday	CLUM4	WLUM4	Building use proportion within a 100m radius from survey spots(%)			
					Residential (%)	Commercial (%)	Office (%)	Other use (%)
Top 10 spots	1	106,186	0.531	**0.618**	0.0	66.6	30.5	2.9
	2	60,197	0.460	**0.530**	0.0	69.6	30.0	0.5
	3	52,298	**0.678**	0.619	0.0	35.8	53.6	10.6
	4	45,139	0.688	**0.777**	0.0	58.1	26.7	15.2
	5	43,882	0.513	**0.698**	0.0	75.5	16.7	7.9
	6	38,271	0.682	**0.762**	0.0	57.5	29.3	13.2
	7	37,466	0.688	**0.734**	0.3	49.8	40.5	9.4
	8	37,131	0.493	**0.660**	0.0	75.3	19.7	5.0
	9	35,321	**0.677**	0.591	0.0	20.0	61.1	19.0
	10	34,134	0.609	**0.742**	0.1	65.1	26.4	8.4
Bottom 10 spots	1	8	**0.057**	0.012	98.6	1.3	0.0	0.1
	2	22	0.000	0.000	100.0	0.0	0.0	0.0
	3	23	0.000	0.000	100.0	0.0	0.0	0.0
	4	23	**0.260**	0.082	91.9	4.3	1.3	2.5
	5	26	**0.235**	0.088	91.9	2.0	0.0	6.1
	6	29	**0.242**	0.137	0.0	10.5	89.5	0.0
	7	41	**0.690**	0.540	37.5	6.8	1.9	53.9
	8	43	**0.156**	0.032	94.6	5.3	0.1	0.0
	9	44	**0.004**	0.001	99.9	0.1	0.0	0.0
	10	44	0.483	**0.492**	33.3	0.5	0.0	66.2

CLUM: conventional entropy-based land use mix index; WLUM: weighted land use mix index.
Note: Bold LUM is higher than another LUM.

These top 10 points must be included in the hidden part of Figure 1, as pointed out by Im and Choi (2019).

In the bottom 10 spots in terms of pedestrian volume, the proportion of residential area exceeds 90% in seven spots, while in the remaining three, the proportion of office or other use is high (see Table 1). Compared to the top 10 spots, we can see that the values of CLUM4 are higher than those of WLUM4 overall. In addition, CLUM4 has a relatively high value of 0.2 or higher at four spots.

From Figure 4 and Table 1, it can be inferred that, compared to CLUM4, WLUM4 is a better indicator of the correlation between land use diversity and pedestrian volume. However, these estimates cannot be confirmed, because other variables may affect the pedestrian volume. Therefore, in the next section, we will build an analytic model that includes these as control variables and discuss the usefulness of WLUM.

Verification of the relationship between pedestrian volume and WLUM4 as well as CLUM4

Descriptive analysis of pedestrian volume and LUM indexes

The average pedestrian volume at the survey sites on a weekday was 3081, with a standard deviation of 3763. This study used CLUM4 and WLUM4, respectively, as diversity indicators, and it sought to verify their effects by reflecting them in each model. The mean value of CLUM4 was 0.58 and that of WLUM4 was 0.40 (see Table 2).

The average street width at the survey spots was 4.0 m, with a standard deviation of 2.2 m. The average number of traffic lanes was 2.9. Survey spots with bus stops within a 50 m radius were 2249 (23.1%) and those having subway entrances at the same distance numbered 680 (7.0%). The average distance to each bus stop was 94.7 m, and for subway station entrances, the average distance was 419.7 m. CBDs represented 33.5% of the research sample, and 53.0% of the sample was selected from north of the Han River. Among total floor areas in the research sample, average residential use included 20,737 m^2 and average commercial use was 11,982 m^2. Residential building use proportion was 49.7%, commercial was 26.7%, and office was 16.3%.

Regression analysis of pedestrian volume using CLUM4 and WLUM4

Logically, CLUM4 has a nonlinear relationship with pedestrian volume and WLUM4 has a relatively linear relationship, as shown in Figure 2. In the descriptive analysis, it is hard to be certain whether the relationship with pedestrian volume is clear. This is because the empirical data also show that, in some spots, CLUM4 and WLUM4 both appear to affect the pedestrian volume, but in other areas, the relationship is not consistent.

This study uses regression analysis to control other dependent variables and builds four different models to compare CLUM and WLUM (see Table 3). Some models in Table 3 also include a squared value of LUM as well as general LUMs as variables, because Im and Choi (2019) suggested that it is logically valid and empirically appropriate to use both CLUM and CLUM squared in whole city models. All models are statistically significant and show a slight difference in adjusted R-squared. The directionality of the LUM indexes and the change in the explanatory power of the models show that the weighted value of WLUM that this study suggests is better than the conventional same value calculation for CLUM. Model 2 with CLUM4 and $CLUM4^2$ together has a higher adjusted R-squared than Model 1 with CLUM4 only, because CLUM4 in Model 2 may act as a control variable and also because many selected survey spots were in residential areas. $CLUM4^2$ has an n-shaped relationship with pedestrian volume, as Im and Choi (2019) found.

Compared to the models with CLUM4, Models 3 and 4 with WLUM4 show a better adjusted R-squared and consistency of variables. Model 4 with WLUM4 and $WLUM4^2$ shows better adjusted R-squared than Model 3 with only WLUM4.[5] The empirical results in Table 3 confirm that, compared to CLUM, WLUM is more stable and consistent in estimating the pedestrian volume, and this study also confirmed that WLUM is more logical than CLUM in estimating the pedestrian volume. These logical and empirical results suggest that a model with WLUM and WLUM squared should be used to calculate the pedestrian volume of the whole city.

Most variables included as control variables also have statistically significant effects on pedestrian volume, maintaining directional consistency in all models (see Table 3). Street maintenance and safety factors such as pedestrian-only streets, street furniture, and braille block width as well as traffic-related factors such as sidewalk width, number of road lanes, and the presence of bus stop and subway entrances have consistently positive relationships to pedestrian volume in all models. Slope at the survey spots negatively affects pedestrian volume. Distance to bus stop and subway stations shows a negative relationship to pedestrian volume. Compared to other areas, survey spots in CBDs and the north of the Han River have higher volumes of pedestrians. Higher density around the survey spots increases volume. Coefficient directions of most of the control variables are reasonable, as expected.

Table 2. Summary statistics.

Variable	Description		Mean	SD	Min.	Max.
Pedestrian volume						
Pv weekday	Average of pedestrian volume on weekdays (person)		3081	3763	8	106,186
Land use mix (diversity, within 100 m)						
CLUM4	Entropy index of residential, commercial, office, and other use		0.58	0.24	0.00	1.00
CLUM4_SQ			0.40	0.25	0.00	0.99
WLUM4			0.40	0.27	0.00	1.00
WLUM4_SQ			0.23	0.25	0.00	1.00
Physical environment at the street level (within 50 m)						
Street width	Street width (m)		4.0	2.2	1.0	24.3
Pedestrian only street[a]	Pedestrian only street	0 = no	4414	45.4	0	1
		1 = yes	5314	54.6		
Street furniture[a]	Existence of street furniture	0 = no	750	7.7	0	1
		1 = yes	8978	92.3		
Braille block[a]	Existence of braille block	0 = no	7293	75.0	0	1
		1 = yes	2435	25.0		
Slope[a]	Existence of a slope	0 = no	7293	75.0	0	1
		1 = yes	2435	25.0		
Fence[a]	Existence of a fence	0 = no	8243	84.7	0	1
		1 = yes	1485	15.3		
No. lanes	Number of traffic lanes		2.9	2.4	1.0	18.0
Bus lane[a]	Existence of a bus-only lane	0 = no	9061	93.1	0	1
		1 = yes	667	6.9		
Bus stop[a]	Existence of a bus stop within a 50 m radius	0 = no	7479	76.9	0	1
		1 = yes	2249	23.1		
Subway[a]	Existence of a subway entrance within a 50 m radius	0 = no	9048	93.0	0	1
		1 = yes	680	7.0		
Crosswalk[a]	Existence of a nearby crosswalk	0 = no	5245	53.9	0	1
		1 = yes	4483	46.1		

(continued)

Table 2. Continued.

Variable	Description		Mean	SD	Min.	Max.
Accessibility						
Distance to bus stop	Distance to the nearest facility (m)		94.7	72.5	0.9	699.0
Distance to subway entrance			419.7	367.5	0.0	3678.3
Regional characteristics						
CBDs[a]	Inclusion of Seoul CBD,	0 = no	6466	66.5	0	1
	YBD, GBD	1 = yes	3262	33.5		
Region type[a]	0 = North of the Han River		5151	53.0	0	1
	1 = South of the Han River		4577	47.0		
Land use characteristics (within 100 m)						
Residential area	Total floor area within a 100 m radius (m^2)		20,737	18,984	0	325,267
Commercial area			11,982	12,057	0	142,213
Office area			9278	18,211	0	182,384
Other use area			2913	5987	0	201,933
Residential per cent	Building use proportion within a		49.7	31.5	0.0	100.0
Commercial per cent	100 m radius (%)		26.7	19.6	0.0	100.0
Office per cent			16.3	20.3	0.0	100.0
Other use per cent			7.3	13.2	0.0	100.0

CBD: central business district in Seoul; YBD: Yeong-deung-po/Yeo-ui-do business district in Seoul; GBD: Gangnam business district in Seoul; CLUM: conventional entropy-based land use mix index; WLUM: weighted land use mix index.

[a]Frequency and percent.

Sources: Seoul Metropolitan Government (2009), Seoul Metropolitan Government's property tax list in 2009; Seoul Metropolitan Government's road network in 2009.

Table 3. Regression analysis of pedestrian volume with CLUM4 and WLUM4.

	Model 1		Model 2		Model 3		Model 4	
Variable	Coef.	β	Coef.	β	Coef.	β	Coef.	β
Land use mix (diversity, within 100 m)								
CLUM4	1054.18***	0.07	4296.10***	0.27				
CLUM4 squared			−3181.08***	−0.21				
WLUM4[a]					2442.70***	0.17	3701.52***	0.26
WLUM4 squared							−1411.48**	−0.09
Physical environment at the street level (within 50 m)								
Sidewalk width(m)	245.18***	0.14	242.88***	0.14	224.17***	0.13	222.75***	0.13
Pedestrian only street (0 = no, 1= yes)	689.29***	0.09	684.08***	0.09	691.73***	0.09	688.23***	0.09
Street furniture (0 = no, 1= yes)	442.90***	0.03	432.97***	0.03	413.86***	0.03	417.16***	0.03
Braille block (0 = no, 1= yes)	327.31***	0.04	329.23***	0.04	303.62***	0.03	298.48***	0.03
Slope (0 = no, 1= yes)	−545.64***	−0.06	−527.64***	−0.06	−464.01***	−0.05	−454.79***	−0.05
Fence (0 = no, 1= yes)	156.70	0.01	174.09*	0.02	144.45	0.01	155.57	0.01
No. of lane	132.45***	0.08	137.65***	0.09	131.98***	0.08	132.91***	0.08
Bus lane (0 = no, 1 = yes)	1150.35***	0.08	1158.73***	0.08	1025.87***	0.07	1029.43***	0.07
Bus stop (0 = no, 1 = yes)	438.17***	0.05	431.80***	0.05	465.37***	0.05	464.68***	0.05
Subway (0 = no, 1= yes)	2279.24***	0.15	2283.07***	0.15	2163.07***	0.15	2183.46***	0.15
Crosswalk (0 = no, 1 = yes)	−287.14***	−0.04	−278.55***	−0.04	−322.66***	−0.04	−327.74***	−0.04
Accessibility								
Distance to bus stop (m)	−2.88***	−0.06	−2.74***	−0.05	−2.21***	−0.04	−2.13***	−0.04
Distance to subway (m)	−1.04***	−0.10	−1.05***	−0.10	−0.82***	−0.08	−0.82***	−0.08
Regional characteristics								
CBDs (0 = no, 1 = yes)	610.09***	0.08	607.96***	0.08	340.71***	0.04	339.99***	0.04
Region type (0= North, 1= South)	−334.41***	−0.04	−309.81***	−0.04	−311.13***	−0.04	−306.80***	−0.04
Land use characteristics (within 100 m)								
Total area (m^2)	0.03***	0.23	0.03***	0.23	0.03***	0.22	0.03***	0.22
Intercept	−594.86***		−1228.88***		−873.89***		−1059.22***	
Number of obs.	9728		9728		9728		9728	
R-squared	0.239		0.242		0.260		0.261	
Adj. R-squared	0.238		0.240		0.259		0.260	
F	179.61***		172.07***		201.09***		190.52***	

CBDs: central business district, Yeong-deung-po/Yeo-ui-do business district, and Gangnam business district in Seoul; CLUM: conventional entropy-based land use mix index; WLUM: weighted land use mix index.

*p<0.1, **p<0.05, ***p<0.01.

The explanatory power of the models presented in this study (R-squared) is relatively low (0.239–0.261). This is presumed to be because this study focused on WLUM as an alternative to CLUM, took control variables related to the built environment, which is the main consideration of urban design, and analysed samples from the entire urban area studied. For example, Liu and Griswold (2009) added variables related to population and employment, and the explanatory power ranged from 0.743 to 0.754. Lee and Koo (2013) targeted specific commercial and business areas, and therefore, their model showed a higher explanatory power than the model in this study. In a future study, socioeconomic variables such as population and employment may be added, or the model's explanatory power might be further improved by adding variables that reflect the characteristics of the region.

Conclusion

Recently, in urban planning and design fields, which emphasize and try to achieve diversity of land use, the issue of how to measure such diversity has been both an academic and a practical concern. When the number of land use categories is more than three, the problem is not easy to solve. As a solution, Cervero and Kockelman (1997) developed the entropy-based land use index – in this study, CLUM – which is a representative index used in various fields to study built environments. Although previous studies have pointed out the limitations of this index, they have not been able to present an appropriate alternative.

The purpose of this study was to address the problem of the entropy-based concept while reflecting the characteristics of each use in the CLUM index. This study was conducted to calculate the weighted values of the differences in effects of each kind of land use on pedestrian volume, a subject which has garnered increasing attention as of late.

The study examined the CLUM and the WLUM to see how they relate to pedestrian volume. The study used regression analysis to assess the relationship between pedestrian volume and the two indexes in 9831 spots in Seoul. This study controlled the design and accessibility factors that would affect pedestrian volume and confirmed that the models with WLUM, as opposed to CLUM, estimated the pedestrian volume more effectively. Based on the empirical results, the WLUM showed a consistent relationship with pedestrian volume compared to the CLUM, and the explanatory power of the regression models that included the WLUM was higher. Thus, when estimating pedestrian volume in the future, it would be preferable to use the WLUM index rather than the CLUM.

Despite this achievement, the study has several limitations. Specifically, it does not offer a method to measure accurate weighted values. The method of measuring weighted values used here is not efficient in terms of cost and time. In addition, the study considered pedestrian volume as the only dependent variable. Future research is necessary to more thoroughly investigate whether the weighted method proposed here would be applicable to an analysis of behaviour mode choice, walking distance and time, and health-related issues. Exact integration effects of different land use types were not discussed in this study; only the effects of individual uses were presented. Future studies should more precisely measure the impact of individual land uses on a dependent variable and analyse the additional effects of the combination and mix of these land uses. They must be more variables than this study.

Acknowledgements

The part of this article was presented in 2017 American Association of Geographers's annual meeting in Boston, Massachusetts, and in 2017 Asian Planning Schools Association's congress in Beijing, China. The authors thank Hyunggun Sung at Chungbuk University, Sugie Lee at Hanyang University, and Sang Hyun Cheon at HongIk University for their valuable advice on developing

WLUM, as well as two anonymous reviewers for their thoughtful comments and suggestions to improve this article's clarity.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science and ICT(NRF-2016R1A2B4010260).

Notes

1. This study uses building use area data because such data have more information on land use intensity than land use classification data.
2. In calculating ln(Pi), when the value of Pi is 0, the log value cannot be calculated. In this case, a practical method is used to calculate the LUM. Applying a very small value that approximates 0 to the unobserved use, the LUM value is also close to 0. For example, in the case of four land use classifications, if one use has 100% of the proposition, the value of LUM4 would be calculated as 0.0000034880245, which is approximate to 0, by calculating 0.0000001 as the value of the other land uses (Im and Choi, 2019).
3. Im and Choi (2018) called the right side of Figure 1 'hidden side' because it was not noted in previous studies using CLUM. They presumed that the starting point of the hidden side would start at the left of the graph rather than at the peak of the CLUM because of the law of diminishing marginal utility and the CLUM limitation at a value of 1, which does not reflect the characteristics of land use change.
4. In this study, land use characteristic data within a 100 m radius were constructed. The analysis radius may vary depending on the purpose of the study and availability of data. This study chose 100 m because this was a simple, clear, and readily imagined measure.
5. The land use classification was set to four in this study. When the number of land use classifications was further divided into five, the R-squared value slightly increased, but the WLUM5 squared variable was not significant in the WLUM5 and WLUM5 squared model. This study only provides information based on four land use categories, because most previous studies divided the land use into four major classification levels. The WLUM squared, which is the main variable of this study, is meaningful in the four category models, and this journal has a word limit. If a future study would clarify which LUM has a positive effect on the pedestrian volume, the following analysis would be recommended. Different land uses' effect on the volume would be estimated, and then the integration effect of different uses mix would be evaluated. In addition, the related land use data (taxation data of Seoul in this study) should be built into the entire survey area. This study has limitations in both of these steps.

References

Brown BB, Yamada I, Smith K, et al. (2009) Mixed land use and walkability: Variations in land use measures and relationships with BMI, overweight, and obesity. *Health and Place* 15: 1130–1141.

Cerin E, Leslie E, du Toit L, et al. (2007) Destinations that matter: Associations with walking for transport. *Health and Place* 13(3): 713–724.

Cervero R and Duncan M (2003) Walking, bicycling, and urban landscapes: Evidence from the San Francisco Bay area. *American Journal of Public Health* 93(9): 1478–1483.

Cervero R and Kockelman K (1997) Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research D* 2(3): 199–219.

Christian HE, Bull FC, Middleton NJ, et al. (2011) How important is the land use mix measure in understanding walking behaviour? Results from the RESIDE study. *International Journal of Behavioral Nutrition and Physical Activity* 8: 55.

Diao M and Ferreira J (2010) Residential property values and the built environment: Empirical study in the Boston, Massachusetts, metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board* 2174: 138–147.

Diao M and Ferreira J (2014) Vehicle miles traveled and the built environment: Evidence from vehicle safety inspection data. *Environment and Planning A* 46(12): 2991–3009.

Duncan M, Winkler E, Sugiyama T, et al. (2010) Relationships of land use mix with walking for transport: Do land uses and geographical scale matter? *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 87(5): 782–795.

Ewing R and Cervero R (2010) Travel and the built environment: A meta-analysis. *Journal of the American Planning Association* 76(3): 265–294.

Forsyth A, Hearst M, Oakes JM, et al. (2008) Design and destinations: Factors influencing walking and total physical activity. *Urban Studies* 45(9): 1973–1996.

Frank LD, Schmid TL, Sallis JF, et al. (2005) Linking objectively measured physical activity with objectively measured urban form. *American Journal of Preventive Medicine* 28(2S2): 117–125.

Gehl J and Svarre B (2013) *How to Study Public Life*. Washington, DC: Island Press.

Hajrsouliha A and Yin L (2015) The impact of street network connectivity on pedestrian volume. *Urban Studies* 52(13): 2483–2497.

Handy S, Cao X and Mokhtarian PL (2006) Self-selection in the relationship between the built environment and walking: Empirical evidence from Northern California. *Journal of the American Planning Association* 72(1): 55–74.

Hess PM, Moudon AV and Lodsdon M (2001) Measuring land use patterns for transportation research. *Transportation Research Record: Journal of the Transportation Research Board* 1780: 17–24.

Im HN and Choi CG (2019) The hidden side of the entropy-based land-use mix index: Clarifying the relationship between pedestrian volume and land-use mix. *Urban Studies* 56(9): 1865–1881.

Jacobs J (1961) *The Death and Life of Great American Cities*. New York: Random House.

Jang S, An Y, Yi C, et al. (2017) Assessing the spatial equity of Seoul's public transportation using the Gini coefficient based on its accessibility. *International Journal of Urban Sciences* 21(1): 91–107.

Kockelman KM (1996) *Travel behavior as a function of accessibility, land use mixing, and land use balance: evidence from the San Francisco Bay area*. Master's Thesis, University of California, USA.

Korea Transport Institute (2001) *Korea transport Surveys and DB Construction Projects*. Koyang, Korea: Korea Transport Institute.

Lee JA and Koo JH (2013) The effect of physical environment of street on pedestrian volume: Focused on central business district (CBD, GBD, YBD) of Seoul. *Journal of the Korea Planning Association* 48(4): 269–286 (in Korean).

Liu XH and Griswold J (2009) Pedestrian volume modeling: A case study of San Francisco. *Yearbook of the Association of Pacific Coast Geographers* 72: 164–181.

Pan D (2013) Key transport statistics of world cities. *Journeys*, September, 105–112.

Seoul Metropolitan Government (2017) *Seoul Statistical Yearbook, the 57th Edition*. Seoul, Korea: Seoul Metropolitan Government (in Korean).

Song Y, Merlin L and Rodriguez D (2013) Comparing measures of urban land use mix. *Computers, Environment and Urban Systems* 42: 1–13.

Sung H and Choi CG (2017) The link between metropolitan planning and transit-oriented development: An examination of the Rosario Plan in 1980 for Seoul, South Korea. *Land Use Policy* 63: 514–522.

Sung H, Lee S and Cheon SH (2015) Operationalizing Jane Jacobs's urban design theory: Empirical verification from the great city of Seoul, Korea. *Journal of Planning Education and Research* 35(2): 117–130.

van Eck JR and Koomen E (2008) Characterising urban concentration and land-use diversity in simulations of future land use. *Annals of Regional Science* 42: 123–140.

Yoon Y, Choi CG, Vongpraseuth T, et al. (2014) Analyzing an impact of the location and type of parks on street vitality in Seoul. *Journal of the Korea Planning Association* 49(6): 95–107 (in Korean).

Yun NY and Choi CG (2013) Relationship between pedestrian volume and pedestrian environmental factors on the commercial streets in Seoul. *Journal of the Korea Planning Association* 48(4): 135–150 (in Korean).

Zhang M (2004) The role of land use in travel mode choice: Evidence from Boston and Hong Kong. *Journal of the American Planning Association* 70(3): 344–361.

Ha Na Im is an invited associate research fellow at Department of Urban Planning and Design Research, The Seoul Institute, Korea from 2016. She received her BA (2007) and MS (2010) from Suwon University, and PhD (2015) from Hanyang University, Seoul. She has received many academic prizes including Best paper award from The Korean Federation of Science and Technology Societies in 2015. She has authored more than 20 articles in academic journals such as *Journal of Korea Planners Association, Urban Studies*, and etc. Her research interests are built environments and its effect, walking, and entropy-based land use mix index.

Chang Gyu Choi is a Professor at Hanyang University's Graduate School of Urban Studies, Korea. With diverse academic and practical experiences, he joined Graduate School of Urban Studies, Hanyang University in 2009. He received his BA (1990), MS (1992) and PhD (1997) in Urban Planning from Hanyang University, Seoul and MsRED (2004) from Columbia University, New York. He has authored more than 60 articles in academic journals such as *Journal of Korea Planners Association, Cities, Land Use Policy, Urban Studies* and etc. His recent research interests are mixed land use, density, walking, livability, and transit-oriented development.