# Development and Deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to Measure Health-Related Neighborhood Conditions

**Michael D. M. Bader**,
Department of Sociology and Center on Health, Risk and Society, American University

**Stephen J. Mooney**,
Department of Epidemiology, Mailman School of Public Health, Columbia University

**Yeon Jin Lee**,
Department of Sociology and Population Studies Center, University of Pennsylvania

**Daniel Sheehan**,
Department of Epidemiology, Mailman School of Public Health, Columbia University

**Kathryn M. Neckerman**,
Columbia Population Research Center, Columbia University

**Andrew G. Rundle**, and
Department of Epidemiology, Mailman School of Public Health, Columbia University

**Julien O. Teitler**
School of Social Work, Columbia University

## Abstract

Public health research has shown that neighborhood conditions are associated with health behaviors and outcomes. Systematic neighborhood audits have helped researchers measure neighborhood conditions that they deem theoretically relevant but not available in existing administrative data. Systematic audits, however, are expensive to conduct and rarely comparable across geographic regions. We describe the development of an online application, the Computer Assisted Neighborhood Visual Assessment System (CANVAS), that uses Google Street View to conduct virtual audits of neighborhood environments. We use this system to assess the inter-rater reliability of 187 items related to walkability and physical disorder on a national sample of 150 street segments in the United States. We find that many items are reliably measured across auditors using CANVAS and that agreement between auditors appears to be uncorrelated with neighborhood demographic characteristics. Based on our results we conclude that Google Street

**Corresponding Author:** Michael D. M. Bader, Department of Sociology, American University, Battelle-Tompkins T-15, 4400 Massachusetts Ave., NW, Washington, D.C. 20016; bader@american.edu; 202-885-2433 (telephone); 202-885-2477 (fax).

View and CANVAS offer opportunities to develop greater comparability across neighborhood audit studies.

## Keywords

Neighborhood; walkability; disorder; measurement error; neighborhood audit; systematic social observation; Google Street View

As research investigating the influence of neighborhood conditions on individual outcomes proliferates, the need for methods to efficiently measure theoretically relevant aspects of neighborhood environments has increased (Cummins, Macintyre, Davidson, & Ellaway, 2005; Brownson, Hoehner, Day, Forsyth, & Sallis, 2009; Saelens & Glanz, 2009). Systematic audits of neighborhood environments represent one of the most adaptable tools available to measure neighborhood features reliably (Pikora et al., 2002; Reiss, 1971; Sampson & Raudenbush, 1999). The adaptability of neighborhood audits (also called systematic social observations) comes from the fact that researchers define pertinent aspects to measure based on theories about how neighborhoods affect health.

Researchers' ability to define their own measures provides a unique advantage of neighborhood audits compared to the georeferencing of existing administrative data that are collected for purposes other than research. Administrative sources often do not contain measures of neighborhood features thought to affect health or the measures do not correspond well to neighborhood features thought to affect health. Using neighborhood audits, researchers create an instrument containing items they deem relevant to observe, train auditors how to observe those items, and then record observations of those items in neighborhoods across their study area. The systematic nature of the data collection allows researchers to use standard statistical techniques to construct scales and model potential influences of neighborhood attributes on individual outcomes (Mujahid, Diez Roux, Morenoff, & Raghunathan, 2007; Raudenbush & Sampson, 1999). Systematic neighborhood audits have been used in several fields including criminology (Perkins, Florin, Rich, Wandersman, & Chavis, 1990; Reiss, 1971; Taylor, Gottfredson, & Brower, 1984), sociology (Raudenbush & Sampson, 1999; Sampson & Raudenbush, 1999), urban design and planning (Ewing, Handy, Brownson, Clemente, & Winston, 2006), and public health (Clifton, Livi Smith, & Rodriguez, 2007; Pikora et al., 2002).

Most neighborhood audits have been conducted with raters visiting streets in person, an expensive undertaking that limits the use of neighborhood audits and hinders the generalizability of studies beyond relatively small geographic areas. Consequently, most systematic neighborhood audits have been limited to areas no larger than a single city or metropolitan area, frequently those in which the researchers reside. An early adaptation to overcome the problem of sending raters to the field used video cameras mounted to trucks that tape-recorded while being driven down Chicago streets (Sampson & Raudenbush, 1999; Raudenbush & Sampson, 1999; Earls, Raudenbush, Reiss, & Sampson, 1995). Auditors then coded neighborhood conditions by viewing the videos. Video recording improved reliability because auditors coded streets in a central facility and reduced costs by eliminating travel time of auditors to streets (Carter, Dougherty, & Grigorian, 1995). At the same time, video

recording presented logistical obstacles including equipment malfunction and correctly identifying segments to ensure ratings were assigned to the appropriate location (Carter et al., 1995).

One of the most promising advances in neighborhood research has been the validation of neighborhood audits using Google Street View, which provides still imagery free of charge for streets in much of the United States. Recent studies have shown that "virtual audits" done using Google Street View have acceptable levels of concurrent validity and inter-rater reliability while eliminating geographic constraints imposed by travel costs and field logistics (Anguelov et al., 2010; Badland, Opit, Witten, Kearns, & Mavoa, 2010; Clarke, Ailshire, Melendez, Bader, & Morenoff, 2010; Curtis, Duval-Diop, & Novak, 2010; Odgers, Caspi, Bates, Sampson, & Moffitt, 2012; Rundle, Bader, Richards, Neckerman, & Teitler, 2011; Vargo, Stone, & Glanz, 2011; Wilson et al., 2012).

The promise of systematic neighborhood audits rests not only with ensuring valid and reliable data *within* a study, but also ensuring valid and reliable data that can be compared *across* studies. To date, the inconsistency of measurement across sites has meant that neighborhood audits have not lived up to their promise (Brownson et al., 2009). Google Street View audits are uniquely capable of overcoming inconsistency because of Google Street View's large geographic reach, consistent method of image acquisition, and existing integration with geographic information systems. The tool described in this paper, the Computer Assisted Neighborhood Visual Assessment System, or CANVAS, is an online application with an efficient and user-friendly interface built around Google Street View that allows researchers to deploy audits and manage data collection using Street View. The primary contribution of CANVAS is that it supports the development of large-scale, generalizable studies of neighborhood effects by improving the efficiency and quality of data collection using the secondary data offered by Google Street View. In this paper we describe the development and features of the CANVAS software and report the results of a virtual field test that measured walkability and physical disorder on a nationwide sample of street segments. We assessed rating times, inter-rater reliability, and potential measurement bias on 187 items.

## METHODS

### Development of the Computer Assisted Neighborhood Visual Assessment System

Here we describe the Computer-Assisted Neighborhood Visual Assessment System (CANVAS) web-based software application, including a summary of the technical specifications and design features for study managers. Three priorities guided the design and implementation of CANVAS. The first was to reduce measurement error due to controllable factors such as auditors rating the wrong street, misinterpretation of the question wording, or inconsistent application of rating instructions. The second was to create a system that allows study managers with limited technical knowledge to deploy and oversee data collection. The last was to develop a standard set of data collection protocols and items covering a variety of domains for use by researchers.

We developed an initial prototype for CANVAS using a combination of Google Forms, a simple computer gateway interface (CGI), and the Google Maps application programming interface (API). High school interns used this prototype to rate streets in New York City during the summer of 2009. This prototype served as a proof of concept, but revealed the necessity of an integrated web framework for efficient and reliable data collection. The CANVAS application was built using version 1.4 of the Django web framework (Django Software Foundation, 2012) with a MySQL version 5.1 database (MySQL AB, 2005). Compared with the prototype, CANVAS improved the integration of the interfaces for auditors and study manager and allowed programming of item skip patterns, grouping of related questions into modules, and inclusion of on-screen help for auditors. The authors and four interns – three undergraduates and one graduate student – tested and critiqued the usability of the new version of CANVAS both for data collection and for study administration. Revisions based on these critiques were incorporated into CANVAS and the system was used in the data collection reported below.

### Items Rated

Items included in CANVAS were adapted from several existing audit instruments. We began by incorporating the full inventory of items in three existing audits, the Irvine-Minnesota Inventory (Day, Boarnet, Alfonzo, & Forsyth, 2006), the Pedestrian Environment Data Scan (PEDS, Clifton et al., 2007), and the Maryland Inventory of Urban Design Qualities (MIUDQ, Ewing, Clement, Handy, Brownson, & Winston, 2005; Ewing et al., 2006). We also incorporated select items related to physical disorder from two other inventories, the systematic social observation of the Project on Human Development in Chicago Neighborhoods (Carter et al., 1995; Sampson & Raudenbush, 1999) and the New York Housing and Vacancy Survey (U.S. Bureau of the Census, 2011). We edited the list to reduce redundancy as well as items not measureable using imagery (e.g. noise).

In addition to the items measuring features of the neighborhood environment, we developed a series of items to evaluate image quality and the prevalence of obstructions blocking views of the sidewalk. The goal of this module was to assess aspects of street imagery that might affect reliability. Unlike neighborhood audit measures, this module had no in-person analog and so had not previously been pilot tested. This module included the zoom level at which the auditor first perceived pixelation, the camera technology used ('bright' (high-resolution) or 'dark' (low-resolution)) (Anguelov et al., 2010), and the legibility of street signs on the segment (clear/blurry/unreadable), as well as the total length of the segment in "steps" (mouse clicks needed to advance along the entire segment) and the number of those steps for which the view was obstructed.

The resulting inventory contained 187 items. To increase the efficiency of rating, we included skip patterns so that auditors would not be required to spend time entering redundant results. For example, auditors were asked if there were any commercial uses on the block; if they answered "no," then they were not asked to rate whether specific kinds of commercial land use were present. This reduced auditor burden and increased rating speed.

**Study deployment interface—**The CANVAS study deployment interface allows researchers to designate the street locations to be audited, designate the audit items to be rated, and assign street segments to auditors. Locations can be identified in either 'manual-confirm' mode or 'auto-select' mode. In manual-confirm mode, study directors upload a CSV file listing pairs of longitude and latitude values representing the start and end points for each segment, and CANVAS then interacts with Google's API to display the Street View imagery at start and end point and a small map showing the connection between them. Study directors may modify the start and end points using the Street View interface before choosing to confirm or reject the segment as selected for the study. In auto-select mode, study directors upload a CSV file listing locations near which an auditable street is sought. CANVAS will match to the nearest viewable street segment by issuing calls to Google's API. If Street View is not available with 50 meters, CANVAS will search up to 5 times in random directions for a segment on which Street View is available, up to 125 meters away. The auto-selector process presents a map indicating segments selected and locations where no segment could be found. At present, auto-selector settings are not configurable, but future CANVAS development may allow study directors to specify offset distance or highlight points that fall at a distance from the originally supplied point. Using auto-select mode, study directors may choose a random array of points, a systematic grid of points, or a table of points of interest. In principle, street segments could also be specified via a point-and-click map interface, though this was not implemented in CANVAS's initial deployment due to time constraints.

After selecting segments, researchers may choose items to rate by selecting from audit protocols already adapted for CANVAS (i.e., Irvine-Minnesota Inventory, PEDS, MIUDQ) or by uploading and editing new items and corresponding rating protocols. Using this interface, researchers are able to tailor rating locations and items to their specific research needs.

**Auditor interface—**CANVAS is designed for use on a computer connected to two monitors. After logging into the CANVAS application, the auditor is presented with a list of assigned street segments to rate. When the auditor clicks on an assignment, the browser opens a form with the items the auditor should rate. Meanwhile, the application also opens a new browser window that locates the auditor on the correct street in Google Street View; a green marker indicates the beginning of the segment to be rated, and a red marker indicates the end. In addition, an inset frame provides an aerial view of the street segment in Google Maps. Auditors can re-arrange the windows between the two screens as they prefer. If dual monitors are unavailable, the windows can be organized on a single monitor. Since using a single monitor reduces image size or requires shuffling between browser windows, a dual-monitor setup is advised. The auditors view the assigned street segment via Street View on one screen and enter the item scores into the form on the second screen. CANVAS provides links to item-specific help boxes with rating instructions, definitions, and, in some cases, images to help auditors classify neighborhood features. After finishing the assigned street segment, the auditor is then directed to the next assignment and the browser loads a new form for the instrument and opens a new Street View window with the next street segment to be rated.

The design of the auditor interface helps reduce coding errors in three ways. First, "placing" the auditor on the correct street and marking the start and end points minimizes the likelihood of rating the wrong street. This was a relatively common experience in initial tests, particularly in settings where it was unclear where segments started and ended. Second, by displaying the street and instrument form simultaneously, CANVAS allows the auditor to observe the street while referring to the coding form, reducing keying errors that might result from switching between browser windows. Finally, the form framework used by Django requires all items to receive a rating before the auditor can move on in the instrument or to the next street segment. This reduces item nonresponse.

**Analytics Interface—**CANVAS also includes an analytics interface that provides data collection reports to the study manager in real time. Reports include data on: each auditor's progress through the assigned street segments, inter-rater kappa ($k$) statistics, and the minimum, maximum, and mean rating times for each module in the study. These reports help study managers monitor data collection, identify problems associated with particular items or auditors, and adjust data collection as necessary to ensure completion. These data can be downloaded in text files and imported into statistical software packages for analysis.

### National Reliability Study

Using the CANVAS application, we implemented a study in the summer of 2011 to assess the reliability of the measures in a sample of street segments in US metropolitan areas.

**Sample—**Street segments were drawn from a sample of segments falling in U.S. metropolitan areas and, to ensure a large number of central city locations, a second sample of segments falling in U.S. central cities. To obtain the metropolitan sample, we randomly sampled, without replacement, 300 Census tracts that fell within U.S. metropolitan areas based on the Office of Management and Budget's 2000 definitions of metropolitan statistical areas (these are the definitions used by the Census Bureau). The central city sample of 150 tracts was drawn without replacement from tracts in which the Census Designated Place Federal Information Processing Standards code (FIPS code) was defined as a central city in the metropolitan statistical area definitions, and was drawn independently of the metropolitan sample. The tracts were drawn from the Neighborhood Change Database, a national dataset of Census tracts from 1970 to 2000 developed by the Urban Institute and distributed by Geolytics; data for 2000 were used. A street segment falling within each Census tract was randomly drawn using a geographic information system (GIS) with the Census TIGER/Line street file; one end of the segment was selected to be the start based on the direction of the street feature vector in the GIS. The segments drawn from the GIS were then loaded into CANVAS using 'manual-confirm' mode as described above. A segment was confirmed as auditable if it contained Street View imagery at both ends of the segment and contained at least one image between the two end points to characterize the street. The analytic samples were drawn from the randomly ordered metropolitan and central city samples. To obtain the 112 segments, we tested the first185 segments in the Google Maps API. We then performed a similar procedure for the central city sample, with the first 56 segments sampled to obtain 38 "street viewable" segments. Longer and non-arterial streets were less likely to have Street View imagery as well as tracts with lower street intersection

density. Holes in the urban grid could also result from policy decisions by Google to, for example, not provide imagery of women's shelters (Mills, 2007). For each selected segment, a trained auditor (the second author) used a Street View window embedded in the CANVAS interface to determine the latitude and longitude values of the points in which Street View imagery showed intersecting streets diverging from the segment of interest; these values were then set as the beginning and end points of the selected segments. Auditors were asked to rate the right side of the street as they "walked" from beginning to end of the segment.

A total of five auditors conducted the ratings over the course of the summer and fall of 2011. Two undergraduate auditors dropped out in the fall and did not complete all 150 sampled segments; our analyses are based only on the three auditors who completed all segments. We conducted ten rounds of pre-testing in which auditors (N=3–5) rated the items on up to ten blocks and then met to discuss discrepancies, which resulted in a total of 291 street faces being tested before conducting the reliability test. These street faces were not used in the reliability study. The team discussed the correct ratings with the project manager and, if necessary, the project manager updated the help text to provide definitions or images to guide coding. We corrected problems with the auditor interface after each round as well.

## Analytic plan

The analyses for this project proceeded in three steps. First, we assessed the descriptive statistics of the segment that auditors rated including the length of the segment, rating time, and indicators of image quality. Second, we analyzed inter-rater reliability. For items with five or fewer response categories or categorical response categories, we assessed inter-rater reliability (IRR) using Fliess' κ statistic (Fleiss, 1971). IRR for continuous items was measured using the intra-class correlation (ICC) of item values from a one-way analysis of variance that measures the proportion of total variation in the rating that exists between segments (Raudenbush & Bryk, 2002). The ICC measures the clustering of items within streets across raters and therefore accounts for the repeated measurement of items.

Fliess' κ assesses the agreement between auditors conditional on the prevalence of a given item across auditors; in other words, κ will penalize disagreement on a given item more if an item is either very frequent or very infrequent in the data. This property of κ is attractive given that we wish to include these items in future statistical analysis where measurement error will cause a larger loss of efficiency on infrequently observed items; it does, however, mean that an item can have a low value of κ even with high levels of inter-rater agreement. Note also that Fliess' κ reflects the probability of agreement above that expected by chance among arbitrary auditors; average pairwise Cohen's κ, an alternate IRR statistic, reflects the probability of agreement above that expected by chance among the auditors who produced the data. In effect, Fliess' κ treats any auditor bias as error rather than a systematic effect, and thus produces κ values that can be interpreted as reflecting what would be expected in a random draw of auditors. These κ values are strictly equal to or less than those produced by average pairwise Cohen's κ (Artstein & Poesio, 2008). Items skipped due to the built-in skip patterns were coded as not present because the negative response to the branching question (e.g., "Are there any non-residential land uses?") implied that the response to the skipped questions would also be negative (e.g., "Is there a community center or library?").

In the third step, we assessed measurement bias by examining whether inter-rater agreement varied by neighborhood characteristics. To assess bias for a categorical item, we coded each street as having perfect agreement when all three raters gave the same response to that item. We then estimated a logistic regression model in which perfect agreement among the three auditors on street segments was regressed on neighborhood characteristics. The purpose of this analysis is to determine whether ratings differed *within* segments, in contrast to the previous analysis that used Fliess' κ and ICC to evaluate whether items were reliably measured *across* segments. With three raters, a disagreement on an item with only two response categories (e.g., present/not present) will mean that two raters agree, leaving only perfect agreement as a suitable measure to test for inter-rater agreement. Since a large number of items we evaluate have only two response categories, we assessed measures using perfect agreement.

Statistically significant results indicate that agreement is associated with the neighborhood characteristic for that item. We considered five neighborhood characteristics measured using 2000 Census tract level data from the NCDB: the *population density* of the census tract measured in 10,000 people per kilometer squared, racial and ethnic neighborhood composition was measured with the *percent non-Latino African American* and *percent Latino*, and tract socioeconomic status was measured using *percent poor* and *percent of homes occupied by homeowners*.

# RESULTS

## Characteristics of Segments and Image Quality on Street View

Characteristics of the rated segments and the image quality of Street View on those segments are reported in Table 1. Auditors reported that segments contained a mean of 19.5 virtual steps, though there was a great deal of variation; the number of virtual steps ranged from 2 to 101 with a standard deviation of 13.28 corresponding, to a large degree, with the length of the segment rated. Inter-rater reliability was high for reported number of steps per segment. The proportion of steps reported to have obstructed views varied considerably, ranging from no obstruction to 100% of steps being obstructed (we excluded two observations on which an auditor reported more obstructed steps than total steps); on average, about one in four steps was obstructed. Obstructions were classified as objects obscuring the rater's line of site to the side of the road (image quality was assessed with different items). The most common obstructions were parked cars and foliage. Consistency across auditors was low for this measure, suggesting that auditors might have had different thresholds for perceiving obstruction.

Mean rating time (based on automatically generated time stamps when an auditor started rating the block on CANVAS until they submitted the final ratings) was about 17 minutes (median=14.6) per segment, with substantial variation between auditors rating the same segment and substantial variation between segments. The mean length was in the midrange of reported completion times for existing in-person audits (Clifton et al., 2007), though of course CANVAS audits require no travel time between segments. Some very long rating times occurred when auditors took breaks in middle of rating segments, thus the mean rating

time reported (based on time-stamps) reflects a conservative estimate of time required to rate segments.

We also report summary statistics of audit time for the eight different data collection modules. Module audit time estimates represent the time spent between confirming item responses with a module, which can be only a rough approximation of the time spent assessing the items in the Street View imagery; auditors often gathered the information to complete items from multiple modules simultaneously. The building condition module took the longest to rate at 4.68 minutes. This module had 52 items (meaning each item took approximately 5 seconds to rate) that spanned the aesthetics and design, disorder, and building characteristics categories. The next longest was the "meta" module that contained items related to the segment and image quality. Although it contained only 7 items auditors took an average of 3.65 minutes to complete (31 seconds per item) this module for two reasons. First, 'meta' was the first module addressed for each street segment, and auditors typically spent some time investigating the street generally before settling in to answer specific items. Second, the 'meta' module included items in which auditors were required to manipulate the Street View interface more completely than when simply assessing presence or absence of features on the segment (by counting virtual steps, counting the number of obstructed views, and zooming to assess pixelation and legibility). The biking characteristics module took the shortest amount of time to complete, taking an average of 0.46 minutes for 6 items (5 seconds per item).

There was little agreement between raters regarding the level of zoom that creates a pixilated image (ICC=0.06), though very few saw pixelation in the initial zoom level (3%). At this zoom level, raters were usually able to read street signs when they were visible (55% of the time a street sign was visible, i.e. 49% of the 88% of segments where street signs were visible). Raters found a moderate level of agreement on this measure (ICC=0.53). Google's camera technologies varied over time (Anguelov 2010); while the specific camera technology could not be determined from the picture, imagery could be empirically categorized into higher- and lower- resolution imagery reliably ($\kappa$=0.82). Twenty-nine percent of blocks in our sample used high-resolution imagery (which had lighter imagery than the lower-resolution imagery) at the start point of the street segment.

### Reliability of Measures included in CANVAS

Values of IRR statistics are plotted in Figure 1; circles represent $\kappa$ values for categorical items while ICC triangles represent values for continuous items. In addition, Table 2 reports a summary of $\kappa$ values for categorical items by measurement category including the category average, standard deviation, and items with the highest and lowest $\kappa$ values (a table of all measures is available as Supplemental Table S1) and Table 3 reports the ICC values for all continuous items. Overall, 64 of 187 items had IRR scores above 0.80 and an additional 33 with IRR scores above 0.60. These kappa statistics are comparable to existing in-person audits (Clifton et al., 2007; Pikora et al., 2002; Raudenbush, 2003).

As one can see from Figure 1, however, the IRR varied considerably across measurement categories. Items measuring the presence of specific land uses were the most reliably measured with an average kappa score of 0.815; most IRR scores were above 0.80. Those

items with the lowest IRR, light industrial and residential types not classified elsewhere (i.e., residential types that were not single-family homes, apartments, etc.), were not prevalent and were therefore sensitive to very small differences between auditors.

Items measuring features of the sidewalk and buildings on the segment also showed relatively high IRR. Half the items in both categories had IRR ratings above 0.80. All the sidewalk items had κ values higher than 0.50; items with the lowest reliability were curb cuts and sidewalk width. The most reliably measured categorical items within the building category asked the auditor to identify a type of building present. The single continuous variable in the building category, a count of buildings on the segment, showed very high reliability (ICC=0.970).

The average measures of IRR for the nature and traffic design categories were relatively low, but the reliability of items within both categories varied considerably, as the large standard deviations in Table 2 demonstrate. The least reliable items in both categories were low-prevalence items. The presence of an ocean was, while not prevalent, the most reliably measured item in the nature category (κ=0.898) and the other natural features that were reliably measured were items measuring foliage buffering the walking path.

Infrequently observed and unreliably measured items in the traffic design category include traffic-calming devices like rumble strips (κ=-0.014) and curb extensions (κ=-0.009). Many traffic design items were measured reliably including items describing traffic patterns (e.g., type of intersection, cul-de-sacs, one-way streets, presence of driveways and angled parking) and pedestrian accommodations (e.g., traffic signals, pedestrian crossing signals, and crosswalks). CANVAS therefore appears to measure traffic patterns and pedestrian accommodations more reliably than traffic-calming devices. High levels of reliability were found for both continuous measures of traffic design: the posted speed limit (ICC=0.903) and number of vehicle lanes (ICC=0.898).

Items measuring disorder, aesthetics, and amenities had low levels of reliability with less variance than natural features or traffic design, resulting in fewer very reliably rated items. Among the disorder items, those measuring the physical condition of buildings on the block were most reliably measured while items measuring small items like litter or bottles were among the least reliably measured. In the aesthetic and design category, the continuous measures counting the number of non-rectangular buildings (ICC=0.868) had a very high reliability and the number of building colors had a moderate level of reliability (ICC=0.704). Categorical items were generally not measured very reliably with the least reliable being items measuring the architectural concepts. Only three items measuring amenities fell in the moderate level of IRR: the presence of bike route signs (κ=0.779), presence of public garbage cans (κ=0.719), and the type of bus stop present (κ=0.626). Besides these, other amenities on the sidewalk were not reliably measured.

Finally, items measuring pedestrian access like barriers to walking and the ease of crossing streets were not reliably measured. None of the 16 measures in the category attained even a moderate level of IRR.

Most measures of IRR were comparable across the central city and metropolitan samples. Figure 2 plots the central city versus metropolitan measurement of kappa for each item (a table of item-level comparisons is available in Supplemental Table S2). The dashed line indicates where the kappa statistic would be identical and the dotted line indicates the line of best fit. The two measures were correlated at $\rho=0.84$. Of items with sufficient variation to measure kappa statistics in both central city and metropolitan samples (N=143), 81% differed by 0.20 points or less. Many of those that varied by more were items where the expected frequency of observation differed substantially across locations (e.g., golf courses, parking structures).

### Measurement Bias Associated with Neighborhood Characteristics

Table 4 reports logistic regression coefficients for items for which at least one neighborhood characteristic showed a statistically significant association (at the $p<0.05$ level) with perfect auditor agreement on the item (Supplemental Table S3 shows the results for all items measured). Positive coefficients indicate that the neighborhood characteristic is associated with a greater probability of perfect agreement among the auditors while a negative coefficient reflects the inverse. We measured agreement on 155 out of 164 categorical items (4 were excluded because they had perfect agreement and 5 because there was insufficient disagreement to draw statistical conclusions) across five independent variables resulting in a total of 765 statistical comparisons. Of these, only 38 comparisons, or 4.9%, were statistically significant at the $p<0.05$ level, precisely the level of significant results that we would expect by chance due to Type I error at $\alpha=0.05$. Similarly, 5 of the comparisons (0.007%) were significant at the *p<0.01* level, approximately 2 less than we would expect by chance.

The significant associations that are present appear to be relatively randomly distributed across measurement categories and independent variables. The only possible exception is that agreement tends to be significantly higher on segments in tracts with a larger percent of homeowners. For 13 items the percent of owner-occupied housing units was statistically significantly different from zero, almost twice the number of significant associations with any of the other four independent variables. We asked auditors to identify whether non-residential land use existed on the block. If none existed, auditors then skipped questions on specific types of non-residential land use, all of which were coded as not present. This could explain why home ownership predicted agreement more than other neighborhood characteristics. Even the results of owner occupied housing are neither extreme nor unreasonable for what one would expect to occur due to random chance. Because the percent owner-occupied predicted perfect agreement frequently, there were fewer systematic biases in the central city neighborhoods, where the percentage of owner-occupied housing tended to be lower, than the metropolitan sample (results of models stratified by sample can be found in Supplemental Tables S4 and S5). It also the case that statistically significant coefficients measuring the association of agreement with both population density and percent black were more often than not negatively correlated, though the frequency of statistically significant relationships for both is about what we would expect by chance. Results of sensitivity tests regressing available Census block level characteristics against agreement were substantively similar to tract-level estimates (results available in

Supplemental Table S6). Thus we have reasonable confidence that measurement error is not systematically biased by neighborhood characteristics.

## DISCUSSION

To our knowledge, this is the first study to examine the reliability of Street View measures on a nationwide sample of street segments in the United States (see Odgers et al., 2012 for a nationwide study in the United Kingdom; and Griew et al., 2013 for the development of a Street View tool developed in the UK). Approximately half of items had good or excellent levels of inter-rater reliability and those items measuring land use, the quality and availability of sidewalks, and building features were the most reliably measured. Pedestrian access features were the only category of items with consistently unreliable measures. The low reliability of pedestrian access features likely reflects the low prevalence of measures in a nationwide sample. The level of inter-rater reliability was consistent with the inter-rater reliability of in-person audits (see Supplemental Table S7 for a comparison of inter-rater reliability on PEDS items between the national reliability sample and those reported by Clifton et al., 2007 in their PEDS validation study). In addition, this study is unique in that it examines potential measurement *bias* by other social, economic, and demographic characteristics of neighborhoods. There appear to be few systematic associations between auditor disagreement and the characteristics of the neighborhood surrounding the segment being rated.

The results suggest that researchers should have reasonable confidence in many measures rated using Street View. Using Street View permits studies to acquire theoretically motivated measures of neighborhood conditions for studies that span large geographic areas. The CANVAS web application provides a cost-effective, standardized method of measuring features of the neighborhood environment. Our raters conducted a comprehensive protocol in an average of 17 minutes per block. On average, this translated to rating an item every five seconds during the reliability field study. Questions may be selected from existing protocols or developed by researchers to measure relevant neighborhood features within budget and time constraints; given the comprehensiveness of our reliability study that included a nearly complete inventory of three audit instruments and questions from additional instruments, 17 minutes should be viewed as an upper limit on the rating time per street since other studies would likely only rate a subset of items. Reducing the number of items will also substantially reduce the time required to train raters and further reduce the cost of neighborhood audits. By reducing travel and protocol design costs, CANVAS creates new opportunities to obtain data and study the influence of neighborhood characteristics on individuals.

The extensive training of the raters in our study improved the reliability of responses. While we would advise researchers to conduct training before using the items on CANVAS, future studies would likely not require as much training as we conducted. We trained raters to rate 187 items; most studies would likely reduce the number of items and increase the efficiency of training raters. In addition, our training included iterations that adapted existing in-person protocols for virtual audits and modifications to the CANVAS application. We plan to

further improve the user interface of CANVAS to reduce auditor fatigue, which became a problem in the study.

## Implications for Future Research

The greatest benefit of CANVAS is that it overcomes the cost of constructing consistent neighborhood measures across places. As Brownson, et al. (2009) note in there review of the state of neighborhood research methods, inconsistent measurement impedes generalization across studies. Such generalization is important to determine whether associations exist between neighborhood conditions and individual outcomes. By increasing the geographic scope of neighborhood studies, CANVAS allows researchers to investigate how city or metropolitan conditions might influence the relationship between neighborhoods and individuals. These studies might reveal why policies to reduce disorder or promote walking might vary in efficacy or cost-effectiveness depending on location.

Our study points to a number of areas of future research. First, future research should consider how to assign auditors to segments and items: specifically, whether reliability might be improved by having each auditor specialize in rating a small number of items, gaining skill and enhancing consistency. Second, future research could adapt the protocol so that it reflects the urban design and cultural conditions of other countries. Third, data collected through CANVAS can be archived and used to build longitudinal data on places, thus addressing one of the major shortcomings of research on place-based effects (Brownson et al., 2009; Lovasi & Goldsmith, 2014).

## Limitations and Considerations for Future Studies

Although CANVAS provides a promising tool for neighborhood research, there are some limitations to its use. First, what researchers can gain from the geographic breadth of CANVAS they lose in the depth of measures typically obtained through in-person audits. For example, noise, social incivilities, and the presence of aggressive dogs – all measures that have been shown to influence physical activity – cannot be measured by virtual audits. Additionally, items that are small or that have a large amount of temporal variability are also not well measured (Rundle et al., 2011). Second, virtual neighborhood assessments are limited to the locations where Google Street View (or similar imagery) data are available. In our national reliability study, up to 40% of segments selected in the metropolitan sample were not "Street Viewable." This limitation will become less of a problem as Google and its competitors (e.g., Microsoft Bing's Streetside application) expand the geographic coverage to more streets and even off-street pedestrian trails. Google's release of historical Street View imagery in the spring of 2014 offers new opportunities for understanding temporal change at locations where historical imagery exists (Google, 2014). Third, CANVAS requires that observation locations be passed through an Internet connection. Because researchers cannot control the use of data passed through this connection, studies sampling streets based on where respondents live represent a violation of human subjects' confidentiality. For this reason, we believe that the combination of geographic sampling techniques with spatial interpolation methods like kriging should be explored (e.g., Bader & Ailshire, 2014). Beyond the limitations of virtual audits, rater fatigue became a problem during the study, potentially reducing the quality of ratings. Finally, the sample size of 150

segments is relatively small, reducing the statistical power of our models and making it more difficult to discern differences in agreement by neighborhood characteristics.

## Conclusions

Given the results of this study, we recommend CANVAS to researchers who want to include neighborhood observations in their study. We particularly recommend CANVAS for studies where participants are geographically dispersed (e.g., nationwide studies). Virtual audits conducted through CANVAS cannot, and we believe should not, replace in-person audits where non-visual (e.g., sounds, smells, textures, etc.) data can be collected. When researchers risk revealing specific addresses or locations of respondents to online APIs, in-person audits should be used in order to protect human subjects' identity. If, however, in-person audits are prohibitively costly, the pairing collection of CANVAS with spatial analytic methods to estimate neighborhood conditions (e.g., Savitz & Raudenbush, 2009; Bader & Ailshire, 2014) provides a method of obtaining neighborhood data. Our findings support the further development and application of virtual audit tools to overcome current problems measuring neighborhood environments.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

## REFERENCES

Anguelov D, Dulong C, Filip D, Frueh C, Lafon S, Lyon R, Weaver J. Google Street View: Capturing the World at Street Level. Computer. 2010; 43(6):32–38.

Artstein R, Poesio M. Inter-Coder Agreement for Computational Linguistics. Computational Linguistics. 2008; 34(4):555–596.

Bader MDM, Ailshire JA. Creating Measures of Theoretically Relevant Neighborhood Attributes at Multiple Spatial Scales. Sociological Methodology. 2014 0081175013516749.

Badland HM, Opit S, Witten K, Kearns RA, Mavoa S. Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits? Journal of Urban Health. 2010; 87(6):1007–1016. [PubMed: 21104331]

Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the Built Environment for Physical Activity: State of the Science. American Journal of Preventive Medicine. 2009; 36(4, Supplement 1):S99–S123. e12. [PubMed: 19285216]

Carter, W.; Dougherty, J.; Grigorian, K. Apendix III: NORC Narrative, Videotaping Neighborhoods. In: Earls, Felton J.; Raudenbush, Stephen W.; Reiss, Albert J., Jr; Sampson, Robert J., editors. Project on Human Development in Chicago Neighborhoods (PHDCN): Systematic Social Observation, ICPSR Study 13578. Ann Arbor, MI: Interuniversity Consortium for Political and Social Research (distributor; 1995. p. 215-229.Retrieved from http://www.icpsr.umich.edu/cocoon/PHDCN/STUDY/13578.xml

Clarke P, Ailshire J, Melendez R, Bader M, Morenoff J. Using Google Earth to Conduct a Neighborhood Audit: Reliability of a Virtual Audit Instrument. Health & Place. 2010; 16(6):1224–1229. [PubMed: 20797897]

Clifton KJ, Livi Smith AD, Rodriguez D. The development and testing of an audit for the pedestrian environment. Landscape and Urban Planning. 2007; 80(1–2):95–110.

Cummins S, Macintyre S, Davidson S, Ellaway A. Measuring neighbourhood social and material context: generation and interpretation of ecological data from routine and non-routine sources. Health & Place. 2005; 11(3):249–260. [PubMed: 15774331]

Curtis A, Duval-Diop D, Novak J. Identifying Spatial Patterns of Recovery and Abandonment in the Post-Katrina Holy Cross Neighborhood of New Orleans. Cartography and Geographic Information Science. 2010; 37:45–56.

Day K, Boarnet M, Alfonzo M, Forsyth A. The Irvine-Minnesota Inventory to Measure Built Environments: Development. American Journal of Preventive Medicine. 2006; 30(2):144–152. [PubMed: 16459213]

Django Software Foundation. Django (Version 1.4). Python. 2012. Retrieved from https:// www.djangoproject.com/

Earls, FJ.; Raudenbush, SW.; Reiss, AJ.; Sampson, RJ. Project on Human Development in Chicago Neighborhoods (PHDCN): Systematic Social Observation, 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor; 1995. (No. ICPSR 13578-v1) Retrieved from http://www.icpsr.umich.edu/cocoon/PHDCN/STUDY/13578.xml

Ewing, R.; Clement, O.; Handy, S.; Brownson, RC.; Winston, E. Identifying and Measuring Urban Design Qualities Related to Walkability -- Final Report. Princeton, NJ: Robert Wood Johnson Foundation; 2005. Retrieved from http://www.activelivingresearch.org/files/ FinalReport_071605.pdf

Ewing R, Handy S, Brownson RC, Clemente O, Winston E. Identifying and Measuring Urban Design Qualities Related to Walkability. Journal of Physical Activity & Health. 2006; 3(Suppl 1):S223–S240.

Fleiss J. Measuring nominal scale agreement among many raters. Psychological Bulletin November 1971. 1971; 76(5):378–382.

Google. Go back in time with Street View. 2014 Apr 23. Retrieved from http:// googleblog.blogspot.com/2014/04/go-back-in-time-with-street-view.html

Griew P, Hillsdon M, Foster C, Coombes E, Jones A, Wilkinson P. Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. International Journal of Behavioral Nutrition and Physical Activity. 2013; 10(1):1–7. [PubMed: 23281722]

Lovasi GS, Goldsmith J. Invited Commentary: Taking Advantage of Time-Varying Neighborhood Environments. American Journal of Epidemiology. 2014; 180(5):462–466. [PubMed: 25117659]

Mills, E. Google's street-level maps raising privacy concerns. 2007 Jun 4. Retrieved May 14, 2014, from http://usatoday30.usatoday.com/tech/news/internetprivacy/2007-06-01-google-maps-privacy_n.htm

Mujahid MS, Diez Roux AV, Morenoff JD, Raghunathan T. Assessing the Measurement Properties of Neighborhood Scales: From Psychometrics to Ecometrics. American Journal of Epidemiology. 2007; 165(8):858–67. [PubMed: 17329713]

MySQL AB. MySQL (Version 5.1). MySQL AB. 2005. Retrieved from http://www.mysql.com

Odgers CL, Caspi A, Bates CJ, Sampson RJ, Moffitt TE. Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. Journal of Child Psychology and Psychiatry. 2012; 53(10):1009–1017. [PubMed: 22676812]

Perkins DD, Florin P, Rich RC, Wandersman A, Chavis DM. Participation and the social and physical environment of residential blocks: Crime and community context. American Journal of Community Psychology. 1990; 18(1):83–115.

Pikora TJ, Bull FCL, Jamrozik K, Knuiman M, Giles-Corti B, Donovan RJ. Developing a reliable audit instrument to measure the physical environment for physical activity. American Journal of Preventive Medicine. 2002; 23(3):187–194. [PubMed: 12350451]

Raudenbush, S. The Quantitative Assessment of Neighborhood Social Environments. In: Kawachi, I.; Berkman, LF., editors. Neighborhoods and health. Oxford: Oxford University Press; 2003. p. 112-131.Retrieved from http://catalog.wrlc.org/cgi-bin/Pwebrecon.cgi?BBID=4202011

Raudenbush, SW.; Bryk, AS. Hierarchical linear models : applications and data analysis methods. Thousand Oaks, CA: Sage Publications; 2002.

Raudenbush SW, Sampson RJ. Ecometrics: Toward a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods. Sociological Methodology. 1999; 29:1–41.

Reiss AJ. Systematic observation of natural social phenomena. Sociological Methodology. 1971; 3:3–33.

Rundle AG, Bader MDM, Richards CA, Neckerman KM, Teitler JO. Using Google Street View to Audit Neighborhood Environments. American Journal of Preventive Medicine. 2011; 40(1):94–100. [PubMed: 21146773]

Saelens BE, Glanz K. Work Group I: Measures of the Food and Physical Activity Environment: Instruments. American Journal of Preventive Medicine. 2009; 36(4, Supplement 1):S166–S170. [PubMed: 19285209]

Sampson RJ, Raudenbush SW. Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. American Journal of Sociology. 1999; 105(3):603–51.

Savitz NV, Raudenbush SW. Exploiting Spatial Dependence to Improve Measurement of Neighborhood Social Processes. Sociological Methodology. 2009; 39(1):151–183.

Taylor RB, Gottfredson S, Brower S. Block Crime and Fear: Defensible Space, Local Social Ties, and Territorial Functioning. Journal of Research in Crime and Delinquency. 1984; 21(4):303–331.

U.S. Bureau of the Census. New York City Housing and Vacancy Survey (NYCHVS). 2011. Retrieved from http://http://www.census.gov/housing/nychvs/

Vargo, J.; Stone, B.; Glanz, K. Google Walkability: A New Tool for Local Planning and Public Health Research?. Journal of Physical Activity & Health. 2011. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21946250

Wilson JS, Kelly CM, Schootman M, Baker EA, Banerjee A, Clennin M, Miller DK. Assessing the Built Environment Using Omnidirectional Imagery. American Journal of Preventive Medicine. 2012; 42(2):193–199. [PubMed: 22261217]

## Research Highlights

- Develops online application to improve neighborhood data collected using Google Street View

- Application provides reliable measures of neighborhood walkability and disorder from a national sample

- Rater disagreements are largely uncorrelated with neighborhood characteristics

- Application can be used to improve consistency across studies and lower technological barriers

**Figure 1.**
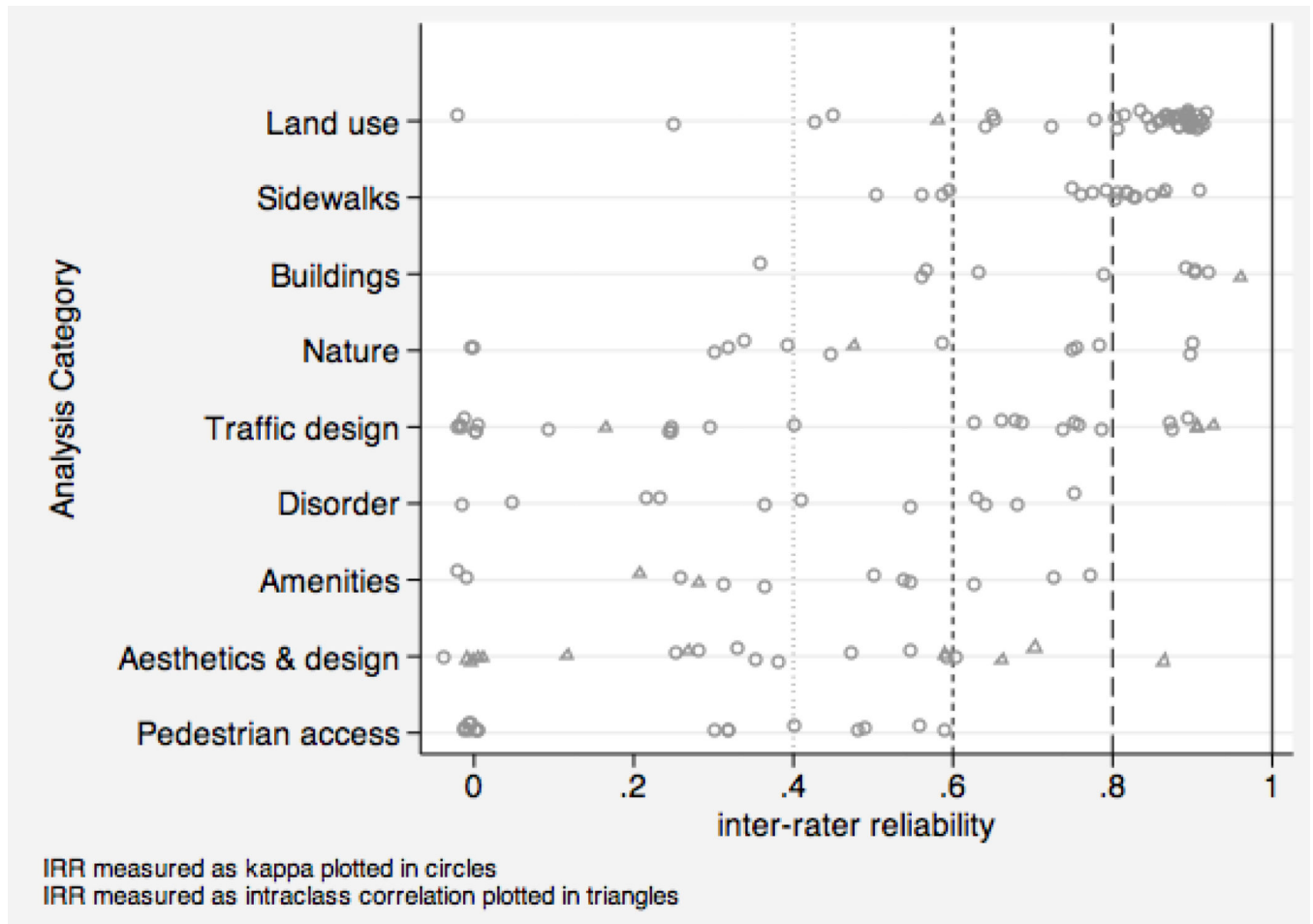Inter-rater reliability of items by category, kappas for categorical items plotted as circles and
intraclass correlation coefficients for continuous items plotted as triangles
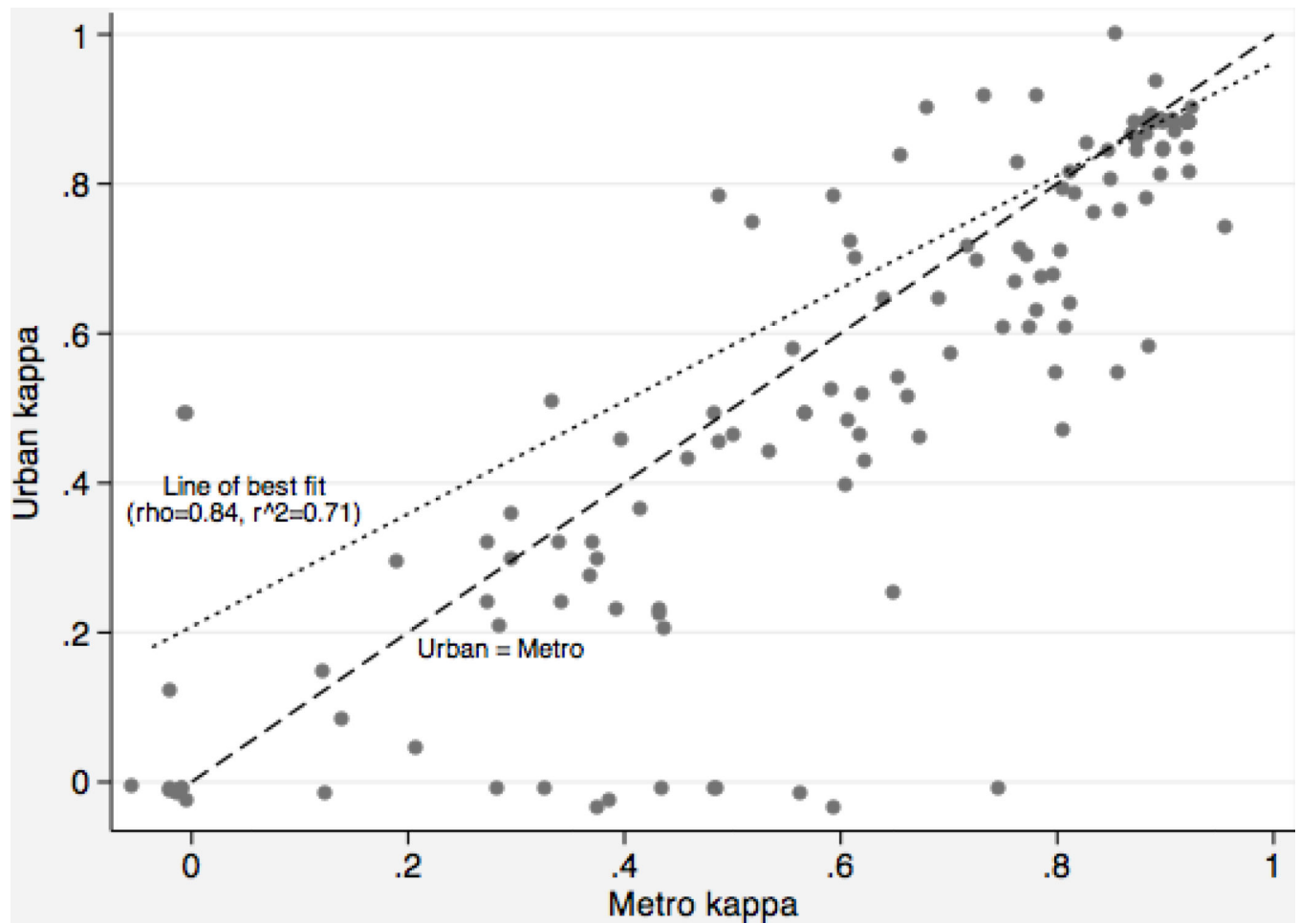
**Figure 2.**
Kappa values of items measuring inter-rater reliability in the metropolitan sample compared to the central city sample. The dashed line indicates no difference between metropolitan and central-city values of IRR.

**Table 1**

Characteristics of segments and image quality rated in CANVAS

| A. *Segment Characteristics* | | Mean | S.D. | Median | Min. | Max. | ICC |
|---|---|---|---|---|---|---|---|
| Number of steps | | 19.48 | 13.28 | 17 | 2 | 101 | 0.99 |
| Percentage of steps obstructed[a] | | 26.98 | 27.10 | 20 | 0.00 | 100.00 | 0.10 |
| **B. *Time to Rate (in minutes)*** | | Mean | S.D. | Median | Min. | Max. | ICC |
| Entire segment | | 17.11 | 10.00 | 14.59 | 4.72 | 108.78 | 0.07 |
| Building conditions module | | 4.68 | 4.46 | 3.88 | 0.77 | 78.08 | 0.03 |
| Meta characteristics module | | 3.65 | 4.30 | 2.63 | 0.63 | 54.37 | 0.04 |
| Land use module | | 2.22 | 2.12 | 1.48 | 0.27 | 12.15 | 0.00 |
| Road and parking module | | 1.63 | 1.16 | 1.44 | 0.28 | 14.25 | 0.00 |
| First intersection module | | 1.58 | 1.65 | 1.27 | 0.37 | 21.43 | 0.04 |
| Sidewalk characteristics module | | 1.52 | 3.10 | 1.20 | 0.05 | 61.13 | 0.06 |
| Pedestrian access module | | 1.38 | 1.29 | 1.00 | 0.25 | 11.23 | 0.00 |
| Biking characteristics module | | 0.46 | 0.78 | 0.25 | 0.07 | 10.92 | 0.12 |
| **C. *image Quality*** | | **Freq.** | **Pct.** | | | | ICC |
| Zoom level (number of clicks on zoom) where pixelation occurs | | | | | | | 0.06 |
| | 0 | 14 | 3.11 | | | | |
| | 1 | 191 | 42.44 | | | | |
| | 2 | 118 | 26.22 | | | | |
| | 3 | 97 | 21.56 | | | | |
| | 4 | 30 | 6.67 | | | | |
| Legibility of street signs at initial zoom level | | | | | | | 0.53[b] |
| | Clearly legible | 136 | 30.22 | | | | |
| | Blurry but legible | 84 | 18.67 | | | | |
| | Unreadable | 178 | 39.56 | | | | |
| Camera technology | | | | | | | 0.82[c] |
| | Dark | 319 | 70.89 | | | | (kappa) |
| | Light | 131 | 29.11 | | | | |

Notes:

[a] $x^2$ of segments had percentage of obstructed streets greater than 100 and were set to missing

[b] The category "No sign visible" not included in intra-class correlation analysis

[c] Kappa value reported because camera technology included only two response categories

**Table 2**

Mean kappa statistic, lowest reliability, and highest reliability items of categories in CANVAS study

| Category (N) | Mean kappa | [S.D.] | Lowest items (Rank) Item | kappa | %agree | Highest items (Rank) Item | kappa | %agree |
|---|---|---|---|---|---|---|---|---|
| Land use (48) | 0.815 | [0.184] | (1) Other residential pres. | −0.009 | 0.982 | (46) Amt. used for libraries or bookshops | 0.911 | 0.960 |
| | | | (2) Light industrial land use pres. | 0.259 | 0.964 | (47) Car dealership pres. | 0.912 | 0.960 |
| | | | (3) Other industrial land use pres. | 0.441 | 0.973 | (48) Agricultural/ranching land use pres. | 0.914 | 0.960 |
| Sidewalks (17) | 0.760 | [0.117] | (1) Curb cuts | 0.514 | 0.658 | (15) Pedestrian path is continuous | 0.86 | 0.924 |
| | | | (2) Condition of pedestrian path | 0.574 | 0.729 | (16) Fence buffers road & path | 0.878 | 0.951 |
| | | | (3) Width of sidewalk | 0.586 | 0.709 | (17) Grass buffers road from path | 0.905 | 0.938 |
| Buildings (9) | 0.722 | [0.199] | (1) Prominence of garage doors | 0.358 | 0.584 | (7) Drive-through pres. | 0.894 | 0.951 |
| | | | (2) Amt. of segment with front porches | 0.564 | 0.747 | (8) Shopping mall pres. | 0.902 | 0.956 |
| | | | (3) Amt. of segment with garage doors | 0.58 | 0.709 | (9) Big-box stores pres. | 0.911 | 0.960 |
| Nature (13) | 0.499 | [0.310] | (1) Lake or pond pres. | −0.009 | 0.982 | (11) Landscaping buffers road & path | 0.787 | 0.898 |
| | | | (2) Stream, river or canal pres. | −0.009 | 0.982 | (12) Hedges buffers road & path | 0.889 | 0.956 |
| | | | (3) Mountain or hills pres. | 0.296 | 0.929 | (13) Ocean pres. | 0.898 | 0.996 |
| Traffic design (26) | 0.420 | [0.355] | (1) Rumble strips or bumps pres. | −0.014 | 0.973 | (24) Cul-de-sac or perm. street closure | 0.885 | 0.982 |
| | | | (2) Curb extension pres. | −0.009 | 0.982 | (25) Traffic signal pres. | 0.891 | 0.933 |
| | | | (3) Overpass or underpass pres. | −0.004 | 0.991 | (26) Type of intersection | 0.92 | 0.956 |
| Disorder (11) | 0.414 | [0.258] | (1) Empty beer bottles visible in street | −0.003 | 0.991 | (9) Amt. of abandoned buildings or lots | 0.643 | 0.902 |
| | | | (2) Graffiti painted over pres. | 0.058 | 0.951 | (10) Dilapidated exterior walls pres. | 0.677 | 0.933 |
| | | | (3) Amt. of litter apparent | 0.223 | 0.609 | (11) Broken or boarded windows pres. | 0.751 | 0.933 |
| Amenities (11) | 0.420 | [0.264] | (1) Water fountain pres. | −0.016 | 0.969 | (9) Type of bus stop pres. | 0.626 | 0.956 |
| | | | (2) Amt. of street vendors or stalls | −0.003 | 0.991 | (10) Public garbage cans pres. | 0.719 | 0.938 |
| | | | (3) Amt. of benches on segment | 0.262 | 0.916 | (11) Bike route signs pres. | 0.779 | 0.987 |
| Aesthetics & design (11) | 0.359 | [0.198] | (1) Degree of enclosure | −0.041 | 0.307 | (9) Significant open view of object pres. | 0.55 | 0.884 |
| | | | (2) Number of long sight lines | 0.129 | 0.551 | (10) Bldgs. pres. on 1/2+ of seg. | 0.596 | 0.876 |
| | | | (3) Level of interesting architecture | 0.265 | 0.536 | (11) Walk through parking to most bldgs. | 0.6 | 0.911 |

| Category (N) | Mean kappa | [S.D.] | Lowest items | | | | Highest items | | | |
| | | | (Rank) Item | kappa | %agree | | (Rank) Item | kappa | %agree |
|---|---|---|---|---|---|---|---|---|---|
| Pedestrian access(16) | 0.215 | [0.240] | (1) Drainage ditch causes barrier | −0.009 | 0.982 | | (14) Road of 6+ lanes causes barrier | 0.492 | 0.947 |
| | | | (2) Impassable land use causes barrier | −0.009 | 0.982 | | (15) Railroad track causes barrier | 0.558 | 0.973 |
| | | | (3) Amt. of barrier caused by impassable use | −0.006 | 0.982 | | (16) Highway causes barrier | 0.596 | 0.991 |

**Table 3**

Intra-class correlations of continuous items included in CANVAS study, by category

| Category | Item | ICC |
|---|---|---|
| Land use | Proportion of segment with active uses | 0.595 |
| Sidewalks | Number of connections on pedestrian path | 0.854 |
| Buildings | Number of buildings visible | 0.970 |
| Nature | Number of small planters | 0.472 |
| Traffic design | Number of vehicle lanes | 0.898 |
| | Posted speed limit | 0.903 |
| Amenities | Number of other street items | 0.198 |
| | Number of pieces of street furniture | 0.291 |
| Aesthetics & design | Proportion of windows at street level | 0.000 |
| | Number of accent colors | 0.000 |
| | Number of pieces of public art | 0.000 |
| | Proportion of segment with historic building frontage | 0.000 |
| | Proportion of view that is sky | 0.273 |
| | Number of buildings with identifiers | 0.585 |
| | Proportion of segment with street wall | 0.662 |
| | Number of basic building colors | 0.704 |
| | Number of buildings with non-rectangular shapes | 0.868 |

**Table 4**

Items with statistically significant coefficient from logistic regression of perfect inter-rater agreement on categorical neighborhood characteristics measured in CANVAS, by category

| No. | Item | N | Kappa | Pop. Dens. | Pet. black | Pet. poor | Pet. foreign-born | Pet. Home-owners | Const. |
|---|---|---|---|---|---|---|---|---|---|
| **Land use** | | | | | | | | | |
| 2 | Light industrial land use present | 2 | 0.26 | 11.720 | 0.001 | −0.034 | −0.038 | 0.053 * | 1.777 |
| 3 | Other industrial land use present (Minn–Irvine) | 2 | 0.44 | 9.513 | 0.000 | −0.038 | −0.021 | 0.062 * | 1.595 |
| 4 | Medium or heavy industrial land use present | 2 | 0.44 | 9.513 | 0.000 | −0.038 | −0.021 | 0.062 * | 1.595 |
| 7 | Duplex home present | 2 | 0.65 | 0.170 | −0.006 | −0.005 | −0.040 * | −0.006 | 3.371 * |
| 11 | Town home, condo or apartment with 3+ units pres. | 2 | 0.81 | 5.777 | −0.013 | 0.011 | −0.060 ** | 0.014 | 2.612 |
| 16 | Amount of segment used for restaurants | 3 | 0.86 | −6.340 | 0.023 | 0.089 | −0.003 | 0.042 * | −1.053 |
| 18 | Amount of segment used for corner stores | 3 | 0.87 | −1.718 | 0.015 | 0.047 | −0.003 | 0.036 * | −0.338 |
| **Sidewalks** | | | | | | | | | |
| 1 | Curb cuts | 4 | 0.51 | 13.042 | −0.014 | −0.013 | −0.005 | −0.024 * | 1.436 |
| 3 | Width of sidewalk | 3 | 0.59 | 1.026 | −0.026 ** | 0.050 * | −0.025 | 0.027 * | −1.162 |
| **Buildings** | | | | | | | | | |
| 4 | Amount of segment with barred windows | 4 | 0.62 | −2.697 | −0.016 | −0.032 | −0.044 * | −0.015 | 4.149 ** |
| **Nature** | | | | | | | | | |
| 7 | Amount of trees shading walking area | 3 | 0.45 | −2.499 | −0.004 | −0.006 | 0.039 * | −0.005 | 0.239 |
| 10 | Trees buffering road & path | 2 | 0.76 | −1.263 | −0.007 | 0.075 * | −0.025 | 0.029 * | −0.886 |
| **Traffic design** | | | | | | | | | |
| 10 | Parking structure visible | 2 | 0.24 | 52.406 | −0.068 * | 0.139 | −0.093 | 0.050 | 1.326 |
| 11 | Street-level use of parking structure | 3 | 0.24 | 52.406 | −0.068 * | 0.139 | −0.093 | 0.050 | 1.326 |
| 12 | Freeway over- or underpass (Minn–Irvine) | 4 | 0.25 | 9.918 | −0.024 | 0.087 | 0.093 | 0.073 * | −1.723 |
| 15 | Median or traffic island present (PEDS) | 2 | 0.64 | −3.894 | −0.003 | 0.132 * | 0.056 | 0.028 | −1.268 |
| 25 | Traffic signal present | 2 | 0.89 | −7.148 * | 0.003 | 0.013 | 0.036 | 0.012 | 1.209 |
| **Disorder** | | | | | | | | | |
| 2 | Graffiti painted over present | 2 | 0.06 | −9.006 * | −0.039 ** | 0.000 | −0.024 | 0.010 | 4.564 * |
| 5 | Level of cleanliness and building maintenance | 3 | 0.37 | −10.239 * | −0.022 * | 0.017 | −0.008 | 0.003 | 1.232 |
| 7 | Amount of graffiti apparent | 3 | 0.55 | −11.415 ** | −0.014 | −0.013 | −0.025 | −0.007 | 4.157 * |

| No. | Item | N | Kappa | Pop. Dens. | Pct. black | Pct. poor | Pct. foreign-born | Pct. Home-owners | Const. |
|---|---|---|---|---|---|---|---|---|---|
| 8 | Level of maintenance of buildings | 4 | 0.63 | -5.388 | -0.003 | -0.015 | -0.032 * | -0.004 | 2.532 * |
| 9 | Amount of abandoned buildings or lots on segment | 4 | 0.64 | -1.847 | -0.012 | -0.056 * | 0.004 | -0.016 | 4.079 ** |
| 10 | Dilapidated exterior walls present | 2 | 0.68 | -1.132 | -0.014 | -0.045 | -0.010 | -0.021 | 4.704 ** |
| **Amenities** | | | | | | | | | |
| 4 | Bike lanes present | 2 | 0.32 | -9.660 * | 0.011 | 0.046 | 0.027 | 0.017 | 1.762 |
| 5 | Amount of outdoor dining facilities on segment | 3 | 0.37 | 11.729 | -0.010 | 0.015 | -0.010 | 0.055 ** | -0.012 |
| 10 | Public garbage cans present | 2 | 0.72 | -0.090 | 0.010 | 0.031 | 0.022 | 0.039 * | -0.685 |
| 11 | Bike route signs present | 2 | 0.78 | -15.648 * | -0.002 | 0.051 | 0.031 | 0.004 | 4.117 |
| **Aesthetics & design** | | | | | | | | | |
| 5 | Amount of in-building designs present | 3 | 0.34 | -2.227 | 0.004 | 0.010 | 0.029 * | 0.003 | -0.789 |
| 6 | Building setback from sidewalk | 3 | 0.36 | 1.904 | -0.018 * | 0.007 | -0.016 | -0.017 | 0.828 |
| 9 | Significant open view of object present | 2 | 0.55 | -0.763 | 0.040 * | -0.005 | 0.011 | 0.030 * | -0.576 |
| **Pedestrian access** | | | | | | | | | |
| 10 | Convenience of crossing | 3 | 0.32 | -7.523 * | -0.004 | 0.072 * | 0.008 | 0.023 | -1.052 |
| 13 | Amount of barrier caused by railroad track | 3 | 0.49 | 0.370 | -0.022 | 0.062 | 0.009 | 0.057 * | -0.319 |