

# UC San Diego

## UC San Diego Previously Published Works

### Title

Inter-rater reliability of streetscape audits using online observations: Microscale Audit of Pedestrian Streetscapes (MAPS) global in Japan

### Permalink

<https://escholarship.org/uc/item/5834828t>

### Authors

Saito, Yoshinobu

Oguma, Yuko

Inoue, Shigeru

et al.

### Publication Date

2022-12-01

### DOI

10.1016/j.pmedr.2022.102043

Peer reviewed



## Inter-rater reliability of streetscape audits using online observations: Microscale Audit of Pedestrian Streetscapes (MAPS) global in Japan

Yoshinobu Saito<sup>a,\*</sup>, Yuko Oguma<sup>b,c</sup>, Shigeru Inoue<sup>d</sup>, Raoul Breugelmans<sup>e</sup>, Hiroyuki Kikuchi<sup>d</sup>, Koichiro Oka<sup>f</sup>, Shinpei Okada<sup>g</sup>, Noriko Takeda<sup>h</sup>, Kelli L. Cain<sup>i,j</sup>, James F. Sallis<sup>i,j</sup>

<sup>a</sup> Faculty of Sport Management, Nippon Sport Science University, Yokohama, Kanagawa, Japan

<sup>b</sup> Sports Medicine Research Center, Keio University, Yokohama, Kanagawa, Japan

<sup>c</sup> Graduate School of Health Management, Keio University, Fujisawa, Kanagawa, Japan

<sup>d</sup> Department of Preventive Medicine and Public Health, Tokyo Medical University, Shinjuku, Tokyo, Japan

<sup>e</sup> Department of English, Kansai Medical University, Hirakata, Osaka, Japan

<sup>f</sup> Faculty of Sport Sciences, Waseda University, Tokorozawa, Saitama, Japan

<sup>g</sup> Physical Education and Medicine Research Foundation, Tomi, Nagano, Japan

<sup>h</sup> Center for Promotion of Higher Education, Kogakuin University, Shinjuku, Tokyo, Japan

<sup>i</sup> Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, Australia

<sup>j</sup> Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, USA

### ARTICLE INFO

#### Keywords:

Built environment  
Physical activity  
Walkability  
Urban design

### ABSTRACT

This study aimed to evaluate the inter-rater reliability of streetscape audits among online observations using the Microscale Audit of Pedestrian Streetscapes-Global version (MAPS-Global) in Japan.

MAPS-Global observations were conducted on routes with distances ranging from 400 to 725 m from a residence toward a non-residential destination. Google Street View audits were independently conducted by two trained raters on each route. A tiered scoring system was applied to summarize the items at multiple levels of aggregation. Positive and negative valence scores were created based on the expected association with physical activity. Inter-rater reliability analyses were performed using kappa statistics or intraclass correlation coefficients (ICC).

Of the 32 older adults participating in an intervention study in the community-wide physical activity promotion project in Fujisawa City, 19 addresses were used, excluding those with nearby addresses. Results demonstrated “excellent” agreement for most of the summary scores analyzed (kappa or ICC values of 0.75 or higher [80.4 %]), while 6.5 % of items exhibited “good” agreement (ICC = 0.60–0.74). By contrast, only 13.0 % of the scales had ICC values lower than 0.60 (“fair” or “poor” reliability). The results illustrated high reliability for the grand summary scores and composite subscale measures. However, caution should be exercised when interpreting subscale scores for less frequently observed negative attributes and aesthetic/social characteristics. The results presented in this study support the application of online observations using MAPS-Global in urban areas of Japan, which could be implemented to inform decisions related not only to physical activity but also to traffic safety.

### 1. Introduction

According to ecological models of health behaviors, the determinants of physical activity (PA) operate at multiple levels of influence, including environmental aspects. (Sallis et al., 2006) Environmental factors can be categorized into macrolevel built environments, which are large scale and difficult to change quickly, and

microscale elements of the environment, which are small scale and can be altered more rapidly and at lower costs. (Sallis et al., 2011) Specific indicators for macrolevel built environments relevant for physical activity include mixed land-use, street connectivity, residential density, and proximity to recreation facilities, which can be assessed using geographic information systems. In addition, a walkability index (calculated primarily using residential density, retail floor area ratio,

\* Corresponding author at: Mailing address: 1221-1 Kamoshida-cho, Aoba-ku, Yokohama, Kanagawa 227-0033, Japan.

E-mail address: [ysaito@nittai.ac.jp](mailto:ysaito@nittai.ac.jp) (Y. Saito).

<https://doi.org/10.1016/j.pmedr.2022.102043>

Received 19 June 2022; Received in revised form 2 November 2022; Accepted 5 November 2022

Available online 7 November 2022

2211-3355/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

land use mix, and intersection density) has been developed to measure walkability in cities. (Frank et al., 2010) The evaluation of microscale environmental features can be done with audit tools based on observations that include the width and quality of sidewalks and bicycle paths and the facilities and landscape at intersections. (Clifton et al., 2007; Brownson et al., 2009; Cerin et al., 2011; Eyler et al., 2015; Millstein et al., 2013; Sallis et al., 2015; Bader et al., 2015; Steinmetz-Wood et al., 2019) Several studies have illustrated that microscale environmental features have strong associations with PA, especially active transport to destinations, independent of walkability evaluated in the context of a macrolevel built environment. (Cain et al., 2014; Molina-Garcia et al., 2020; Sallis et al., 2015) In Japan, where this study was conducted, half of all fatal accidents involving pedestrians and bicyclists occur within 500 m of the victim's home. (Ministry of Land and Infrastructure, Transport and Tourism, 2022) Therefore, investigation and improvement of neighborhood microscale environments can contribute not only to promoting physical activity but also to traffic safety.

A typical audit tool for assessing microscale environmental features is the Microscale Audit of Pedestrian Streetscapes (MAPS), which has been found to have moderate to excellent inter-rater reliability. (Cain et al., 2014; Millstein et al., 2013) MAPS has also been validated for PA across multiple age groups. (Cain et al., 2014) Furthermore, based on MAPS, items developed across several continents were utilized to create the global version (MAPS-Global), which is suitable for international use. (Cain et al., 2018) In a study of five cities worldwide (Melbourne, Ghent, Curitiba, Hong Kong, and Valencia) using MAPS-Global, the inter-rater reliability of on-street, live and online, and local and remote online observations was tested and found to have moderate to excellent reliability for all data collection modes. (Cain et al., 2018; Fox et al., 2021; Queralt et al., 2021; Vanwolleghem et al., 2016) Queralt et al. conducted a study using MAPS-Global in the above five cities and confirmed the high reliability of almost all the items for both on-street and online observations. However, aesthetic and social environment variables had the lowest overall reliability values, although they were still in the "good to excellent" category. (Queralt et al., 2021) Fox et al. reported a significant association between measurements of local on-street observations conducted in the same five cities and remote online observations conducted in the United States. (Fox et al., 2021) Thus, online observations using MAPS-Global have been confirmed to be highly reliable.

Remote observations using online imagery such as Google Street View are typically less time-consuming and less expensive than on-street observations. In addition to reducing travel costs, these remote observations are particularly useful for assessing geographically dispersed or international sites. (Taylor et al., 2011; Wilson et al., 2012).

The MAPS-Global tool was the first audit instrument designed for international use. Although evidence of its utility has accumulated, further evaluation of its performance in countries with diverse built environments and cultural characteristics is imperative. Therefore, this study aimed to evaluate the inter-rater reliability of MAPS-Global using online observations in one region of Japan.

## 2. Methods

### 2.1. Residential addresses

This study sought responses from older people participating in a group exercise intervention study of the 'Fujisawa Plus Ten' project, a community intervention to promote PA in Fujisawa City, Japan. (Saito et al., 2018; Saito et al., 2021) The city is located in an urban area approximately 50 km from Tokyo, with a population of 441,708 and a population density of 6,350/km<sup>2</sup> (as of January 1, 2022). The southern part of the city faces the sea and is a popular tourist destination. Several public transportation systems are available within the city, including buses, trains, subways, and monorails. The target number of participants was set at 65, as in the previous studies. (Cain et al., 2018) Thirty-two

participants in two groups consented to the present study. The study participants only provided information on their address and basic attributes, as obtained in the previous studies. (Saito et al., 2018; Saito et al., 2021) Study staff identified likely walking routes from home addresses, as was done with other MAPS Global studies. (Cain et al., 2018) Participants were randomly excluded if they had the same address (e.g., the same apartment building, or lived together as a couple) or were in the same neighborhood with the same route to destinations such as stores, services, parks, and schools. The final evaluation target was 19 addresses.

Written consent was obtained from all participants, and the study was approved by the Research Ethics Review Committee of the Graduate School of Health Management, Keio University (Acceptance No. 2019–21).

### 2.2. MAPS-Global tool

The development of MAPS-Global was part of the International Physical Activity and the Environment Network (IPEN) Adolescent study and led by the IPEN Coordinating Center. *International Physical Activity and Environment Network (2022)* MAPS-Global was based on the original MAPS tool developed and validated in the US. (Cain et al., 2014; Millstein et al., 2013) MAPS-Global was modified substantially by drawing on items from built environment instruments developed on multiple continents. (Millstein et al., 2013; Pikora et al., 2002; Spittaels et al., 2010; Cerin et al., 2011; Dunstan et al., 2005; Griew et al., 2013; Jones et al., 2010; Katzmarzyk et al., 2013; Oyeyemi et al., 2013; Adlakha et al., 2016) Wording and scoring were altered for greater international applicability and consistency within MAPS-Global. Based on numerous international data, the instrument was found to be feasible and reliable. (Cain et al., 2018; Fox et al., 2021; Queralt et al., 2021).

This study used the Japanese version of the MAPS-Global tool to collect microscale environmental features required for the assessment. The adaptation of the tool involved the following steps: 1) translation into Japanese, 2) a pilot round with data collection on five sample routes and translation modifications, and 3) the main data collection round with the final tool. The Japanese version was prepared in consultation with the initial development team (YS, YO, and SI), and a Japanese translation was performed by the research team who also referred to previous studies. (Vanwolleghem et al., 2016) The Japanese translation by the research team was checked by a bilingual expert specializing in medical English (RB), and the expressions were modified to prevent compromising the original version.

The online pilot data collection for the five pilot routes was determined by YS and conducted by three observers other than the author. These were different routes from the main data collection, which included all four MAPS-Global sections. A training manual, translated into Japanese, was provided to the observers. After the pilot data collection was completed, the observers were interviewed to obtain their input for improving the measure, and the final Japanese version was prepared by simplifying the wording without compromising the original meaning.

The main data collection was conducted using online Google Street View images by two graduate students (two of the pilot route observers) who had completed training to meet the certification requirements of the MAPS-Global tool. Details of the development and design features of the MAPS tool and certification requirements can be found in other papers. (Cain et al., 2018; Millstein et al., 2013).

In an international context, MAPS-Global captures a wide range of pedestrian- and bicycle-centric environmental features of streets and surrounding areas. The tool has 123 items, which are divided into four sections: (1) the route section tracks land-use characteristics and features of the entire route defined by origin–destination pairs, (2) segment-level microscale characteristics assess block faces between intersections, (3) the crossing section collects intersection information, and (4) the cul-de-sac section tracks dead ends and cul-de-sac features.

The route and cul-de-sac sections captured features of the built environment on either side of the street. By contrast, the segment and crossing sections mainly collected attributes on one side of the street to simulate exposure to environmental factors between the home and nearby destinations. MAPS-Global identified mid-block crosswalks at the segment level. In the route section, land-use, streetscape, aesthetic and social characteristics (e.g., speed limits, social environment, and aesthetics) that generally applied to the route were obtained. Segment-level measurements assessed characteristics that are likely to vary along the route and included sidewalk characteristics, buffers between the street and sidewalk, trees, and building composition. The crossing section analyzed pedestrian protection features (e.g., the presence of crosswalks, traffic signals, and walk signs) and the width of crossings. The average MAPS-Global route for this study included multiple crossing and segment sections. For routes with multiple segments and crossings, each variable was averaged across segments or crossings and used in the scoring and reliability analysis described below. (Cain et al., 2018; Fox et al., 2021; Queralt et al., 2021).

### 2.3. Route selection and data acquisition

Residential neighborhoods have been extensively investigated in PA literature to quantify exposure to the built environment. Therefore, routes beginning at residential address sites were selected as the most suitable locations for assessing microscale environments. (Adlakha et al., 2015; Fox et al., 2021).

One route per participant was defined and observed starting from the participant's home and heading toward the nearest predefined destination along the street network (e.g., stores, services, parks, and schools). Observations were conducted on the shortest walkable routes from home addresses to the nearest cluster of destinations, as determined by study staff. Network distances ranged from 400 m to 725 m along the road network accessible to pedestrians. Alleys, non-motorized, and informal paths adjacent to the street network were not used to create routes. However, these pedestrian facilities were coded within MAPS-Global when they were observed. Routes were identified using Google Earth (Microsoft Windows, 2013, Google Inc.) and Google Street View (Audit date: July 22–September 12, 2020). These routes were considered a sample of walking or bicycling routes likely to be used often by each participant. This approach is more efficient, feasible, and tailored to individual participants than observing all street segments in a neighborhood.

### 2.4. Scoring and data analysis

MAPS-Global scoring largely followed the original MAPS scoring structure, which has been previously described. (Cain et al., 2014; Millstein et al., 2013) To ensure relative weight equality in creating scales, all items (except land use) were dichotomized or trichotomized because different response formats were used. Land-use items were rated on a Likert-type scale with values of 0, 1, 2, 3, 4, or 5+. After rescoring the various items, the subscales were calculated by adding them together. Summing the subscales together, and based on their expected relevance to PA, positive and negative valence scores were created. To determine an overall section score, the negative-valence scores were subtracted from the positive-valence scores. The last step involved calculating the grand score. This was done by subtracting the overall negative- from the overall positive-valence score. The MAPS-Global version includes three new subscales drawn from items in various sections: pedestrian infrastructure, pedestrian design, and bicycle facilities. (Cain et al., 2018; Fox et al., 2021; Queralt et al., 2021) These subscales are expected to positively influence physical activity. Although the items are conceptually related, they are selected from different sections of the tool (see Table 5 for sample items and overall subscale description). Additional information on item recoding and subscale development can be downloaded. MAPS-Global (2021).

### 2.5. Statistical analysis

Inter-rater reliability analysis was performed using the kappa statistic for dichotomous variables and the intraclass correlation coefficient (ICC) for continuous or ordinal variables. ICC was calculated using a one-way random model of the average measures with 95 % confidence intervals. (Cain et al., 2018; Cain et al., 2014; Cicchetti, 2001; Fox et al., 2021; Pikora et al., 2002) In this study, the numerical range and descriptors of Cicchetti's ICC were examined for test–retest reliability. ICC was classified to indicate the test–retest reliability, which was as follows: “excellent” (ICC  $\geq$  0.75), “good” (0.60–0.74), “fair” (0.40–0.59), and “poor” ( $<$ 0.40). (Cain et al., 2018; Cain et al., 2014; Cicchetti, 2001; Fox et al., 2021; Pikora et al., 2002) Items that were seldom observed or had low score variability (mostly zero or “never”) but had an inter-rater agreement of 75 % or higher were considered reliable. (Cain et al., 2018; Pikora et al., 2002).

For each item (original and recoded), the researchers calculated the mean and standard deviation for each rater and inter-rater reliability. Statistical analyses were performed using SPSS version 27 (SPSS Inc., Tokyo, Japan).

## 3. Results

The mean assessment time (standard deviation) was 47 (31) min/route. Observed routes had a mean of 4.6 segments and 4.1 crossings. The sample sizes for the four sections and the evaluation times from the present study were compared to those of previous studies, as presented in Table 1. Despite the small number of routes, the sample sizes for segments ( $n = 88$ ) and crossings ( $n = 78$ ) were comparable to those of the Hong Kong study (segments:  $n = 115$ , crossings:  $n = 73$ ). The average distance length (standard deviation) was 484 (71) meters.

Tables 2–5 illustrate the ICCs among online observers. The tables include the upper and lower limits of the 95 % confidence intervals and descriptive statistics for the main MAPS-Global variables: route section (Table 2), crossing section (Table 3), segment and cul-de-sac section (Table 4), and grand score (Table 5). The descriptive statistics indicate the number of individual items included and the range of potential scores and central tendency for each subscale.

The ICC for the overall grand score was 0.98 (95 % confidence interval: 0.97–0.99). All the grand scores reflected “excellent” reliability (Table 5).

Several positive scales exhibited stronger results than negative scales. In the route section (Table 2), the ICC value of the overall aesthetics and social positive subscale (ICC = 0.90) was significantly greater than that of the overall aesthetics and social negative subscale (ICC = 0.33). Similarly, in the segment section (Table 4), the ICC value of the overall segments' positive subscale (ICC = 0.75) was significantly greater than that of the overall segments' negative subscale (ICC = 0.22).

In the subscales of the route section (Table 2), the pedestrian street subscale displayed “good” results (ICC = 0.67), while the other subscales were “excellent” (ICC = 0.87 to 1.00). The crossing section (Table 3) was also “excellent” (ICC = 0.83 to 0.99), except for the bicycle features' positive subscale (ICC = 0.56). In the segment section (Table 4), the building height and setbacks' positive subscale (ICC = 0.35)—which evaluates building heights and the smallest and largest setbacks—and the building aesthetics and design's positive subscale (ICC = 0.38)—which evaluates the ratio of street windows to the total segment—were in the “fair” and “poor” reliability range, respectively. The cul-de-sac section scores (Table 4) exhibited “excellent” reliability (ICC = 0.95).

## 4. Discussion

In this study, online inter-rater reliability was measured using the MAPS-Global environment observation tool for walking routes from residential addresses to the nearest destination in an urban area of

**Table 1**  
Study locations, sample sizes and assessment times for MAPS-Global online evaluation<sup>a</sup>

Country	City	Sample size				Assessment time, min/routeMean (SD), range
		Routes	Segments (sample size/route)	Crossings (sample size/route)	Cul-de-sacs	
Japan (this study)	Fujisawa	19	88 (4.6)	78 (4.1)	11	47 (31), 5–120
5-city study ( below)		349	1228 (3.5)	799 (2.3)	16	22 (12), 2–78 <sup>b</sup>
Australia	Melbourne	65	208 (3.2)	91 (1.4)	10	21 (7), 9–45
Belgium	Ghent	81	236 (2.9)	156 (1.9)	6	26 (17), 4–78
Brazil	Curitiba	82	319 (3.9)	213 (2.6)	0	11 (4), 2–29
China	Hong Kong SAR	40	115 (2.9)	73 (1.8)	0	– <sup>c</sup>
Spain	Valencia	81	350 (4.3)	266 (3.3)	0	32 (11), 8–62

SD: standard deviation.

a Data for five cities other than this study were taken from “Queralt et al. *Int J Health Geogr.* 2021”.

b This overall value does not include Hong Kong assessment time due to missing data.

c Start and end assessment times were not collected in Hong Kong for online assessments.

**Table 2**  
MAPS-Global Route section reliability.

Variable Description <sup>a</sup>	# items (range of scores)	Rater’s mean (SD)		Inter-rater reliability between online ratings	Sample items and overall subscale description
		Rater A	Rater B		
Residential Mix	4 (0–3)	2.00 (0)	1.89 (0.32)	17/19 (89.5 %) Agreement <sup>a</sup>	Single family, multi-family only and any other mix, apartment over retail only
Shops	8 (0–28)	1.26 (1.94)	1.16 (1.71)	0.96 (0.91, 0.98)	Grocery, convenience store, bakery, drugstore, other retail, shopping mall, strip mall, open-air market
Restaurant-Entertainment	4 (0–20)	1.26 (1.52)	1.21 (1.47)	0.99 (0.98, 0.99)	Fast food, sit-down, café, entertainment
Institutional-Service	3 (0–15)	3.00 (2.71)	2.53 (2.89)	0.97 (0.92–0.98)	Bank, health-related professional, other service
Institutional-Place of Worship	1 (0–5)	0.11 (0.32)	0	17/19 (89.5 %) Agreement <sup>a</sup>	Place of worship
Institutional-School	1 (0–5)	0.16 (0.38)	0.11 (0.32)	0.87 (0.68–0.95)	School land use
Public Recreation Facilities	4 (0–20)	0.47 (0.70)	0.47 (0.70)	0.94 (0.85–0.97)	Public indoor, public outdoor facility, park, trail
Private Recreation Facilities	2 (0–10)	0.05 (0.23)	0.05 (0.23)	1.00	Private indoor, private outdoor facility
Pedestrian Street	1 (0–5)	0.26 (0.73)	0.68 (0.95)	0.67 (0.17, 0.87)	Pedestrian street/zone
Age-restricted bar or nightclub	1 (0–5)	0	0	1.00	Age-restricted bar/nightclub
Liquor or alcohol store	1 (0–5)	0.16 (0.38)	0.16 (0.38)	1.00	Liquor or alcohol store
Positive destinations & land use	28 (0–111)	8.47 (5.28)	7.58 (5.18)	0.98 (0.96–0.99)	Sum of the positive DLU subscales
Negative destinations & land use	2 (0–10)	0.16 (0.37)	0.16 (0.37)	1.00	Sum of the negative DLU subscales
Overall destinations & land use	30	8.32 (5.27)	7.42 (5.19)	0.98 (0.96–0.99)	Positive DLU – Negative DLU
Overall Streetscape Positive	22 (0–29)	2.37 (2.34)	3.00 (2.49)	0.90 (0.76–0.96)	Transit, traffic calming, trash bins, benches, bike racks, bike lockers, bike docking stations, kiosks, hawkers.
Overall Aesthetics and Social Positive	4 (0–4)	1.32 (0.58)	1.21 (0.54)	0.90 (0.76–0.96)	Hardscape, water, softscape, landscaping
Overall Aesthetics and Social Negative	6 (0–6)	1.0 (0.58)	1.58 (0.61)	0.33 (0.68–0.74)	Buildings not maintained, graffiti, litter, dog fouling, physical disorder, highway near
Overall Aesthetics and Social Score	10	0.32 (0.67)	–0.37 (0.81)	0.51 (–0.24–0.81)	Positive Aesthetics/Social – Negative Aesthetics/Social

a Too rare to calculate ICC, reporting percent agreement.

DLU destination and land use.

Japan. The results demonstrated “excellent” agreement (kappa or ICC values of 0.75 or higher) for 80.4 % of the analyzed summary scores, while 6.5 % of items exhibited “good” agreement (ICC = 0.60–0.74). By contrast, only six items (13.0 %) of the scales had ICC values of 0.60 or less (“fair” or “poor” reliability). Thus, this study documented high reliability of online audits among independent observers. A previous study on the reliability of MAPS-Global by two independent observers using on-street and online measures showed equivalent results, with only 23.3 % of the subscales having kappa or ICC values below 0.70. (Queralt et al., 2021) This study’s results add to the evidence that the MAPS-Global online tool can be utilized in diverse countries, including

Japan, provided that online imaging data are available and sufficiently recent.

The average assessment time per route in this study was 47 min, which is longer than the average assessment time of 22 min in the five-city study. One possible reason for this difference is that the number of segments and crossings per route in the present study (4.6 and 4.1, respectively) was higher than that in the 5-city study (3.5 and 2.3, respectively). (Queralt et al., 2021) Further, audit times tend to decrease as auditors gain experience and become more familiar with the process. Therefore, had a larger sample been achieved, the average assessment time might have decreased as the number of audits increased.

**Table 3**  
MAPS-Global Crossing section reliability.

Variable Description <sup>a</sup>	# items (range of scores)	Rater's mean (SD)		Inter-rater reliability between online ratings	Sample items and overall subscale description
		Rater A	Rater B		
Crosswalk Amenities Positive	7 (0–7)	0.92 (0.83)	0.63 (0.53)	0.83 (0.57, 0.93)	Crossing aids, marked crosswalk, high visibility striping, different material, curb extension, raised crosswalk, refuge islands
Curb Quality Positive	3 (0–6)	2.92 (2.35)	2.74 (2.21)	0.99 (0.98, 0.99)	Curb presence, curb ramps lined up, tactile paving
Intersection Control and Signage Positive	7 (0–7)	0.85 (0.73)	0.58 (0.74)	0.90 (0.76, 0.96)	Yield signs, stop signs, traffic signal, traffic circle, pedestrian walk signals, push buttons, countdown signal
Bicycle Features Positive	3 (0–3)	0.05 (0.14)	0.02 (0.08)	0.56 (-0.10, 0.83)	Waiting area, bike lane crossing the crossing, bike signal
Pedestrian Overpass Positive	1 (0–1)	0	0.01 (0.11)	77/78 (98.7 %) Agreement <sup>a</sup>	Crossing on pedestrian overpass, bridge
Road Width Negative	1 (0–2)	0.14 (0.32)	0.11 (0.26)	0.96 (0.91, 0.98)	Distance of crossing leg
Overall Crossings Positive	21 (0–24)	4.74 (3.45)	3.98 (3.03)	0.97 (0.92, 0.98)	Sum of the positive crossing subscales
Overall Crossings Score	22	4.61 (3.30)	3.87 (2.91)	0.97 (0.92, 0.98)	Positive Crossing – Road Width Negative

a Too rare to calculate Kappa, reporting percent agreement.

The subscales of overall segments negative, overall aesthetics and social negative, building height and setbacks positive, and building aesthetics and design positive had lower ICC scores. This is generally consistent with prior studies. Reviewing the literature highlights the potential of these characteristics to introduce subjectivity into the observers' responses. In addition, some of these items are more difficult to see and may be obscured in online imagery. (Cain et al., 2018; Fox et al., 2021; Phillips et al., 2017; Queralt et al., 2021).

Negative subscales have been reported to generally have greater variability and limited ICC strength because of the small number of items. (Fox et al., 2021) There was a trend in each of the sections that microscale features with a limited frequency of occurrence (overall segments negative, overall aesthetics and social negative) had lower ICC values than the more commonly observed features.

When conducting online observations with MAPS-Global, the

quality, currency, and extent of online images must be sufficiently high. (Fox et al., 2021) In the present study, conducted in 2020, all images were available at high resolution and did not hinder the observer's accurate interpretation. In rare cases, details of dead-end locations could not be seen. However, such challenges were usually resolved by observing the features from various angles. Nonetheless, temporary discrepancies in the seasons of the images used in the audit (e.g., shading by trees) may have occurred, so limitations of Street View imagery could have contributed to errors in observations. Although technology has advanced globally, some countries still have limited online images depicting their regions, with some streets (and often whole rural areas) unavailable. (Rzotkiewicz et al., 2018) Over time, many of these gaps will need to be addressed to make online observations possible in all locations. Unfortunately, low-income countries and some countries that have banned or severely restricted image collection programs (such as

**Table 4**  
MAPS-Global Segment and Cul-De-Sac section reliability.

Variable Description <sup>a</sup>	# items (range of scores)	Rater's mean (SD)		Inter-rater reliability between online ratings	Sample items and overall subscale description
		Rater A	Rater B		
Building Height and Setbacks Positive	4 (0–10)	3.74 (0.75)	4.34 (0.77)	0.35 (-0.65, 0.74)	Building height, smallest and largest setback
Building Height: Road Width and Setback Ratio Positive	5 (0–3)	1.77 (0.68)	1.78 (0.75)	0.89 (0.72, 0.95)	Building height, setback and road width
Buffers Positive	2 (0–5)	1.36 (1.33)	1.64 (1.16)	0.73 (0.32, 0.89)	Parking along street, buffer
Bicycle Infrastructure Positive	3 (0–5)	0.51 (1.28)	0.58 (1.32)	0.99 (0.99, 0.99)	Bike lane presence, quality, signage
Shade Positive	3 (0–6)	0.88 (0.94)	0.59 (0.65)	0.70 (0.25, 0.88)	Number of trees, sidewalk coverage, shade
SidewalkQualities Positive	2 (0–6)	3.19 (2.00)	3.64 (1.53)	0.84 (0.59, 0.93)	Sidewalk presence and width
Pedestrian Infrastructure Positive	5 (0–5)	1.28 (0.42)	1.17 (0.47)	0.86 (0.65, 0.94)	Mid-segment crossing, pedestrian bridge, covered place to walk, street lights
Building Aesthetics and Design Positive	1 (0–2)	0.44 (0.22)	0.26 (0.30)	0.38 (-0.56, 0.76)	Street windows
Informal Path or Shortcut Positive	1 (0–1)	0	0.10 (0.30)	79/88 (89.8 %) Agreement <sup>a</sup>	Informal path connecting to something else
Hawkers/Shops Positive	1 (0–2)	0	0	1.00	Hawkers/shops on sidewalk/pedestrian zone
Overall Segments Positive	27 (0–45)	13.17 (3.75)	14.16 (3.13)	0.75 (0.38, 0.90)	Sum of the positive segment subscales
Overall Segments Negative	7 (0–13)	0.63 (0.49)	1.09 (0.49)	0.22 (-0.97, 0.70)	Non-continuous sidewalk, trip hazards, obstructions, cars blocking walkway, slope, gates, driveways
Overall Segment Score	34	12.54 (4.15)	13.07 (3.40)	0.79 (0.46, 0.91)	Positive Segment – Negative Segment
Overall Cul-de-sac/Dead-end Score	6 (0–4)	1.09 (0.94)	1.09 (1.04)	0.95 (0.83, 0.98)	closeness to participant's home, total amenities, visibility of cul-de-sac area from participant's home

a Too rare to calculate Kappa, reporting percent agreement.

**Table 5**  
MAPS-Global Grand Scores and conceptual scale reliability.

Variable Description <sup>a</sup>	# items (range of scores)	Rater's mean (SD) <sup>a</sup>		Inter-rater reliability between online ratings	Sample items and overall subscale description
		Rater A	Rater B		
Overall Positive	102 (0–210)	31.04 (11.12)	29.93 (10.42)	0.99 (0.97, 0.99)	Positive DLU, positive streetscape, positive aesthetics/social, positive segment (mean of all segments), positive crossing (mean of all segments).
Overall Negative	16 (0–22)	2.40 (1.00)	2.93 (0.93)	0.80 (0.50, 0.92)	Negative DLU, negative aesthetics/social, negative segment (mean of all segments), negative crossing (mean of all crossings).
Overall Grand Score	118	28.64 (11.15)	26.99 (10.60)	0.98 (0.97, 0.99)	Overall Positive – Overall Negative
Pedestrian Infrastructure	13 (0–27)	5.75 (3.93)	6.22 (3.32)	0.94 (0.84, 0.97)	Trail, pedestrian zone, sidewalk presence/width, buffer, shortcut, mid-segment crossing, pedestrian bridge, air-conditioned place to walk, low lights, overpass, crosswalk, refuge island
Pedestrian Design	13 (0–22)	4.79 (2.26)	4.93 (1.34)	0.86 (0.65, 0.94)	Open-air market, trash cans, benches, kiosks, hawkers and shops, setback, visibility, pedestrian walk signals, push buttons, countdown signals, ramps, crossing aids
Bicycle Facilities	9 (0–11)	0.56 (1.27)	0.52 (1.15)	0.98 (0.97, 0.99)	Bike racks, docking stations, lockers, bike lane, bike lane quality, signs, bike signal, bike box, bike lane perpendicular to the crossing

DLU destination and land use.

Google Street View) may continue to experience the knowledge gap. (Queralt et al., 2021).

In Japan, in response to a series of accidents in which children were killed or injured, including a traffic accident in which a car drove into a line of children on their way to and from school, the government conducted an inspection of the school routes of approximately 19,000 schools nationwide in 2021. Consequently, 76,404 dangerous spots were found, such as “heavy traffic of large vehicles” and “intersections with poor visibility,” and the government promoted measures such as the construction of sidewalks and speed restrictions for cars. (Cabinet Office, Government of Japan, 2022) Thus, observation of the built environment is important not only for physical activity and chronic diseases, but also from the perspective of traffic safety. Therefore, developing an objective tool for online evaluation, such as MAPS-Global that has confirmed reliability, is essential.

The present study's strength lies in the way it documented evidence of the inter-rater reliability in a country where MAPS-Global had not been used before. The study described distinct constructs to allow for comparisons across countries, clear scoring guidelines and training procedures, and conceptually meaningful summary variables that can be used for analysis. However, in-person observations on-street were not conducted, and alternate-form reliability could not be confirmed. In addition, the number of routes was small (19 routes in the two groups), and the addresses within the same group were in proximity, which prevented the assessment of the local environment's full diversity. In other words, reliability estimates may have been underestimated or overestimated for elements that were rarely observed, regardless of their positive or negative impact on physical activity. Given this convenience sampling, a study with a larger number of participants and more systematic sampling from diverse areas of a city is necessary to render the data representative of the target area of this study. Though the small sample of routes in one Japanese city demonstrates the Japanese version of MAPS-Global is feasible, we cannot determine whether the results would generalize to other Japanese cities. We recommend broader use of this instrument throughout Japan to examine generalizability and generate data relevant to important scientific and societal questions.

The limitations of MAPS-Global include a large number of items and the need for observer training and continuous monitoring, which increases the cost of data collection and the researcher's burden. In this study, the built environment in Japan was characterized by a higher number of segments per route than in other countries, (Queralt et al., 2021) which may have contributed to the extended observation times. If the data collected by MAPS-Global are accumulated in the future and if conducting evaluation using technologies such as artificial intelligence and deep learning becomes possible, these issues can be overcome. Several methods of auditing using deep learning have been developed and applied to microscale observations, (Koo et al., 2021; Lu, 2018; Nagata et al., 2020) and combining these methods with MAPS-Global for

further study may be possible.

## 5. Conclusions

This study examined the inter-rater reliability using online observations with MAPS-Global in an urban area of Japan. The results demonstrated high reliability for the grand and composite subscale measures. However, caution should be exercised when interpreting subscale scores for less frequently observed negative factors and aesthetic/social characteristics. These results are comparable to those of previous studies and support the use of online observations using MAPS-Global in urban areas of Japan. Online observations of the built environment are expected to be useful for studying not only physical activity but also traffic safety.

## CRedit authorship contribution statement

**Yoshinobu Saito:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Yuko Oguma:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing, Supervision, Funding acquisition. **Shigeru Inoue:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Raoul Breugelmans:** Methodology, Writing – review & editing. **Hiroyuki Kikuchi:** Writing – review & editing. **Koichiro Oka:** Writing – review & editing. **Shinpei Okada:** Writing – review & editing. **Noriko Takeda:** Writing – review & editing. **Kelli L. Cain:** Resources, Writing – review & editing, Supervision. **James F. Sallis:** Resources, Writing – review & editing, Supervision. All the authors have read and agreed to the published version of the manuscript.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The authors do not have permission to share data.

## Acknowledgments

We are grateful to the residents who participated in this study. The authors would like to acknowledge Dr. Masamitsu Kamada and Dr. Haruka Murakami who provided feedback on the development of the Japanese version of MAPS-Global tool, and Naoyuki Ozawa, Kanako Kikuchi, and Natsue Doihara for their assistance with the data

collection. This study was conducted with the support of The Japanese Association of Exercise Epidemiology's Project Research for organizing the research team and with the assistance of experts. This study was partially supported by JSPS KAKENHI Grant Numbers: JP17K01795, JP17H06151, JP18K11055, and JP19H03910.

## References

- Adlakha, D., Hipp, A.J., Marx, C., Yang, L., Tabak, R., Dodson, E.A., Brownson, R.C., 2015. Home and workplace built environment supports for physical activity. *Am J Prev Med* 48, 104–107. <https://doi.org/10.1016/j.amepre.2014.08.023>.
- Adlakha, D., Hipp, J.A., Brownson, R.C., 2016. Adaptation and Evaluation of the Neighborhood Environment Walkability Scale in India (NEWS-India). *Int J Environ Res Public Health* 13, 401. <https://doi.org/10.3390/ijerph13040401>.
- Bader, M.D., Mooney, S.J., Lee, Y.J., Sheehan, D., Neckerman, K.M., Rundle, A.G., Teitler, J.O., 2015. Development and deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neighborhood conditions. *Health Place* 31, 163–172. <https://doi.org/10.1016/j.healthplace.2014.10.012>.
- Brownson, R.C., Hoehner, C.M., Day, K., Forsyth, A., Sallis, J.F., 2009. Measuring the built environment for physical activity: state of the science. *Am J Prev Med* 36 (S99–123), e12. <https://doi.org/10.1016/j.amepre.2009.01.005>.
- Cabinet Office, Government of Japan. White paper on traffic safety in Japan 2022. [https://www8.cao.go.jp/koutu/taisaku/r04kou\\_haku/index\\_zenbun\\_pdf.html](https://www8.cao.go.jp/koutu/taisaku/r04kou_haku/index_zenbun_pdf.html) (Accessed Sept 20 2022). in Japanese.
- Cain, K.L., Millstein, R.A., Sallis, J.F., Conway, T.L., Gavand, K.A., Frank, L.D., Saelens, B.E., Geremia, C.M., Chapman, J., et al., 2014. Contribution of streetscape audits to explanation of physical activity in four age groups based on the Microscale Audit of Pedestrian Streetscapes (MAPS). *Soc Sci Med* 116, 82–92. <https://doi.org/10.1016/j.socscimed.2014.06.042>.
- Cain, K.L., Geremia, C.M., Conway, T.L., Frank, L.D., Chapman, J.E., Fox, E.H., Timperio, A., Veitch, J., Van Dyck, D., et al., 2018. Development and reliability of a streetscape observation instrument for international use: MAPS-global. *Int J Behav Nutr Phys Act* 15, 19. <https://doi.org/10.1186/s12966-018-0650-z>.
- Cerin, E., Chan, K.W., Macfarlane, D.J., Lee, K.Y., Lai, P.C., 2011. Objective assessment of walking environments in ultra-dense cities: development and reliability of the Environment in Asia Scan Tool-Hong Kong version (EAST-HK). *Health Place* 17, 937–945. <https://doi.org/10.1016/j.healthplace.2011.04.005>.
- Cicchetti, D.V., 2001. The precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. *J Clin Exp Neuropsychol* 23, 695–700. <https://doi.org/10.1076/j.jcen.23.5.695.1249>.
- Clifton, K.J., Livi Smith, A., Rodriguez, D., 2007. The development and testing of an audit for the pedestrian environment. *Landsc Urban Plan* 80, 95–110. <https://doi.org/10.1016/j.landurbplan.2006.06.008>.
- Dunstan, F., Weaver, N., Araya, R., Bell, T., Lannon, S., Lewis, G., Patterson, J., Thomas, H., Jones, P., et al., 2005. An observation tool to assist with the assessment of urban residential environments. *J Environ Psychol* 25, 293–305. <https://doi.org/10.1016/j.jenvp.2005.07.004>.
- Eyler, A.A., Blanck, H.M., Gittelsohn, J., Karpyn, A., McKenzie, T.L., Partington, S., Slater, S.J., Winters, M., 2015. Physical activity and food environment assessments: implications for practice. *Am J Prev Med* 48, 639–645. <https://doi.org/10.1016/j.amepre.2014.10.008>.
- Fox, E.H., Chapman, J.E., Moland, A.M., Alfonsin, N.E., Frank, L.D., Sallis, J.F., Conway, T.L., Cain, K.L., Geremia, C., et al., 2021. International evaluation of the Microscale Audit of Pedestrian Streetscapes (MAPS) Global instrument: comparative assessment between local and remote online observers. *Int J Behav Nutr Phys Act* 18, 84. <https://doi.org/10.1186/s12966-021-01146-3>.
- Frank, L.D., Sallis, J.F., Saelens, B.E., Leary, L., Cain, K., Conway, T.L., Hess, P.M., 2010. The development of a walkability index: application to the Neighborhood Quality of Life Study. *Br J Sports Med* 44, 924–933. <https://doi.org/10.1136/bjsm.2009.058701>.
- Griew, P., Hillsdon, M., Foster, C., Coombes, E., Jones, A., Wilkinson, P., 2013. Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. *Int J Behav Nutr Phys Act* 10, 103. <https://doi.org/10.1186/1479-5868-10-103>.
- Jones, N.R., Jones, A., van Sluijs, E.M., Panter, J., Harrison, F., Griffin, S.J., 2010. School environments and physical activity: The development and testing of an audit tool. *Health Place* 16, 776–783. <https://doi.org/10.1016/j.healthplace.2010.04.002>.
- Katzmarzyk, P.T., Barreira, T.V., Broyles, S.T., Champagne, C.M., Chaput, J.P., Fogelholm, M., Hu, G., Johnson, W.D., Kuriyan, R., et al., 2013. The International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE): design and methods. *BMC Public Health* 13, 900. <https://doi.org/10.1186/1471-2458-13-900>.
- Koo, B.W., Guhathakurta, S., Botchwey, N., 2021. Development and validation of automated microscale walkability audit method. *Health Place* 73, 102733. <https://doi.org/10.1016/j.healthplace.2021.102733>.
- Lu, Y., 2018. The Association of Urban Greenness and Walking Behavior: Using Google Street View and Deep Learning Techniques to Estimate Residents' Exposure to Urban Greenness. *Int J Environ Res Public Health* 15. <https://doi.org/10.3390/ijerph15081576>.
- MAPS-Global webpage. [https://drjimalls.org/measure\\_maps.html#MAPSGLOBAL](https://drjimalls.org/measure_maps.html#MAPSGLOBAL) (Cited Nov 25 2021).
- Millstein, R.A., Cain, K.L., Sallis, J.F., Conway, T.L., Geremia, C., Frank, L.D., Chapman, J., Van Dyck, D., Dipzinski, L.R., et al., 2013. Development, scoring, and reliability of the Microscale Audit of Pedestrian Streetscapes (MAPS). *BMC Public Health* 13, 403. <https://doi.org/10.1186/1471-2458-13-403>.
- Ministry of Land, Infrastructure, Transport and Tourism. Traffic accident situation. Portal site for traffic safety measures on community roads. <https://www.mlit.go.jp/road/road/traffic/sesaku/jiko.html> (Accessed Sep 11 2022). in Japanese.
- Molina-Garcia, J., Campos, S., Garcia-Masso, X., Herrador-Colmenero, M., Galvez-Fernandez, P., Molina-Soberanes, D., Queralt, A., Chillón, P., 2020. Different neighborhood walkability indexes for active commuting to school are necessary for urban and rural children and adolescents. *Int J Behav Nutr Phys Act* 17, 124. <https://doi.org/10.1186/s12966-020-01028-0>.
- Nagata, S., Nakaya, T., Hanibuchi, T., Amagasa, S., Kikuchi, H., Inoue, S., 2020. Objective scoring of streetscape walkability related to leisure walking: Statistical modeling approach with semantic segmentation of Google Street View images. *Health Place* 66, 102428. <https://doi.org/10.1016/j.healthplace.2020.102428>.
- International Physical Activity and Environment Network. <http://www.ipenproject.org>. (Cited Oct 1 2022).
- Oyeyemi, A.L., Sallis, J.F., Oyeyemi, A.Y., Amin, M.M., De Bourdeaudhuij, I., Deforche, B., 2013. Adaptation, test-retest reliability, and construct validity of the Physical Activity Neighborhood Environment Scale in Nigeria (PANES-N). *J Phys Act Health* 10, 1079–1090. <https://doi.org/10.1123/jpah.10.8.1079>.
- Phillips, C.B., Engelberg, J.K., Geremia, C.M., Zhu, W., Kurka, J.M., Cain, K.L., Sallis, J.F., Conway, T.L., Adams, M.A., 2017. Online versus in-person comparison of Microscale Audit of Pedestrian Streetscapes (MAPS) assessments: reliability of alternate methods. *Int J Health Geogr* 16, 27. <https://doi.org/10.1186/s12942-017-0101-0>.
- Pikora, T.J., Bull, F.C., Jamrozik, K., Knuiaman, M., Giles-Corti, B., Donovan, R.J., 2002. Developing a reliable audit instrument to measure the physical environment for physical activity. *Am J Prev Med* 23, 187–194. [https://doi.org/10.1016/s0749-3797\(02\)00498-1](https://doi.org/10.1016/s0749-3797(02)00498-1).
- Queralt, A., Molina-Garcia, J., Terron-Perez, M., Cerin, E., Barnett, A., Timperio, A., Veitch, J., Reis, R., Silva, A.A.P., et al., 2021. Reliability of streetscape audits comparing on-street and online observations: MAPS-Global in 5 countries. *Int J Health Geogr* 20, 6. <https://doi.org/10.1186/s12942-021-00261-5>.
- Rzotkiewicz, A., Pearson, A.L., Dougherty, B.V., Shortridge, A., Wilson, N., 2018. Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health Place* 52, 240–326. <https://doi.org/10.1016/j.healthplace.2018.07.001>.
- Saito, Y., Oguma, Y., Tajima, T., Kato, R., Kibayashi, Y., Miyachi, M., Takebayashi, T., 2018. Association of high individual-level of social capital with increased physical activity among community-dwelling elderly men and women: a cross-sectional study. *Jpn J Phys Fitness Sports Med* 67 (177–85). <https://doi.org/10.7600/jspfsm.67.177> in Japanese.
- Saito, Y., Tanaka, A., Tajima, T., Ito, T., Aihara, Y., Nakano, K., Kamada, M., Inoue, S., Miyachi, M., et al., 2021. A community-wide intervention to promote physical activity: A five-year quasi-experimental study. *Prev Med* 150, 106708. <https://doi.org/10.1016/j.ypmed.2021.106708>.
- Sallis, J.F., Certero, R.B., Ascher, W., Henderson, K.A., Kraft, M.K., Kerr, J., 2006. An ecological approach to creating active living communities. *Annu Rev Public Health* 27, 297–322. <https://doi.org/10.1146/annurev.publhealth.27.021405.102100>.
- Sallis, J.F., Slymen, D.J., Conway, T.L., Frank, L.D., Saelens, B.E., Cain, K., Chapman, J.E., 2011. Income disparities in perceived neighborhood built and social environment attributes. *Health Place* 17, 1274–1283. <https://doi.org/10.1016/j.healthplace.2011.02.006>.
- Sallis, J.F., Cain, K.L., Conway, T.L., Gavand, K.A., Millstein, R.A., Geremia, C.M., Frank, L.D., Saelens, B.E., Glanz, K., et al., 2015. Is Your Neighborhood Designed to Support Physical Activity? A Brief Streetscape Audit Tool. *Prev Chronic Dis* 12, E141. <https://doi.org/10.5888/pcd12.150098>.
- Spittaels, H., Verloigne, M., Gidlow, C., Gloanec, J., Titze, S., Foster, C., Oppert, J.M., Rutter, H., Oja, P., et al., 2010. Measuring physical activity-related environmental factors: reliability and predictive validity of the European environmental questionnaire ALPHA. *Int J Behav Nutr Phys Act* 7, 48. <https://doi.org/10.1186/1479-5868-7-48>.
- Steinmetz-Wood, M., Velauthapillai, K., O'Brien, G., Ross, N.A., 2019. Assessing the micro-scale environment using Google Street View: the Virtual Systematic Tool for Evaluating Pedestrian Streetscapes (Virtual-STEPS). *BMC Public Health* 19, 1246. <https://doi.org/10.1186/s12889-019-7460-3>.
- Taylor, B.T., Fernando, P., Bauman, A.E., Williamson, A., Craig, J.C., Redman, S., 2011. Measuring the quality of public open space using Google Earth. *Am J Prev Med* 40, 105–112. <https://doi.org/10.1016/j.amepre.2010.10.024>.
- Vanwolleghem, G., Ghekiere, A., Cardon, G., De Bourdeaudhuij, I., D'Haese, S., Geremia, C.M., Lenoir, M., Sallis, J.F., Verhoeven, H., et al., 2016. Using an audit tool (MAPS Global) to assess the characteristics of the physical environment related to walking for transport in youth: reliability of Belgian data. *Int J Health Geogr* 15, 41. <https://doi.org/10.1186/s12942-016-0069-1>.
- Wilson, J.S., Kelly, C.M., Schootman, M., Baker, E.A., Banerjee, A., Clennin, M., Miller, D.K., 2012. Assessing the built environment using omnidirectional imagery. *Am J Prev Med* 42, 193–199. <https://doi.org/10.1016/j.amepre.2011.09.029>.