

Identifying and Measuring Urban Design Qualities Related to Walkability

Reid Ewing, Susan Handy, Ross C. Brownson,
Otto Clemente, and Emily Winston

Background: In active living research, measures used to characterize the built environment have been mostly gross qualities such as neighborhood density and park access. This project has developed operational definitions and measurement protocols for subtler urban design qualities believed to be related to walkability. *Methods:* Methods included: 1) recruiting an expert panel; 2) shooting video clips of streetscapes; 3) rating urban design qualities of streetscapes by the expert panel; 4) measuring physical features of streetscapes from the video clips; 5) testing inter-rater reliability of physical measurements and urban design quality ratings; 6) statistically analyzing relationships between physical features and urban design quality ratings, 7) selecting of qualities for operationalization, and 8) developing of operational definitions and measurement protocols for urban design qualities based on statistical relationships. *Results:* Operational definitions and measurement protocols were developed for five of nine urban design qualities: imageability, visual enclosure, human scale, transparency, and complexity. *Conclusions:* A field survey instrument has been developed, tested in the field, and further refined for use in active living research.

Key Words: visual assessment survey, active living, streetscapes, pedestrians

A growing body of research provides evidence of a link between the built environment and active living.¹ However, to date, measures used to characterize the built environment in travel behavior and physical activity research have been mostly gross qualities such as neighborhood density, street connectivity, and distance to parks.²⁻⁷ Audit instruments have proliferated for assessing the walkability and bikeability of environments, but these too have characterized the built environment with crude measures such as number of travel lanes and presence of marked crosswalks.⁸

Urban designers point to subtler qualities that may influence choices about active travel and active leisure time. These are sometimes referred to as perceptual qualities of the urban environment or, alternately, just urban design qualities. Classic readings in urban design are filled with references to these qualities (for

Ewing and Clemente are with the National Center for Smart Growth, University of Maryland, College Park, MD 20742. Handy is with the Dept of Environmental Science and Policy, University of California Davis, Davis, CA 95616. Ross C. Brownson is with the Prevention Research Center, Saint Louis University School of Public Health, St. Louis, MO 63104. Winston is with the Institute of Transportation Studies, University of California Davis, Davis, CA 95616.

example,⁹⁻¹⁴). Urban designers presume that these qualities are important for active street life, but have little empirical evidence to back the claim. Until urban design qualities can be measured, and the propensity to walk can be related empirically to these measures, this presumption remains untested.

For routine physical activity, no element of the urban environment is more important than streets. This is where active travel to work, shop, eat out, and engage in other daily activities takes place, and where walking for exercise mostly occurs. While parks, plazas, trails, and other public places have a role in physical activity, this study focuses on urban design qualities that make one street more walkable than another.

The overall goal of this project was to develop operational definitions and measurement protocols for key urban design qualities of streetscapes. The operational definitions take the form of statistically-derived equations that link objectively measured physical features of the environment to ratings of urban design qualities. The main purpose of the project was not to test the association between these qualities and walkability or walking itself, but rather to develop measurement methods that can be used by researchers to study these relationships.

Conceptual Framework and Definitions

The conceptual model underlying this study considers the role of perceptions as they intervene (or mediate) between the physical features of the environment and walking behavior. We hypothesize that the perceptions lie on the causal path between objective measurements and subjective reactions, as diagrammed in Figure 1.

Physical features can be measured objectively and are, in a sense, “facts” about the environment.^{8, 15-17} Urban design qualities may be assessed with a degree of objectivity by outside observers, though not with the objectivity of physical features; these qualities should be strongly associated with objective physical features and vary relatively little from person to person. Individual reactions, in contrast, are influenced by physical features and perceptions of physical features but cannot be measured by observing the physical environment alone but instead must be measured on an individual basis. All of these factors—physical features, urban design qualities, and individual reactions—determine the overall walkability of a street, by which we mean the way individuals feel about the street as a place to walk. By examining the first part of this model—the link between physical features and urban design qualities—we hope to contribute to a better understanding of the way physical features of the built environment affect walking behavior.

Urban Design Qualities

The urban design literature identifies numerous perceptual qualities of the urban environment that may influence walking behavior.¹⁸⁻¹⁹ The visual assessment literature, which attempts to measure how individuals perceive their environments and better understand what individuals value in their environments, adds other potentially important qualities.²⁰⁻²⁵ The visual assessment literature goes beyond the boundaries of urban design to the fields of architecture, landscape architecture, park planning, environmental psychology, etc., as perceptual qualities of the environment figure prominently in these fields as well.

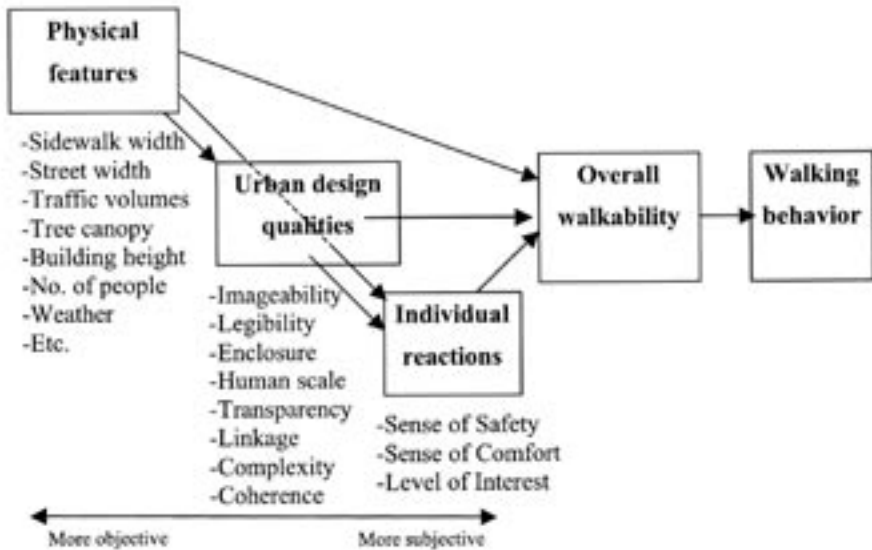


Figure 1—Conceptual framework

The long list of perceptual qualities described in the literature includes: adaptability, ambiguity, centrality, clarity, compatibility, comfort, complementarity, continuity, contrast, deflection, depth, distinctiveness, diversity, dominance, expectancy, focality, formality, identifiability, intelligibility, interest, intimacy, intricacy, meaning, mystery, naturalness, novelty, openness, ornateness, prospect, refuge, regularity, rhythm, richness, sensuousness, singularity, spaciousness, territoriality, texture, unity, upkeep, variety, visibility, and vividness. In examining the literature, we looked for qualities that were most frequently discussed and that empirical evidence showed are important to users of urban space. From this list, we combined similar concepts and identified distinct concepts. This effort led to the selection of eight urban design qualities for subsequent study: imageability, legibility, visual enclosure, human scale, transparency, linkage, complexity, and coherence. A ninth quality, tidiness, was added when a review of video clips, as described below, indicated that one potentially important dimension of scenes was not captured by the eight qualities selected initially. We developed definitions of these nine qualities based primarily on the urban design literature, and refined with the help of an expert panel (Table 1).

Methods

Our challenge was to move from these highly qualitative definitions of urban design qualities to operational definitions that capture the essence of each quality and can be measured with a degree of reliability across raters, including those without training in urban design. Our general approach was to measure these qualities for a sample of streetscapes and identify detailed physical features associated with each

Table 1 Urban Design Qualities

Imageability

Imageability is the quality of a place that makes it distinct, recognizable, and memorable. A place has high imageability when specific physical elements and their arrangement capture attention, evoke feelings, and create a lasting impression.

Legibility

Legibility refers to the ease with which the spatial structure of a place can be understood and navigated as a whole. The legibility of a place is improved by a street or pedestrian network that provides travelers with a sense of orientation and relative location and by physical elements that serve as reference points.

Enclosure

Enclosure refers to the degree to which streets and other public spaces are visually defined by buildings, walls, trees, and other elements. Spaces where the height of vertical elements is proportionally related to the width of the space between them have a room-like quality.

Human Scale

Human scale refers to a size, texture, and articulation of physical elements that match the size and proportions of humans and, equally important, correspond to the speed at which humans walk. Building details, pavement texture, street trees, and street furniture are all physical elements contributing to human scale.

Transparency

Transparency refers to the degree to which people can see or perceive what lies beyond the edge of a street or other public space and, more specifically, the degree to which people can see or perceive human activity beyond the edge. Physical elements that influence transparency include walls, windows, doors, fences, landscaping, and openings into midblock spaces.

Linkage

Linkage refers to physical and visual connections from building to street, building to building, space to space, or one side of the street to the other which tend to unify disparate elements. Tree lines, building projections, marked crossings all create linkage. Linkage can occur longitudinally along a street or laterally across a street.

Complexity

Complexity refers to the visual richness of a place. The complexity of a place depends on the variety of the physical environment, specifically the numbers and kinds of buildings, architectural diversity and ornamentation, landscape elements, street furniture, signage, and human activity.

Coherence

Coherence refers to a sense of visual order. The degree of coherence is influenced by consistency and complementarity in the scale, character, and arrangement of buildings, landscaping, street furniture, paving materials, and other physical elements.

Tidiness

Tidiness refers to the condition and cleanliness of a place. A place that is untidy has visible signs of decay and disorder; it is in obvious need of cleaning and repair. A place that is tidy is well maintained and shows little sign of wear and tear.

quality, a process we call a “visual assessment survey.” Specific steps included: 1) recruitment of a panel of urban design and planning experts; 2) creation of a library of video clips of streetscapes; 3) selection of video clips; 4) rating of urban design qualities of streetscapes by the expert panel; 5) measurement of physical features of streetscapes through a content analysis of video clips; 6) inter-rater reliability testing of physical measurements and urban design quality ratings; 7) statistical analysis of relationships between physical features and urban design quality ratings; 8) selection of qualities for operationalization, and 9) development of operational definitions and measurement protocols for urban design qualities based on statistical relationships. The details of these methods are described in more detail in the final report for this project,²⁶ which can be found at www.activelivingresearch.org.

1. Expert Panel

Because the concepts we sought to operationalize are not familiar to the average person, we could not simply ask a random sample of individuals to rate streetscapes as to their “legibility,” “transparency,” and so on. Instead, we chose to consult experts who employ these constructs in their work. We assembled a panel of 10 urban design and planning experts from professional practice as well as academia. The panel members helped us qualitatively define urban design qualities of streetscapes, rated different scenes with respect to these qualities, submitted to interviews as they assigned their ratings to provide the research team with qualitative insights, met to discuss ways of measuring urban design qualities, and reviewed and commented on the draft field survey manual that presented the measurement instrument in all its detail. We accepted the panelists’ ratings of the urban design qualities for the sample of streetscapes as valid by virtue of their specialized expertise.

2. Library of Video Clips

For practical reasons, we used video clips of streetscapes rather than field visits as the medium for rating urban design qualities. To ensure that reactions to street scenes were not biased by different filming techniques, we developed and used a consistent filming protocol. A great deal of experimentation and dialogue among the investigators went into the development of a protocol that would mimic the experience of pedestrians. Pedestrians are usually in motion, sway a bit as they walk, have peripheral vision, and tend to scan their environments. The protocol specified the starting point on a street block, walking speed, and panning motions; the distance covered and time length of the clips varied somewhat depending on actual walking and panning speeds but averaged between 1 and 1 ¼ min.

A shoot list for the video clips was generated according to a fractional factorial design described below. Working off the shoot list, more than 200 clips were filmed in dozens of cities across the US. Diversity of street scenes was ensured by the different regional locations of the investigators and the travels of the investigators on other business during the course of the study. In shooting clips, we focused on commercial streets in urban or “main street” settings—all places with sidewalks and other pedestrian amenities such as landscaping, pedestrian lighting, street furniture, and businesses or public spaces within view.

3. Sample Selection of Video Clips

Scenes were shot and ultimately selected for the visual assessment survey using a fractional factorial design. The purpose was to capture relevant combinations of the eight urban design qualities chosen for operationalization (tidiness was added later). Without variation across the qualities, it would have been impossible to tease out the contributions of individual physical features to urban design quality ratings. Factorial designs are common in experimental research in which the goal is to isolate the effect of each of multiple factors.

To choose our samples, one investigator and a research assistant rated clips as “high” or “low” with respect to the eight urban design qualities. From the larger set, 48 clips were selected that best matched the combinations of high/low values in a 2^{8-4} fractional factorial design. In contrast to a full factorial design, the fractional factorial design provided a way of choosing an appropriate subset from all 256 (2^8) possible combinations of the eight urban design qualities. The 2^{8-4} sample allowed us to capture the main effects of each urban design quality on overall walkability (as described below), plus two-factor interaction effects.²⁷

Urban design qualities tend to co-vary (that is, appear in certain combinations of high and low values), making perfect matches unlikely starting with any practically sized set of clips. Some of the clips matched high/low patterns perfectly, while others matched on only seven, six, or even five of the qualities, rather than all eight. Although we were not able to exactly match the fractional factorial design in all cases, following the design as closely as possible resulted in the selection of clips that were distinctly different. Where ratings for two or more clips matched factorial designs equally well, clips were selected to maximize geographic diversity.

4. Ratings of Urban Design Qualities by Expert Panel

In addition to helping us refine the definitions of urban design qualities presented earlier, the expert panel rated each of the 48 selected video clips with respect to each of nine urban design qualities. The first wave of the visual assessment survey (32 clips) was conducted remotely. The sample of video clips was recorded in random order onto DVDs, and the DVDs were distributed to expert panel members. A telephone survey was then conducted between each panel member and a research team member following a standardized protocol. Panelists first viewed five specific clips that illustrated the full range of values represented within the sample for each urban design quality. The panelist then viewed each clip, concurrently with the research team member, and assigned a score for each quality on a scale from 1 (low) to 5 (high). Definitions of the qualities and descriptions of low, medium, and high values of each were provided to the panelist for use during the survey. Note that the ratings did not reflect an assessment of whether a particular quality was good or bad for walkability. The research team member recorded scores on a standardized form and audio-taped the comments of the panelist as he or she viewed the video clips.

The remaining 16 video clips were viewed (in random order) and rated in a face-to-face meeting of the expert panel. Four panelists who could not attend the meeting were sent DVDs and subsequently surveyed by phone following the same procedure as for the first 32 clips. The audio-tapes of the interviews and face-to-face meeting were analyzed to assist the research team in identifying physical features of scenes that should be measured in the subsequent content analysis.

In addition to rating urban design qualities for each clip, the panelists also provided an overall assessment of walkability, again on a scale from 1 (low) to 5 (high). The relationships between ratings of urban design qualities and walkability scores were used as one of five criteria for deciding which urban design qualities should be operationalized, as described below. It is quite possible that the walkability ratings of the panelists differed from the ratings that a sample of average pedestrians would have given. Because we were trying to operationalize design concepts, not assess public preferences, we relied on the walkability ratings of the panelists for this limited purpose.

5. Measurement of Physical Features by Research Team

To measure physical features of streetscapes, all 48 video clips were analyzed for content. All told, more than 100 features were measured for each video clip. The process typically required more than an hour for each video clip, and much more for the more complex scenes. Detailed operational rules for measuring each physical feature were developed to ensure consistency.

The physical features measured in this manner were derived from the urban design literature, from earlier visual assessment studies, and most importantly, from interviews with the expert panel. As panelists rated scenes, they also commented on the physical features that caused ratings to be high or low with respect to each urban design quality. Interviews, which had been taped, were reviewed to identify promising features.

One of the investigators and a research assistant measured each physical feature for all 48 clips using a process that might best be described as one of forced consensus. The two independently measured each feature, discussed differences, and finally reached agreement on a single value for each physical feature of each video clip. To assess inter-rater reliability of measured physical features, a random sample of video clips was assigned to three other members of the research team. The sample consisted of 12 clips in all, or four per team member. Sample size was limited by the time required to evaluate more than 100 features of each clip.

6. Inter-Rater Reliability Testing of Physical Measurements and Urban Design Quality Ratings

Various statistical techniques may be used to assess inter-rater reliability in studies like this, where multiple individuals independently rate the same set of cases. We used intra-class correlation coefficients (ICCs), representing the ratio of between-groups variance to total variance.²⁸ ICCs were used to evaluate inter-rater reliability for both the ratings of urban design qualities by the expert panel and the measures of physical features by the research team.

7. Statistical Analysis of Relationships Between Physical Features and Urban Design Quality Ratings

Multivariate statistical methods were used to model urban design ratings in terms of measurable physical features of scenes. Models were specified based on hypothesized relationships between urban design qualities and specific physical features.

The hypothesized relationships were partly based on common sense, partly a reflection of the urban design literature, and partly a product of the interviews with the expert panelists. To keep model building from becoming a data mining exercise, a matrix of hypothesized relationships was created and only the features plausibly linked to urban design qualities were actually tested for predictive power.

We estimated cross-classified random effects models in this analysis. When an outcome varies systematically in two dimensions and random effects are present, the resulting data structure is best represented by a cross-classified random effects model.²⁹ The outcome variable in this analysis was the urban design quality rating assigned by an individual panelist to an individual streetscape. Ratings varied from scene to scene owing to different qualities of the scene itself. Ratings also varied from viewer to viewer owing to differences in judgment; some viewers were more generous in their grading than others. Finally, ratings varied due to unique interactions between scenes and viewers. A particular scene may have evoked a particularly positive or negative reaction in a particular viewer. We viewed such unique reactions as measurement errors.

The more interesting source of variation in scores is that associated with scenes. Indeed, the purpose of this study is to identify the physical features of scenes that give rise to high or low ratings of urban design qualities. In statistical parlance, the “scene effect” gives rise to “scene variance.” While not of much interest, variation also occurs across panelists and must be accounted for. Again in statistical parlance, the “viewer effect” gives rise to “viewer variance.” The unique reactions of individual panelists, and the random variations in their scoring across scenes, produce “measurement error variance.”

In order to bring into focus the interesting variation, that is, the variation across streetscapes, it helps statistically to separate the scene variance from viewer variance and measurement error variance. Doing so, we are able to eliminate viewer effects when evaluating the power of physical features to predict streetscape ratings. If we had simply used the average ratings of scenes as the outcome variable, and the physical features of scenes as explanatory variables, the effect of scene variance might have been confounded by the effect of viewer variance. Cross-classified random effects models were estimated using HLM 6.0 software, a statistical package developed by Raudenbush, Bryk, and Congdon.³⁰

The estimated models included characteristics of viewers and scenes:

$$\text{actual score} = \text{predicted score} + \text{measurement error}$$

and

$$\text{predicted score} = \text{constant} + \text{viewer random effect} + \text{scene random effect} + a * \text{viewer variables} + b * \text{scene variables}$$

where the *viewer random effect* is the portion of the viewer effect left unexplained by viewer characteristics, the *scene random effect* is the portion of the scene effect left unexplained by scene characteristics, *viewer variables* is the vector of relevant viewer characteristics, *a* is the vector of associated coefficients, *scene variables* is the vector of relevant scene characteristics, and *b* is the vector of associated coefficients. These variables capture the “fixed effects” of viewers and scenes on urban design ratings. The equations for the predicted scores were used to operationalize

the measurement of each urban design quality in terms of significant physical features.

8. Criteria for Selecting Urban Design Qualities to Operationalize

Through the course of the study, it became clear that not all urban design qualities could be defined operationally. Some are clearly more amenable to measurement than are others. To decide which urban design qualities would be defined operationally in the field survey instrument, five criteria were established:

- The urban design quality was rated by the expert panel with at least a moderate degree of inter-rater reliability ($ICC \geq 0.4$), following the criteria suggested by Landis and Koch.³¹
- The total variance in ratings of the urban design quality was explained to at least a moderate degree by measurable physical features of scenes (explained portion ≥ 0.3).
- The portion of total variance in ratings attributable to scenes was explained to a substantial degree by measurable physical features of scenes (explained portion ≥ 0.6).
- All physical features related to ratings of a particular urban design quality were measured by the research team with at least a moderate degree of inter-rater reliability ($ICC \geq 0.4$), excluding those for which ICC values could not be computed because of insufficient variation in that quality across sampled scenes.
- The urban design quality as judged by the expert panel had a statistically significant relationship to overall walkability ratings by the expert panel ($P \leq 0.05$).

9. Development and Testing of Measurement Protocols and Field Manual

Measurement protocols were developed for the selected urban design qualities (described below) and incorporated into a field manual. Although the protocols were developed based on video clips, the field manual is of course intended for use in the field. As a result, it was important to validate measurements made in the field. The research team went into the field and completed measurements for a sample of 16 of the video clip sites. Field observations and video clips were compared using a variety of tests. Based on this analysis, several refinements to the field manual were made to improve the reliability of field measurements.

The field manual was then tested for inter-rater reliability using students from the investigators' academic institutions as lay observers. The observers were first trained in the classroom in the use of the field manual. A sample of video clips from the original visual assessment survey was used as the training medium in the classroom, where students could compare their measurements of physical features

with the measures made by the research team. After classroom training, students were taken into the field for observations. Students completed measurements of the physical features used to compute the urban design qualities. Scoring sheets were later analyzed for inter-rater reliability. Final edits to the field manual were made based on the classroom and field experience.³²

Results

Criterion 1: Inter-rater Reliability of Scene Ratings

In this study, the members of the expert panel independently rated each of 48 clips with respect to the nine urban design qualities, and values were compared for inter-rater reliability (see Table 2). According to the ICC values, most urban design qualities demonstrated moderate inter-rater reliability among panelists ($0.6 > \text{ICCs} \geq 0.4$); the exceptions—linkage, coherence, and legibility—showed only fair reliability ($0.4 > \text{ICCs} \geq 0.2$) and failed to meet the criterion of 0.4. For purposes of comparison, Cronbach's alpha is also reported for these ratings.

Table 2 Inter-rater Reliability for Ratings of Urban Design Qualities

	Intra-class Correlation Coefficient	95% CI of ICC	Cronbach's alpha
imageability	0.494	0.385–0.618	0.930
legibility	0.380	0.276–0.509	0.895
enclosure	0.584	0.478–0.697	0.945
human scale	0.508	0.399–0.630	0.928
transparency	0.499	0.390–0.622	0.926
linkage	0.344	0.169–0.621	0.896
complexity	0.508	0.398–0.632	0.926
coherence	0.374	0.271–0.504	0.880
tidiness	0.421	0.314–0.550	0.915
<i>N</i>	48		

Criteria 2 and 3: Relationships Between Physical Features and Urban Design Qualities

Cross-classified random effects models partition total variance in ratings into portions attributable to scenes, viewers, and measurement error. For each urban design quality, Table 3 shows the total variance in ratings and the portions attributable to each source. For all urban design qualities, there was more variance across scenes than across viewers. The fuzzier constructs such as legibility and linkage have the highest proportions attributable to viewer judgment and measurement error.

As an example, for the urban design quality of imageability, the scene variance was 0.67, the viewer variance was 0.16, and the measurement error variance was 0.50. The total variance was thus split in the following proportions:

- 50% scene variance,
- 12% viewer variance, and
- 38% measurement error variance.

In estimating the models for urban design qualities, many combinations of viewer and scene variables were tested. The only available variables characterizing viewers—urban designer or not (1 or 0 dummy) and new urbanist or not (1 or 0 dummy)—proved to have no explanatory power in most analyses. That is to say, neither variable was significant at the 0.10 probability level, except in the model of human scale, in which urban designers assigned marginally higher ratings to scenes than did non-designers. This is consistent with earlier visual assessment studies, which report common environmental preferences across professions.³³

By contrast, many of the variables characterizing scenes proved significant individually and in combination with each other. This again is consistent with the visual assessment literature. The models that reduced the unexplained variance of scores to the greatest degree, and for which all variables had the expected signs and were significant at the 0.10 level or beyond, are presented in the final report for this project.²⁶ In all, 37 physical features proved significant in one or more

Table 3 Variance in Ratings by Source for Each Urban Design Quality (percent of total variance in parentheses)

	Scene Variance	Viewer Variance	Measurement Error	Total Variance
imageability	0.67 (50)	0.16 (12)	0.50 (38)	1.33
legibility	0.46 (39)	0.17 (14)	0.55 (47)	1.18
enclosure	0.83 (59)	0.10 (7)	0.48 (34)	1.41
human scale	0.68 (53)	0.11 (8)	0.50 (39)	1.29
transparency	0.77 (51)	0.13 (8)	0.62 (41)	1.52
linkage	0.51 (34)	0.26 (17)	0.74 (39)	1.51
complexity	0.6 (52)	0.09 (8)	0.47 (40)	1.16
coherence	0.45 (38)	0.11 (9)	0.62 (53)	1.18
tidiness	0.46 (43)	0.17 (16)	0.43 (41)	1.06

models. Six features were significant in two models: long sight lines, number of buildings with identifiers, proportion first floor façade with windows, proportion active uses, proportion street wall—same side, and number of pieces of public art. Two features were significant in three models: number of pedestrians and presence of outdoor dining.

As an example, for imageability, the estimated model with physical features left the measurement error variance unchanged at 0.50, reduced the unexplained viewer variance only slightly from 0.16 to 0.15, but reduced the unexplained scene variance substantially, from 0.67 to 0.19. Altogether, 72% of the variation across scenes and 37% of the overall variation in imageability scores (including variation across viewers and measurement errors) were explained by the significant physical features (see Table 4). These results met both selection criteria: 60% or more of the scene variance and 30% or more of the overall variance were explained by physical features in this model.

Criterion 4: Inter-rater Reliability of Measured Features

The inter-rater reliability test was conducted on a subset of video clips. Two members of the research team jointly measured physical features for 12 clips, while other members of the team independently measured them for each clip. For most features, there was almost perfect agreement ($ICCs \geq 0.8$) or substantial agreement ($0.8 > ICCs \geq 0.6$) between ratings. It is relatively easy to count objects and estimate widths; ICCs were high for these types of features. However, several features had low or even negative ICC values. Of these, features such as the number of landscape elements could probably be rated more consistently with better operational definitions. Other features, such as landscape condition, involve a high degree of judgment and might require training and/or photographic examples to achieve reasonable inter-rater reliability.

Table 4 Best-Fit Imageability Model

Variable	Coefficient	t-statistic	P-value
constant	2.516		
courtyards/plazas/parks (#)	0.393	3.58	0.001
major landscape features (#)	0.735	2.00	0.046
proportion of historic buildings	0.948	4.16	0.000
buildings with identifiers (#)	0.115	1.80	0.072
buildings with non-rectangular silhouettes (#)	0.0745	1.95	0.052
pedestrians (#)	0.0271	4.73	0.000
noise level (rating)	-0.195	-2.11	0.035
outdoor dining (y/n)	0.703	3.97	0.000
Proportion of scene variance explained	0.72		
Proportion of total variance explained	0.37		

The ICCs for each of the physical features significant in the final models of urban design qualities (Step 7) were examined. For example, for the model of imageability, all of the significant scene variables had acceptable levels of inter-rater reliability with intraclass correlation coefficients of 0.40 or above. Major landscape features had insufficient variance across the sample to compute inter-rater reliability.

Criterion 5: Correlates of Overall Walkability

Using mean values for the 48 video clips, we found that overall walkability (as rated by the expert panel) was directly and significantly related to each urban design quality individually. The analysis was complicated, however, by the fact that eight of the nine qualities (the exception being tidiness) were collinear. Tolerance values were unacceptably low when all variables were included in a regression at once.

Linkage and legibility appeared to be largely functions of the other urban design qualities, so they were dropped from further consideration. Of the remaining variables, human scale had the strongest relationship to overall walkability almost regardless of what combination of variables was tested. Tidiness, and to a lesser extent, transparency, enclosure, and imageability, were somewhat independent of human scale, proved significant at the 0.10 level in most model runs, and improved the explanatory power of the model (the adjusted *R*-squared). Coherence was ultimately dropped because it proved insignificant and reduced the explanatory power of the model. Complexity was ultimately dropped even though significant in some model runs, because it altered relationships between other variables and overall walkability, and because it had a low tolerance value itself.

The best-fit equation is presented in Table 5. Urban design qualities explain more than 95% of the variation in mean overall walkability, according to our expert panel. All qualities are directly related to overall walkability, and all are significant at conventional levels except tidiness, which falls just below the 0.10 level. Based on their *t*-statistics, human scale ranks first in significance as a determinant of overall walkability, imageability second, enclosure third, transparency fourth, and tidiness a distant fifth.

Table 5 Regression Model for Overall Walkability

Variable	Coefficient	Standardized coefficient	<i>t</i> -statistic	<i>P</i> -value
constant	-0.226		-1.503	0.140
human scale	0.411	0.420	5.814	0.000
transparency	0.137	0.149	2.366	0.023
tidiness	0.070	0.059	1.598	0.117
enclosure	0.140	0.157	2.504	0.016
imageability	0.307	0.310	5.153	0.000
<i>N</i>	48			
<i>R</i> -square	0.959			
Adjusted <i>R</i> -square	0.954			

Performance on Selection Criteria

The urban design qualities of *imageability*, *enclosure*, *human scale*, and *transparency* met all five criteria for operationalization (see Table 6). The qualities of *legibility*, *linkage*, and *coherence* met only one of five criteria, and were dropped from further consideration. The qualities of *complexity* and *tidiness* met three of five criteria. The research team decided to include these two urban design qualities, along with the first four qualities, in field validation tests. Tidiness was ultimately dropped when field measurements disagreed fundamentally with lab measurements based on video clips. The field manual thus includes protocols for measuring five urban design qualities: *imageability*, *enclosure*, *human scale*, *transparency*, and *complexity*.

Conclusions

This study has demonstrated that qualitative urban design qualities can be quantified. The power of our approach is that it used relatively simple and objective features of the physical environment to measure abstract urban design qualities. The protocols for measuring urban design qualities can be used by lay observers without any training in urban design. The resulting measures could be useful to urban designers interested in pursuing a more quantitative approach to their profes-

Table 6 Performance of Urban Design Qualities Relative to Selection Criteria

	Inter-rater Reliability of Rating of Quality (ICC)	Portion of Scene Variance/Total Variance Explained by Best-Fit Models	Inter-rater Reliability of Significant Variables (number with ICC > 0.4)	Relationship to Walkability in Best-Fit Model (<i>P</i> -value)	Criteria Met
Imageability	0.494	0.72/0.37	7 of 7 (1 missing)	0.000	5 of 5
Legibility	0.380	0.54/0.21	5 of 5 (1 missing)	--	1 of 5
Enclosure	0.584	0.72/0.43	5 of 5	0.016	5 of 5
Human scale	0.508	0.62/0.35	7 of 7	0.000	5 of 5
Transparency	0.499	0.62/0.32	3 of 3	0.023	5 of 5
Linkage	0.344	0.61/0.21	4 of 5	--	1 of 5
Complexity	0.508	0.73/0.38	5 of 6	--	3 of 5
Coherence	0.374	0.67/0.25	3 of 4	--	1 of 5
Tidiness	0.421	0.70/0.30	2 of 3 (1 missing)	0.117	3 of 5

sion. The measures should also be useful to researchers interested in understanding how environmental qualities, as well as patterns and combinations of particular qualities, affect people's perceptions of streetscapes and their willingness to walk and otherwise be active in them. Methodology similar to the one followed in this study could be used to develop measures for perceptual qualities of other physical settings, such as residential streets, plazas, parks, and trails.

Limitations of this study include the focus on urban environments (as opposed to the suburbs or rural areas), the reliance on video clips (rather than in-field ratings), the reliance on experts (rather than end users of streets) to rate walkability, and the moderate degrees of inter-rater reliability achieved for ratings of urban design qualities, measures of physical features, and lab vs. in-field measurements. Although we were unable to operationalize four of the identified urban design qualities (legibility, linkage, coherence, and tidiness), we believe these qualities may be important in explaining walkability and are worthy of further attempts at quantification.

The field instrument, training materials, and final report are now available on-line at www.activelivingresearch.org. The field instrument includes: qualitative definitions of urban design qualities; explanations and photographic illustrations of physical features relating to urban design qualities; procedures for field observation and data collection; and scoring procedures for translating objectively measured physical features into urban design quality scores (see Figure 2). A DVD of video clips and sample scoring sheets are available as part of the training package.

Acknowledgments

The authors gratefully acknowledge the funding support of the Active Living Research Program of The Robert Wood Johnson Foundation. Our panel of leading experts consisted of: Victor Dover, Dover, Kohl & Partners Town Planning; Geoffrey Ferrell, Geoffrey Ferrell Associates; Mark Francis, University of California at Davis; Michael Kwartler, Environmental Simulation Center; Rob Lane, Regional Plan Association of New York and New Jersey; Anne Vernez Moudon, University of Washington; Tony Nelessen, A. Nelessen Associates; John Peponis, Georgia Institute of Technology; Michael Southworth, University of California at Berkeley; and Dan Stokols, University of California at Irvine.

References

1. Committee on Physical Activity, Health, Transportation, and Land Use. Does the Built Environment Influence Physical Activity: Examining the Evidence. Washington: Transportation Research Board, 2005.
2. Brownson RC, Baker EA, Houseman RA, Brennan LK, Bacak SJ. Environmental and policy determinants of physical activity in the United States. *Am J Public Health*. 2001; 91:1995-2003.
3. Ewing R, Cervero R. Travel and the built environment. *Transp Res Record*. 2001;1780: 87-114.
4. Saelens BE, Sallis JF, Frank LD. Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Ann Behav Med*. 2003; 25(2):80-91.
5. Owen N, Humpel N, Leslie E, Bauman A, Sallis J. Understanding environmental influences on walking review and research agenda *Am J Prev Med*. 2004;27(1):67-76..

6. Handy S. Critical Assessment of the Literature on the Relationships Among Transportation, Land Use, and Physical Activity. Department of Environmental Science and Policy, University of California, Davis. Prepared for the Committee on Physical Activity, Health, Transportation, and Land Use. Washington: Transportation Research Board; 2005.
7. Ewing R. Can the physical environment determine physical activity levels? *Exerc Sport Sci Rev*. 2005; 33(2): 69-75.
8. Moudon AV, Lee C. Walking and bicycling: an evaluation of environmental audit instruments. *Am J Health Promot*. 2003; 18(1): 21-37.
9. Lynch K. *The image of the city*. Cambridge, MA: Joint Center for Urban Studies; 1960.
10. Alexander C, Ishikawa S, Silverstein M. *A pattern language—towns buildings construction*. New York: Oxford University Press; 1977.
11. Hedman R. *Fundamentals of urban design*. Chicago: American Planning Association; 1984.
12. Gehl J. *Life Between buildings—using public space*. New York: Van Nostand Reinhold; 1987.
13. Whyte WH. *City—rediscovering the center*. New York: Doubleday; 1988.
14. Jacobs A. *Great streets*. Cambridge, MA: MIT Press; 1993.
15. Pikora T, Bull F, Jamrozik K, Knuiaman M, Giles-Corti B, Donovan R. Developing a reliable audit instrument to measure the physical environment for physical activity. *Am J Prev Med*. Oct 2002;23(3):187.
16. Moudon AV, Lee C. Walking and bicycling: an evaluation of environmental audit instruments. *Am J Health Promot*. Sep-Oct 2003;18(1):21-37.
17. Brownson R, Hoehner C, Brennan L, Cook R, Elliott M, McMullen K. Reliability of two instruments for auditing the environment for physical activity. *J Phys Activity Health*. 2004;1:189-207.
18. Handy S. Regional versus Local Accessibility: Variations in Suburban Form and the Effects on Nonwork Travel. Unpublished dissertation, University of California at Berkeley, November; 1992.
19. Ewing R. Pedestrian- and Transit-Friendly Design. Tallahassee: Florida Dept of Transportation; 1996.
20. Kaplan R, Kaplan S. *The experience of nature—a psychological perspective*. New York: Cambridge University Press; 1989.
21. Sanoff H. *Visual research methods in design*. New York: Van Nostrand Reinhold; 1991.
22. Nasar JL. *The evaluative image of the city*. Thousand Oaks, CA: Sage; 1998.
23. Stamps AE. *Psychology and the aesthetics of the built environment*. Netherlands: Kluwer Publishers; 2000.
24. Ewing R. Asking transit users about transit-oriented design. *Transp Res Record*. 2000;1735:19-24.
25. Ewing R, King M, Raudenbush S, Clemente O. Turning highways into main streets: two innovations in planning methodology. *J Am Planning Assoc*. 2005; 71:269-282.
26. Ewing R, Clemente O, Handy S, Brownson R, Winston E. Identifying and Measuring Urban Design Qualities Related to Walkability—Final Report. Princeton, NJ: Robert Wood Johnson Foundation; 2005.
27. NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, February 12, 2004.
28. Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. New York: John Wiley & Sons; 1981.
29. Raudenbush SW, Bryk AS. *Hierarchical linear models: applications and data analysis methods*. 2nd ed. Thousand Oaks, CA: Sage; 2002.
30. Raudenbush SW, Bryk A, Congdon R. HLM 6: Hierarchical Linear & Nonlinear Modeling. Chicago: Scientific Software International; 2004.

31. Landis J, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-74.
32. Clemente O, Ewing R, Handy S, Brownson R, Winston E. Measuring Urban Design Qualities—An Illustrated Field Manual. Princeton, NJ: Robert Wood Johnson Foundation; 2005.
33. Stamps AE. Demographic effects in environmental aesthetics: A meta-analysis. *J Planning Literature*. 1999;14(2): 155-175.

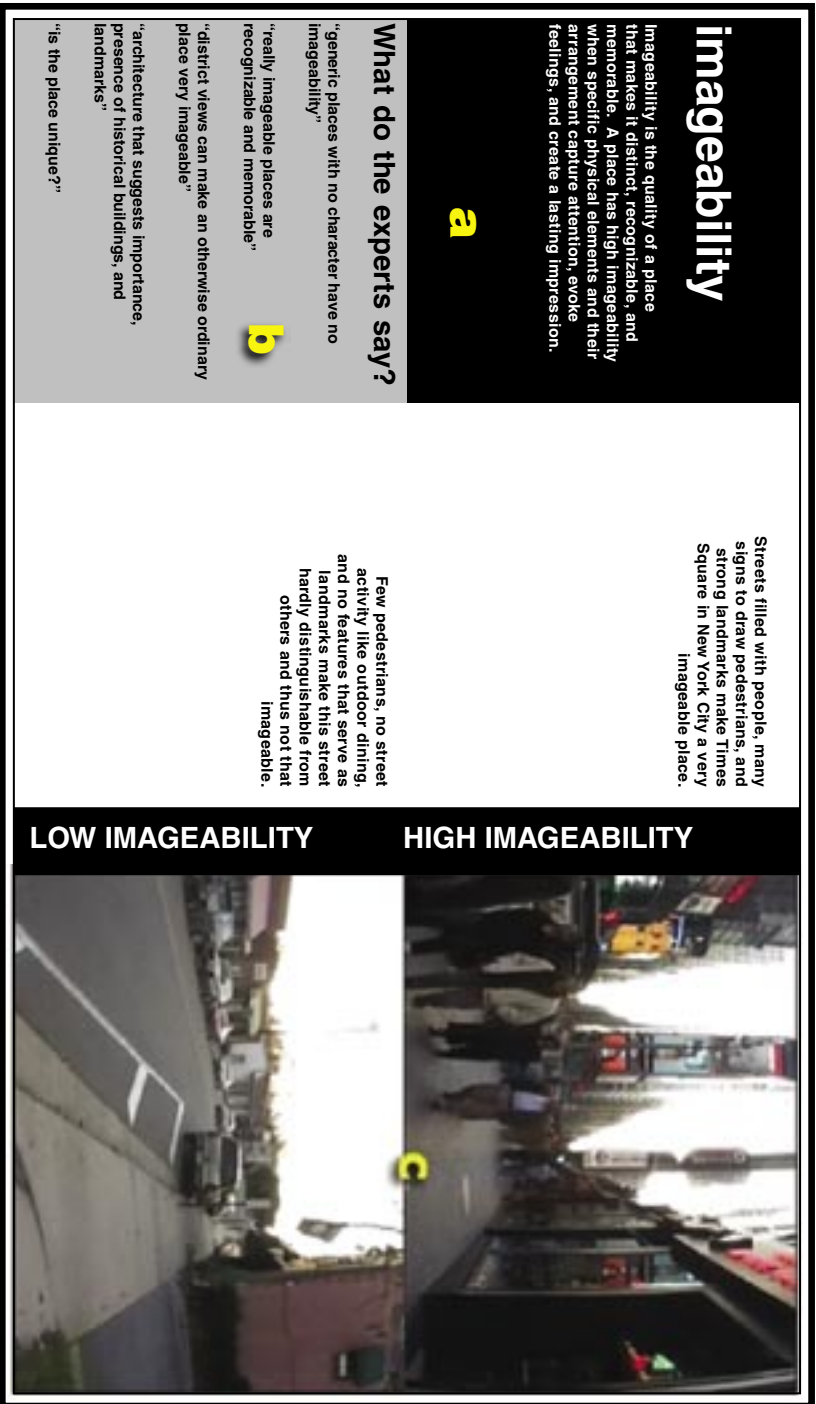


Figure 2—Introduction to one urban design quality, imageability