Article

Urban Studies

Urban Studies
2019, Vol. 56(9) 1865–1881

DOI: 10.1177/0042098018763319
journals.sagepub.com/home/usj
SAGE

The hidden side of the entropy-based land-use mix index: Clarifying the relationship between pedestrian volume and land-use mix

Ha Na Im
The Seoul Institute, Republic of Korea

Chang Gyu Choi
Hanyang University, Republic of Korea

Abstract

This study clarifies the previously unknown limitations of the entropy-based land-use mix index and suggests conditions under which the index is valid. The land-use mix index has an n-shaped relationship to dependent variables, which was evidenced by this study, but previous studies have ignored the problem. This study identified a non-linear relationship between the land-use mix index and a common dependent variable of interest, pedestrian volume. Pedestrian volume is a common measure of the vitality of a district and/or a city and a major goal of urban design and regeneration. Using mathematical analysis, simulation, and empirical analysis, this study found that the land-use mix index had an inconsistent quadratic relationship to pedestrian volume. It was confirmed that an analytical model using the land-use mix index, and that index squared, should be used together when samples representative of entire cities are tested. Otherwise, in samples from predominantly residential areas, the land-use mix index positively relates to pedestrian volume, whereas, in predominantly commercial areas, it will be negative. Previous studies failed to observe the hidden side of the entropy-based land-use mix index in commercial areas because their focus was mainly on residential areas or residents. Future studies should clarify the logical and theoretical relationships between the index and the outcome variable of interest, review the characteristics of the data and, then, implement appropriate statistical analyses by being aware of the hidden side.

Keywords

diversity, entropy-based land use mix index, land-use mix, pedestrian volume, street vitality

Corresponding author:
Chang Gyu Choi, Graduate School of Urban Studies, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, 133-791, Republic of Korea.
Email: cgchoi@hanyang.ac.kr

摘要

本研究阐明了以前未知的基于熵的土地利用综合指数的局限性，并提出了该指数有效的条件。土地利用综合指数与因变量呈 n 型关系，本研究证明了这一点，但以前的研究忽略了这个问题。本研究确定了土地利用综合指数与一个常用的因变量——行人流量——之间的非线性关系。行人流量是衡量区域和/或城市活力的常用指标，也是城市设计和再生的主要目标。利用数学分析、模拟和实证分析，本研究发现土地利用综合指数与行人流量之间有着不一致的二次方关系。我们已经证实，在对代表整个城市的样本进行测试时，土地利用综合指数分析模型和平方指数应一起使用。否则，在居住区占主导的样本中，土地利用综合指数与行人流量正相关，而在商业区主导的样本中则为负相关。先前的研究未能观察到商业区基于熵的土地利用综合指数的隐藏面，因为它们的重点主要集中在居住区或居民。未来的研究应该阐明该指数与相关结果变量之间的逻辑和理论关系，评估数据的特征，然后在明确意识到隐藏面的前提下实施合适的统计分析。

关键词

多样性、基于熵的土地利用综合指数、土地利用综合、行人流量、街道活力

Received January 2017; accepted February 2018

Introduction

Diversity in land use has been a critical target of research and practice on built environments regarding urban planning and design, transportation planning, health and preventive medicine, and property valuation (Ewing and Cervero, 2010; Frank et al., 2005; Hess et al., 2001; Song and Knaap, 2004). Mixed land use has frequently been the focus of scholars, practitioners, developers, and civil servants working with built environments. For example, Calthorpe (1993: 41) emphasised mixed land use for transit-related development, and the Congress of New Urbanism (2000) argued that street vitality could be achieved by pursuing land-use diversity.

The entropy-based land-use mix index (LUM) is likely the most often-used indicator of land-use diversity (Ewing and Cervero, 2010; Song et al., 2013). The LUM, introduced by Cervero (1989), was employed as an objective measure using geographical information systems (GIS) by Frank and Pivo (1994) and standardised by Kockelman (1996) to become an archetype for determining land-use mix. The uses of the LUM have extended from transportation-related research to include studies on real estate.

However, the logical principles for using the LUM have not yet been specifically set forth. Cervero (1989: 77), who initially proposed Shannon and Weaver’s (1949) entropy index to measure land-use diversity, did not sufficiently explain the relationship between the land-use diversity index and the dependent variable, which, in his study, was trip behaviour. Kockelman (1996: 49) was the first to propose standardising the LUM and briefly described its limitations in a note, stating ‘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’. Hess et al. (2001: 18) took a similar position, but the authors of the study focused on the spatial interaction among land-use types. In health studies, Brown et al. (2009: 1131–1132) explained the issues in detail, but they focused on the number of land-use categories.

Song et al. (2013) confirmed the LUM as the most efficient way to represent land with three or more uses through a simulation and comparison of the characteristics of various indices of land-use diversity. If research on built environments aims to derive typical values to indicate the various mixtures of land uses, an index must be usable with sufficient comprehensibility. The application of the LUM, however, continued, despite insufficient evaluation, because of its practical utility. Studies on the built environment must accurately apply this index and find ways to make better use of it.

This study aimed to analyse the LUM's logic by simulating the relationship between the LUM and pedestrian volume and by empirically analysing the two variables. It investigated these two variables because, when the LUM is used to explain variation in pedestrian volume, its characteristics and limitations are clearly displayed. In addition, pedestrian volume has been of significant interest in recent urban studies because Jacobs (1961) emphasised the importance of walking and the diversity of land use to promote the vitality of districts.

Pedestrian volume has been used to measure whether the results of efforts to improve liveability in cities have been successful or not (Gehl and Svarre, 2013). Recent studies estimating pedestrian volume or the influencing factors thereof have sparked theoretical and practical interests (Hajrasouliha and Yin, 2015; Lee and Koo, 2013; Liu and Griswold, 2009; Yun and Choi, 2013). Most of these studies have used the LUM to measure land-use diversity.

The remainder of this paper is organised as follows. The next section reviews the development of the LUM, and the previously stated alternatives to it, and presents the limitations of the previous studies. The following section demonstrates that the LUM has a quadratic nature, and a simulation confirms that the index has a non-linear 'n-shaped' relationship to pedestrian volume, which is the dependent variable of this study. The penultimate section presents the empirical test of the relationship between the LUM and pedestrian volume in Seoul city, which has high population density and land-use diversity. The results found that the coefficient's direction of effects on pedestrian volume was different for the citywide sample than for the land-use type samples. In the final section, this study's contribution is confirmed and its limitations are clarified.

Literature review

The LUM as a measure of land-use diversity

The equation used in the LUM was derived from Shannon and Weaver (1949) in the field of communications, which used Boltzmann's H-theorem in thermodynamics. Wilson (1969: 112) was the first to apply the entropy formula to urban and regional models to estimate traffic distribution between zones. Cervero (1989: 77–78) used the entropy formula to measure the land-use mix in suburban areas of the USA. Frank and Pivo (1994: 13) calculated land-use diversity based on the equation using GIS, which found a positive relationship between the LUM and travel behaviour. Kockelman (1996: 16) divided the diversity index by *ln*(*k*) to standardise it, as shown in equation (1), and Cervero and Kockelman (1997: 206) published the equation in a frequently cited journal article. The LUM scores resulting from equation (1) range from zero (when there is one land use and no mix) to one (all uses are evenly distributed):

$$\text{LUM} = -\sum_{i=1}^{k} \frac{P_i \times ln(P_i)}{ln(k)} \tag{1}$$

where k refers to the number of land-use (herein: building uses) categories, and P_i is the proportion of building area for the use of i.[1]

Depending on data availability or research goals, the area of a plot or a parcel is selected to calculate P_i, or, in other cases, the area of a building. Researchers can select specific land uses to calculate the P_i. For example, Duncan et al. (2010) developed a revised LUM that excluded recreational and other uses and assumed that it was not significantly influenced by behaviour. In most of the previous studies, *k* is constant, except for Hajna et al. (2014), who calculated the index using the precise numbers of observed land uses. The values of the LUM used in this study refer to the number of land-use classifications, which is similar to previous studies; for example, 'LUM4' is the *indicator* of four mixed land uses.

Ewing and Cervero's (2010) meta-analysis found that the LUM had been the most appropriate measure for explaining land-use diversity in studies on travel and the built environment. Some researchers have used the Herfindahl index, which is used in economics to measure concentration, to measure land-use diversity. Van Eck and Koomen (2008: 134) found that the LUM and the Herfindahl index yielded similar results, although they appear to be different. Song et al. (2013) verified that finding after simulating the two indexes. Wilson (2010) pointed out that the entropy principle is useful and extensively used in urban and regional modelling.

However, the early studies on the index, such as Cervero (1989), Frank and Pivo (1994), and Cervero and Kockelman (1997), did not clearly identify the relationship between the LUM and their dependent variables or the limitations of the LUM. Previous studies using the LUM implicitly assumed a positive relationship to the dependent variable. Although Kockelman (1996: 49) admitted that the index has limitations in the paper's note (quoted in the introduction of this study), only a few studies have paid attention to that claim.

Previous alternatives to the LUM

Hess et al. (2001: 18) pointed out that completely different land-use characteristics might have similar LUM values, which supports Kockelman (1996: 49). Brown et al. (2009: 1132)[2] indicated that the LUM remains the same regardless of the proportional land-use mix (e.g. one-quarter residential and three-quarters retail or three-quarters residential and one-quarter retail). However, these studies failed to consider the changes in continuous values, suggest alternatives, or offer ways to effectively use the index.

In health-related studies, discussion on the limitation of the LUM has been relatively active. Three major options for using the LUM have been proposed, although not as alternatives for the index's non-linear nature. First, studies on the types of land use to include in land-use diversity have been conducted. Because healthcare facilities that are relatively less used might have no significant effects on individual behaviour, some studies considered the types of land use worth including and/or suggested alternatives to the LUM (Cerin et al., 2007; Christian et al., 2011; Duncan et al., 2010; Forsyth et al., 2008). Alternatives were suggested by excluding leisure facilities or by examining the self-reported perceptions of the mixed land use. A second proposal was to change the number of land-use categories. Brown et al. (2009) analysed the relevance of such a change to individual health by changing the number of land-use categories (which is the denominator of the LUM) to two, three, or six. Hajna et al. (2014) analysed the influence of setting the denominator to 'four' or varying the number of observed land-use categories based on physical activity. Third, the interval of the LUM value was used to analyse the relationship between the land-use mix and walking behaviour by Rodriguez et al. (2009), who analysed walking to destinations and walking for exercise using the four quartiles of the LUM value.

Unlike the health studies, Manaugh and Kreider (2013) focused on the spatial interaction between different land-use types (devised by Hess et al., 2001) and proposed a new mixed land-use measurement based on the interaction. However, their proposal was limited because it is difficult to account for the mix of building uses and/or blocks found in European and Asian cities and in recent North American New Urbanism developments.

Using the LUM to research pedestrian volume

Previous studies on pedestrian volume using the LUM as the explanatory variable did not review the limitations of the index, but mostly took the conventional position. Liu and Griswold (2009) found that the LUM significantly and positively related to pedestrian volume in San Francisco, California. Lee and Koo (2013) found a positive relationship between the LUM and pedestrian volume in Seoul, South Korea, in regression analyses, but the relationship was not found in their other models. Hajrasouliha and Yin (2015) found a positive effect of the LUM on pedestrian volume using structural equation modelling. However, Yun and Choi (2013) reported a statistically significant negative effect of the LUM on pedestrian volume in Seoul.

The LUM in previous studies on the built environment was hypothesised to positively influence various dependent variables, such as walking time, number of vehicular miles travelled, public transit choice, and pedestrian volume. These studies did not explain the significantly negative or non-significant relationships between the LUM and the dependent variables. Although some previous studies have suggested alternatives to the LUM, they have never discussed whether the index has non-linear relationships with other variables. The LUM's characteristic is even more crucial when its relationships to dependent variables are interpreted, particularly when that outcome is pedestrian volume.

The relationship between the LUM and pedestrian volume

According to equation (1), when the proportional distribution between two land uses changes, the LUM changes. In Figure 1, points 'a' and 'b' have the same LUM2 (LUM with two land uses) value of 0.72, although the two parts have very different land-use mixes: point 'a' is 20% non-residential use and 80% residential use, whereas point 'b' is the reverse. Kockelman (1996: 49), Hess et al. (2001: 18), and Brown et al. (2009: 1132) previously pointed out the limitations of the LUM this way, but they did not consider changes to the index or ways to handle the limitation.

The continuously curved line of the index's values (dotted line in Figure 1) illustrates that the LUM2 values increase until the commercial and residential proportions reach 50%, and they decrease as the commercial and residential land uses change by ±10 percentage points. The LUM2 index in Figure 1 clearly presents a typical quadratic function curve, which can have an n-shaped relationship to dependent variables and is referred to herein as the 'n-shaped characteristic' of the LUM.

If a variable related to the index is added, the limitations of the index are evident. The estimated pedestrian volume calculated by applying the walking generation unit (WGU) by building use[3] yielded the line with triangles in Figure 1 indicating that pedestrian volume increases as residential land-use proportion decreases. According to the Korea Transport Surveys and Database Construction Projects in 2001, daily walking trips are generated by building uses, such that a residential building is 0.11 trips/m^2 per day, a commercial building is 0.85 trips/m^2 per day, an office building is 0.35 trips/m^2 per day, and other types of buildings are 0.33 trips/m^2 per day (Korea Transport Institute, 2001).[4] The results in Figure 1

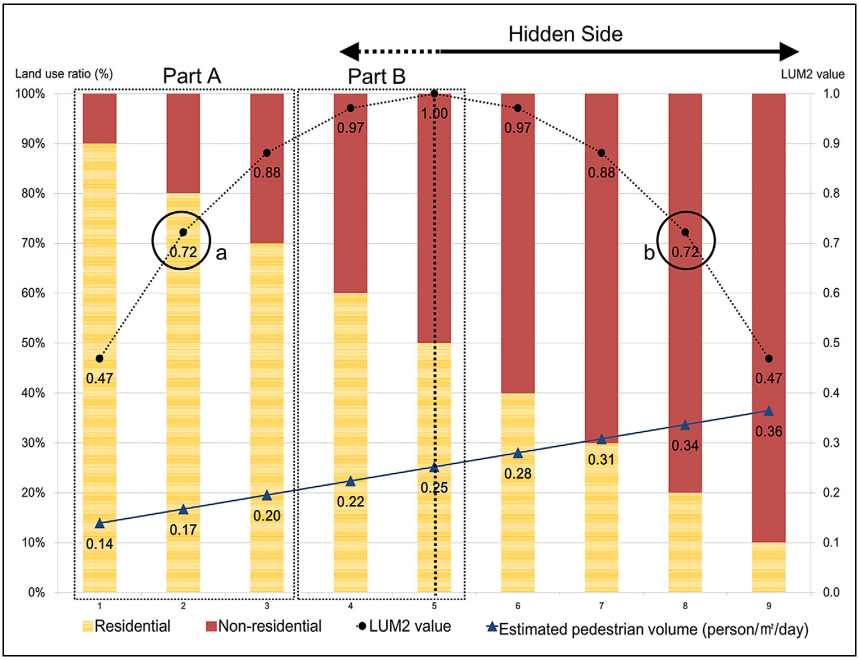

Figure 1. The relationship between LUM2 and estimated pedestrian volume.

illustrate the changes to the relationship between the LUM2 value and pedestrian volume.

The left side of the inflection point in Figure 1 shows a positive relationship between the two variables: in Part A, the LUM2 value and the estimated pedestrian volume positively correlate as the proportion of residential areas decreases. However, the right side of the figure shows a negative relationship, in which the LUM2 decreases as the estimated pedestrian volume increases. Because the previous studies did not clearly address this two-dimensional feature, this study refers to the right side of Figure 1 as the 'hidden side' of the LUM. The starting point of the hidden side, which must not have been the LUM's peak, is discussed below.

The hidden side of the LUM hides diverse land-use mixes and their various effects on pedestrian volume. The values of the previous WGUs for commercial building land uses are about eight times those of residential building land uses, and office building land uses and other building land uses are about three times higher than those of residential building land uses. If just two of the four land uses were selected and mixed to derive a 70:30 ratio (for simplicity), the value of LUM4 (LUM with four land uses) could be fixed at 0.44, which the dotted line in Figure 2 represents. The solid line in Figure 2 illustrates pedestrian volume estimated by randomly selecting two land uses using the WGUs. Therefore, the LUM4 value is always 0.44, although the 11 different land-use mixes must have different characteristics and different effects on pedestrian volume. The estimated pedestrian volume ranges from 0.18 trip/m^2 per day to 0.70 trip/m^2 per day. Thus, the influence of land-use mix on pedestrian volume varies, depending on the proportional use of larger

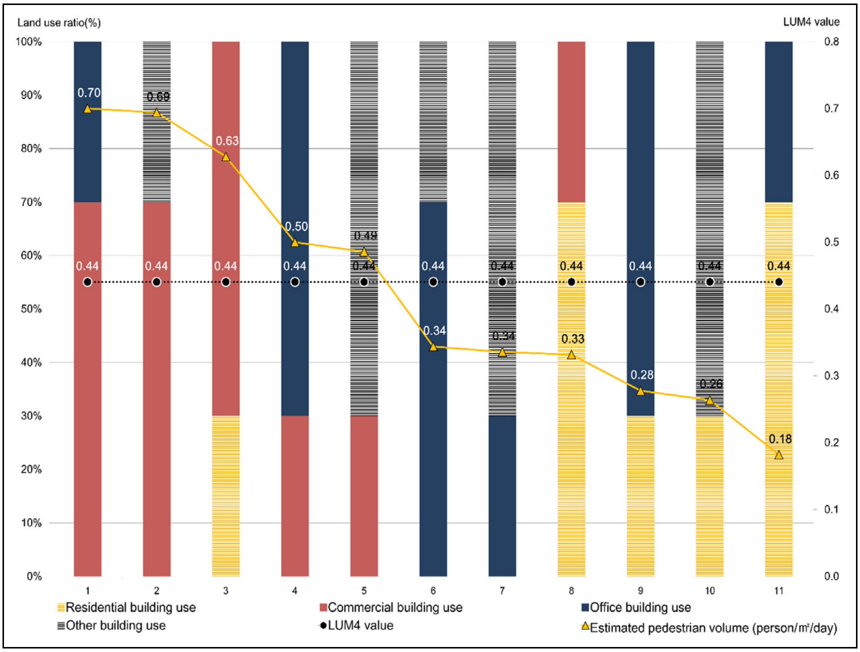

Figure 2. Estimated pedestrian volume from different land-use mixes, but the same value of LUM4.

WGUs rather than on the proportions of residential land-use areas (Figure 2).

The results shown in Figures 1 and 2 confirm, first, the LUM's n-shaped relationship to pedestrian volume and, second, that, depending on the main use of a district or street, the sign of the statistical effect of the LUM on pedestrian volume could change. A mixture of residential land use with predominantly commercial or office land uses has a negative effect on pedestrian volume, whereas a mixture of commercial or office land uses with predominantly residential land use has a positive effect on pedestrian volume.

This argument regarding the LUM could be extended from its relationship to pedestrian volume to other relationships between human behaviours and mixed land uses. Even when a given land-use type has permanently positive effects on an individual behaviour, the LUM value will decrease when the proportion of the land-use type increases above $1/N$ (where '1' is the LUM). If there is a land-use type that permanently negatively influences behaviours, the LUM value will decrease when the proportion of the land-use type increases above $1/N$, too. The two different conditions must oppositely affect people's behaviours, but the LUM value only shows a decrease of the same magnitude. This occurs near the peak of the LUM curve, possibly in Part B near the top of the land-use distribution in Figure 1.

If the influence of a land-use type were to follow the law of diminishing marginal utility, no type of land use would have a permanent and positive effect on behaviour. The marginal effect of a land-use type's proportional increase would be positive to a certain point and, then, it gradually would decrease, depicted as a typical S-shape. For example, in mixed-use developments, the size of a shopping mall could influence the amount of walking time for residents only to a certain point, after which the influence would be in the 'hidden side' of Figure 1. It does not begin at the peak of the LUM curve; its starting point is somewhere in the left side of Figure 1. Therefore, as Figure 1 shows, 'Part B' exists, but it cannot be clearly defined.

The results of previous studies on the built environment that found positive influences of the LUM on behaviour are assumed to be limited by availability heuristic or biased sampling. These studies might have used samples in specific areas, such as residential areas, or samples of certain types of residents, because most of a city's land uses are dominated by residential rather than commercial and/or business land uses (e.g. Cervero and Duncan, 2003; Frank et al., 2008; Zhang, 2004). The research samples having commercial or business districts would not significantly influence the relationship of the LUM to a dependent variable.

Empirical study

Case study area and pedestrian volume data

Seoul, South Korea, was chosen for this study's empirical test because it has a high population density and diverse mixed land uses (Sung et al., 2014, 2015) expected to yield an effect of the LUM on pedestrian volume. Furthermore, the city has previously surveyed pedestrian volume on a large scale and offers parcel-based building land-use data.

Pedestrian volume data were obtained by the city of Seoul through a survey conducted from August to November of 2009. The data provide the numbers of pedestrians passing through 10,000 selected points in Seoul between 7:30 a.m. and 8:30 p.m., Monday through Saturday. The data cover the largest number of locations in pedestrian volume data surveyed by the city government to date.

Variables and descriptive statistics

The dependent variable was the average pedestrian volume on Monday through Friday; Saturday was excluded because of differences in behaviour at weekends. LUM4 was calculated for mixed land uses of residential, commercial, office, and other land-use types. To control for the effects of other factors, independent variables, such as the physical street environment, accessibility, and area characteristics, were included. These variables were chosen based on previous studies of neighbourhood environmental factors and pedestrian volume (Hajrasouliha and Yin, 2015; Lee and Koo, 2013; Liu and Griswold, 2009; Sung et al., 2013; Yun and Choi, 2013).

The 2010 taxation ledger was used to calculate each land-use type's floor area. Because these data contained information on all uses of each building, they provided more detailed land-use information than other building registration data that only have one use of a building as a representative use. The total floor areas by land-use type were calculated for each parcel, and total floor areas were calculated by applying the proportion of the area within the boundary. Total floor areas and proportions by land-use type within the radius were derived. Excluding the missing data and/or vacant areas resulted in 9728 points and buffers as the final sample for analysis.

A radius of 100 m around the survey locations was used to compute the land-use diversity. As Duncan et al. (2010: 783) pointed out, the geographical scale of a land-use mix is critical, but it is impossible to identify precise boundaries. Frank and Pivo (1994) and Cervero and Kockelman (1997) used US census tracts, Cervero and Duncan (2003) applied a one-mile radius, Frank et al. (2005) employed a 1-km radius, Cerin et al. (2007) used Australian census districts, and Rodriguez et al. (2009) used 200 m. To define the radius for this study, the R^2 values of the regression models were compared by changing the 50-m radius up to 200 m; ultimately, 100 m was used because it is the distance that is easiest to imagine,

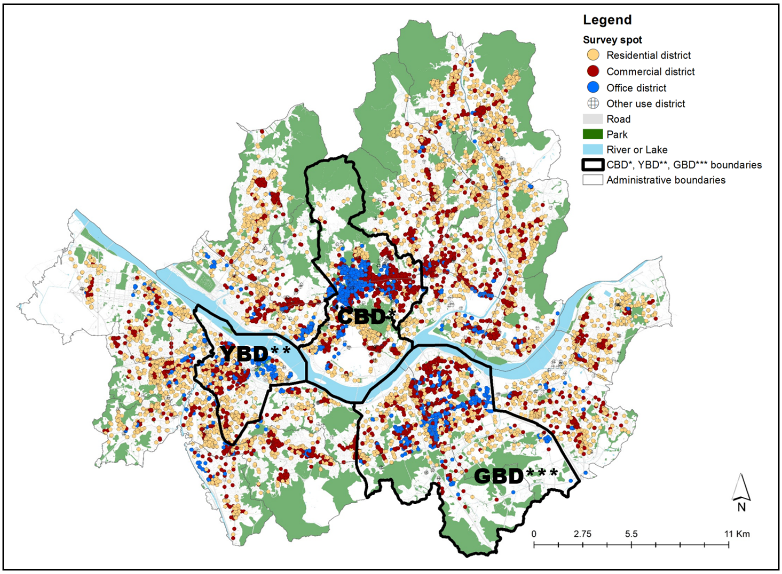

Figure 3. Survey locations of pedestrian volume and district classification in Seoul.
Notes: * CBD is Central business district in Seoul. ** YBD is Yeong-deung-po/Yeo-ui-do business district in Seoul. *** GBD is Gangnam business district in Seoul.

although the R^2 value tends to increase when the radius increases.

The physical environmental data at the street level, such as walkway width/type, slope, fencing, Braille block,[5] and so on, were constructed based on the on-site conditions recorded by surveyors during the pedestrian volume survey.[6] Using cluster analysis of the building floor uses (classification criteria are described below), this study divided the surveyed locations into four land-use types: residential, commercial, office, and other (Figure 3).

Using GIS, accessibility was calculated as the Euclidean distance from the pedestrian volume survey point to the nearest public transit facility, such as subways and bus stops. Distances to public transit from the survey locations were negatively related to pedestrian volume (Table 1). Street type was coded 1 = pedestrian-only (54.6%) and 0 = vehicular/pedestrian mixed (45.4%). The proportions with a bus stop or a subway entrance within a 50-m radius were 23.1% and 7.0%, respectively (Table 1).

Differences in area characteristics were coded 1 = in the central business/commercial districts[7] (33.5%) and 0 = otherwise (66.5%) because commercial/business districts were assumed to experience more pedestrians than others (Table 1 and Figure 3). Because Seoul is divided by the Han River ('big river' in Korean), 1 = Gangnam (South) area and 0 = Gangbuk (North) area. Gangbuk includes the historical capital of Korea, which has been there since the 15th century, and relatively old developed areas, whereas Gangnam is a relatively new developed area, comprising 47.0% of the sample (Table 1 and Figure 3).

The n-shaped relationship between the LUM and pedestrian volume

Previous simulation results suggest that the LUM has an n-shaped relationship to

Table 1. 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, 100 m)						
LUM4	Entropy index of residential, commercial, office, and other use		0.58	0.24	0.00	1.00
LUM4 squared			0.40	0.25	0.00	0.99
Physical environment at the street level						
Sidewalk width	Sidewalk 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. of 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		
Accessibility						
Distance to bus stop	Distance to the nearest facility (m)		94.7	72.5	0.9	699.0
Distance to subway station			419.7	367.5	0.0	3678.3
Regional characteristics						
CBDs[a,b]	Inclusion of Seoul CBD, YBD, GBD	0 = no	6466	66.5	0	1
		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						
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	18211	0	182,384
Other use area			2913	5987	0	201,933
Residential ratio	Building use proportion within a 100 m radius (%)		49.7	31.5	0.0	100.0
Commercial ratio			26.7	19.6	0.0	100.0
Office ratio			16.3	20.3	0.0	100.0
Other use ratio			7.3	13.2	0.0	100.0

Notes: [a] Frequency and percent.

[b] Seoul is believed to have three central business/commercial districts.

Table 2. Empirical regression models on pedestrian volume for whole city of Seoul.

Variable	Model 1		Model 2	
	Coef.	β	Coef.	β
Land-use mix (diversity, 100M)				
LUM4	859.26***	0.05	4828.72***	0.31
LUM4 squared			−3892.05***	−0.26
Physical environment at the street level				
Sidewalk width (m)	274.60***	0.16	271.33***	0.16
Pedestrian-only street (0 = no, 1 = yes)	854.37***	0.11	845.48***	0.11
Street furniture (0 = no, 1 = yes)	427.66***	0.03	415.74***	0.03
Braille block (0 = no, 1 = yes)	360.73***	0.04	362.57***	0.04
Slope (0 = no, 1 = yes)	−554.54***	−0.06	−532.37***	−0.06
Fence (0 = no, 1 = yes)	206.56**	0.02	227.08**	0.02
No. of lanes	94.01***	0.06	100.96***	0.06
Bus lane (0 = no, 1 = yes)	1166.40***	0.08	1176.41***	0.08
Bus stop (0 = no, 1 = yes)	360.76***	0.04	354.14***	0.04
Subway (0 = no, 1 = yes)	2294.98***	0.16	2299.42***	0.16
Crosswalk (0 = no, 1 = yes)	−180.72**	−0.02	−171.83**	−0.02
Accessibility				
Distance to bus stop (m)	−3.23***	−0.06	−3.04***	−0.06
Distance to subway station (m)	−1.45***	−0.14	−1.45***	−0.14
Regional characteristics				
CBDs (0 = no, 1 = yes)	927.29***	0.12	919.85***	0.12
Region type (0 = North, 1 = South)	−170.62**	−0.02	−143.02**	−0.02
Intercept	821.86***		24.58	
Number of obs.	9728		9728	
R-squared	0.191		0.195	
Adj. *R*-squared	0.190		0.194	
F	143.55***		138.55***	

Notes: Dependent variable is pedestrian volume on weekdays.
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

pedestrian volume, and the LUM4^2 has a negative relationship to it. However, when residential building land uses are dominant in a sample, the LUM positively relates to pedestrian volume. To empirically test this, this study tested for a non-linear relationship between the LUM and pedestrian volume net of the effects of the control variables (Table 2).

Model 1 in Table 2 tested the effect of LUM4 and found that the influence of the LUM4 was positive, similar to the results of previous studies (e.g. Sung, 2014; Yoon et al., 2014). However, Model 2, which tested LUM4 and LUM4^2, found a positive effect of LUM4 and a negative effect of LUM4^2 on pedestrian volume. Model 2 had a higher coefficient of determination (R^2) than Model 1. Controlling for the effects of LUM4, pedestrian volume *decreased* as LUM4^2 *increased* (Model 2). This finding *supports* the quadratic n-shaped characteristic of the LUM and the effect of considerable (over)sampling in the residential areas, and it confirmed the study's simulation results. We recommend that future studies that select citywide samples use the approach of Model 2 and test LUM4 and LUM4^2 together.

The effects of the control variables in Model 2 (Table 2) were logical and the

Table 3. District classification by using the cluster analysis based on building use proportions of districts.

	Residential district (N = 4970)	Commercial district (N = 3178)	Office district (N = 1206)	Other use district (N = 374)	F-test
Residential building use (%)	76.22	27.64	8.46	16.59	11,455.07***
Commercial building use (%)	14.82	47.86	23.96	14.70	4609.88***
Office building use (%)	5.38	17.05	61.39	8.89	10,414.36***
Other building use (%)	3.58	7.46	6.19	59.82	5955.25***
Weekday average pedestrian volume (person/day)	1961.9	4181.6	4778.2	3130.1	353.53***

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

variance inflation factors were less than two, satisfying multicollinearity standards.[8] In the physical environment, sidewalk width, pedestrian-only street, street furniture, Braille block,[9] fence, number of traffic lanes, and bus-only lanes positively influenced pedestrian volume, whereas the slope of the road and crosswalks had negative effects. Public transit facilities (bus and subway) within a 50-m radius of the survey points positively influenced pedestrian volume, and the distances to those facilities had negative effects. The CBD, YBD, and GBD had positive relationships with pedestrian volume, and the North of the Han River had a positive effect on pedestrian volume. The directions of the coefficients' signs were constant across the models.

Relationship between the LUM by districts' land-use types and pedestrian volume

A cluster analysis was performed using the ratio of the four land-use classifications in 100-m radiuses of the survey point to distinguish among the districts' land-use types. By referring to the average proportion by land use (see Table 3 and Figure 3), the districts were classified as: (1) residential, (2) commercial, (3) office, or (4) other land use. Commercial and office districts had more than twice the pedestrian volume of the residential districts, and the 'other' land-use category had about 1100 people more than the residential category on average weekdays (Table 3).

The above simulation found that the positive or negative direction of the effect of the LUM on pedestrian volume was determined by the characteristics of land-use mix other than the dominant land-use type. It was expected that pedestrian volume would increase in the residential land-use districts as the land-use diversity increased, whereas it would decrease in the commercial districts, and it is hard to predict in the office districts and other land-use districts (see Table 3).

Table 4 shows that the relationship between the LUM4 and pedestrian volume was significantly positive for residential districts and negative for the commercial districts as expected. Because residential building land use has had the lowest WGU and commercial land use has had the highest WGU (Korea Transport Institute, 2001), pedestrian volume increases when the proportion of non-residential use in predominantly residential areas increases, and the pedestrian volume decreases when the proportion of non-commercial use in the commercial areas increases.

Office land use and other types of land use have similar WGUs, and the estimated pedestrian volume of the two land uses exists between those residential and commercial uses. If the building land-use mix

Table 4. Regression models for district's types of land use.

Variable	Residential district		Commercial district		Office district		Other use district	
	Coef.	β	Coef.	β	Coef.	β	Coef.	β
Land use mix (diversity, 100 M)								
LUM4	1132.21***	0.13	−5628.77***	−0.17	−992.41*	−0.05	1240.01**	0.11
Physical environment at the street level								
Sidewalk width (m)	114.74***	0.11	352.33***	0.16	215.89***	0.12	157.09***	0.15
Pedestrian-only street (0 = no, 1 = yes)	382.24***	0.10	936.99***	0.10	1007.86***	0.10	1247.01***	0.18
Street furniture (0 = no, 1 = yes)	125.37	0.02	843.38***	0.04	277.54	0.02	−62.38	−0.01
Braille block (0 = no, 1 = yes)	161.91**	0.03	433.78**	0.04	513.63*	0.05	457.92	0.07
Slope (0 = no, 1 = yes)	−208.55***	−0.05	−426.88**	−0.03	−990.46***	−0.10	194.61	0.03
Fence (0 = no, 1 = yes)	255.95***	0.05	376.71	0.03	−36.89	0.00	−596.98*	−0.09
No. of lanes	49.58***	0.05	87.99*	0.04	136.79**	0.09	68.64	0.06
Bus lane (0 = no, 1 = yes)	447.72***	0.05	1561.83***	0.09	1274.50***	0.08	884.63*	0.09
Bus stop (0 = no, 1 = yes)	217.65***	0.05	516.67**	0.05	765.98**	0.07	398.05	0.06
Subway (0 = no, 1 = yes)	1904.90***	0.19	2083.21***	0.12	1847.19***	0.14	1729.42**	0.14
Crosswalk (0 = no, 1 = yes)	180.69***	0.05	−502.11**	−0.05	−780.49***	−0.08	−91.12	−0.01
Accessibility								
Distance to bus stop (m)	−1.14***	−0.05	−6.00***	−0.08	−2.31	−0.03	−1.50	−0.04
Distance to subway station (m)	−0.59***	−0.11	−1.79***	−0.13	−2.75***	−0.16	−1.44***	−0.19
Regional characteristics								
CBDs (0 = no, 1 = yes)	−88.54	−0.02	482.86***	0.05	2115.45***	0.21	207.48	0.03
Region type (0 = North, 1 = South)	311.86***	0.08	−26.19	0.00	−694.33***	−0.08	−316.22	−0.05
Intercept	615.32***		6042.95***		2888.22***		1145.42	
Number of obs.	4970		3178		1206		374	
R-squared	0.171		0.171		0.240		0.239	
Adj. *R*-squared	0.168		0.167		0.229		0.205	
F	63.9***		40.66***		23.42***		6.99***	

Note: Dependent variable is pedestrian volume on weekdays.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

were to change in the office districts or other use districts where those uses are relatively dominant, the relationship to pedestrian volume would not be unidirectional; it would be bi-directional. Therefore, it was expected that the office land-use models and the 'other' land-use models would have different results (Table 4). It is likely that the different results on the LUM's effects found by previous studies were because of the LUM's characteristics and differences in sampling.

There was no evidence of multicollinearity with respect to the control variables in Table 4. Higher pedestrian volume was found in wider-sidewalk, pedestrian-only streets, and near subway entrances within a 50-m radius. Where the distances to the subway station were short, pedestrian volume was higher. Some variables, such as Braille blocks, fences, and number of lanes, did not have consistent relationships to pedestrian volume across the land-use types of districts, and further investigation of this finding should be carried out in the future because, although important, it was not within the scope of this study.

Conclusion

Based on growing interest in land-use diversity, the present study examined the characteristics of the LUM, which is a typical measure of that diversity, and investigated its relationship to pedestrian volume as a dependent variable. This study's findings confirmed that the previous studies about the LUM, or that applied it and/or suggested alternatives to it, did not investigate its limitations regarding its non-linear characteristic or consider its so-called 'hidden side'. Based on this study's confirmation of these limitations, the characteristics of the LUM were assessed through simulations and empirical analyses. As a result, the following three conclusions were reached.

First, the LUM had an n-shaped characteristic. Extreme effects of mixed land uses were found, in which there were clear differences in the effects by land-use type regarding pedestrian volume. Previous studies using the LUM did not pay attention to its limitation. It is strongly suggested that, in future studies, logical and theoretical relationships between the LUM and the outcome variables of interest be initially established, followed by clarification of the features of the sample being tested, particularly when the effect of the LUM has been logically unclear.

Second, the LUM and pedestrian volume had a negative quadratic relationship in this study. It is logical that the LUM^2 would negatively influence pedestrian volume when samples are collected representative of a city (to include all possible land-use mixes). However, because the LUM has a strong influence on residential areas and residential building areas, which comprise substantial proportions of cities, it was confirmed that the LUM^2 negatively influenced pedestrian volume when the effects of the LUM were controlled in the multiple regression model.

Third, if a study using the LUM were conducted where the effects of mixed land uses on pedestrian volume were clear, such as predominantly residential or commercial areas, the LUM could be used alone as an independent variable to examine the effects of mixed land use. Based on the WGU used in this study, commercial building land use had about eight times as much pedestrian generation per unit area as a residential building land-use area. In this case, the land-use mix or the LUM value in the residential areas had positive effects on walking, but it had negative effects on walking in the commercial areas. When two variables have clear directions, it is not necessary to use the LUM^2 with the LUM.

Although the present study investigated the logical limitations of the conventional LUM measure and revealed potential problems in interpreting the ways that a land-use mix might influence a dependent variable, it did not identify alternatives to solve the problem. As discussed above, the LUM is one of the best indicators for explaining land-use diversity, but it has a critical problem as an independent variable. To address the problem, future studies on the effects of land-use mix and on the LUM should consider its limitations before implementing it, and, if possible, suggest ways to overcome or minimise this challenging problem. They will obtain more accurate results than in the past if they consider the land-use classifications issues (e.g. Cerin et al., 2007 and Duncan et al., 2010).

Acknowledgements

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 this study's initial version, as well as three anonymous reviewers for their thoughtful comments and suggestions to improve this article's structure and 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 a Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2016R1A2B4010260).

Notes

1. When calculating $\ln(P_i)$, when a value of P_i 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 zero to an unobserved use, the LUM value is also close to zero. For example, in the case of a four-land-uses classification, if one land use had 100% of proposition, the value of LUM4 would be calculated as 0.0000034880245, which is approximate to zero, by calculating 0.0000001 as the value of the other land use.
2. Brown et al. (2009) insisted on six limitations of the LUM, but the explanations have similar arguments.
3. This study used the building uses' area data because they have more information on land-use intensity than land-use classification data.
4. The projects have been surveyed several times by the Korea Transport Institute, a research institute of the Korean government. Because the 2001 survey has more detailed building-use classifications, this study used the 2001 survey data.
5. This is a road sign for blind people, usually using raised dots directly on the pavement.
6. Although 2000 surveyors recorded on-site circumstances at 10,000 survey sites, there is a possibility of some subjective criteria applied to the presence of a slope, street furniture, and so on.
7. Seoul is believed to have three central business/commercial districts: Central Business District (CBD), Gangnam Business Districts (GBD), and Yeong-deung-po/Yeo-ui-do Business District (YBD) (Figure 3).
8. Centre line of road was eliminated in Table 2's models, because it had a 0.7 correlation coefficient in its association with the number of traffic lanes.
9. It possibly means that the street with Braille blocks had been built with the efforts of street design.

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.

Calthorpe P (1993) *The Next American Metropolis: Ecology, Community, and the American Dream.* New York, NY: Princeton Architectural Press.

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

Cervero R (1989) *America's Suburban Centers: The Land Use–Transportation Link*. Boston, MA: Unwin Hyman.

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.

Congress of New Urbanism (2000) *Charter of the New Urbanism*. New York: McGraw Hill.

Duncan M, Winkler E and Sugiyama T (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 and Pivo G (1994) Impacts of mixed use and density on utilization of three-modes of travel: Single-occupant vehicle, transit, and walking. *Transportation Research Record* 1466: 44–52.

Frank LD, Bradley M, Kavage S, et al. (2008) Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation* 35(1): 37–54.

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.

Hajna S, Dasgupta K, Joseph L, et al. (2014) A call for caution and transparency in the calculation of land use mix: Measurement bias in the estimation of associations between land use mix and physical activity. *Health and Place* 29: 79–83.

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

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.

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

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, Berkeley.

Korea Transport Institute (2001) *Korea Transport Surveys and DB Construction Projects*. Goyang, South 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.

Manaugh K and Kreider T (2013) What is mixed use? Presenting an interaction method for measuring land use mix. *The Journal of Transport and Land Use* 6(1): 63–72.

Rodriguez DA, Evenson KR, Roux AVD, et al. (2009) Land use, residential density, and walking. *American Journal of Preventive Medicine* 37(5): 397–404.

Shannon CE and Weaver W (1949) *The Mathematical Theory of Communication*. Urbana, IL: The University of Illinois Press.

Song Y and Knaap GJ (2004) Measuring the effects of mixed land uses on housing values.

Regional Science & Urban Economics 34: 663–680.

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 (2014) Diagnosis on the non-linear association of built environment with walking activity in residential areas: Focused on the difference of walking purposes for physical environment based on Jacobs' life. *Journal of the Korea Planning Association* 49(3): 159–174 (in Korean).

Sung H, Go DH and Choi CG (2013) Evidence of Jacobs's street life in the great Seoul city: Identifying the association of physical environment with walking activity on streets. *Cities* 35: 164–173.

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.

Sung H, Lee S and Jung S (2014) Identifying the relationship between the objectively measured built environment and walking activity in the high-density and transit-oriented city, Seoul, Korea. *Environment and Planning B* 41: 637–660.

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.

Wilson A (2010) Entropy in urban and regional modelling: Retrospect and prospect. *Geographical Analysis* 42: 364–394.

Wilson AG (1969) The use of entropy maximising modes in the theory of trip distribution, mode split and route split. *Journal of Transport Economics and Policy* 3(1): 108–126.

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.