How Travel Mode Influences One's Perception of the Built Environment: Testing a Neighborhood Audit

David Bergmark

Abstract

Neighborhood audits have emerged as an integral tool for gathering objective, street level data on the built environment. However, in order to trust and make use of their results, researchers must understand which factors influence audit results. In this study, researchers investigated for the first time the influence which auditors' travel mode has on audit results. Though most neighborhood audits are conducted by foot, a number of audits have chosen to use "windshield surveys" instead. This study was designed to determine if audit results vary according to auditors' travel mode and if they do, which environmental features are perceived differently.

To answer these questions, the PIN III Neighborhood Audit tool was used to re-evaluate 79 audit segments on foot which had been previously rated a year before by car. All re-evaluated audit segments were designated as 'urban' and located with Durham or Orange County, NC. A pair of trained auditors conducted this test, with one of the two auditors participating in both audit sessions. Audit results from these two sessions were then compared, and percent agreement and kappa scores were generated in order to determine where significant differences occurred. The results of this analysis showed that while most environmental features included in this audit were not perceived differently, the audit results of 11 questions showed significant signs of changing due to travel mode. Alternative explanations for observed variations in audit results, such as the year long time delay between audit sessions, were tested and ruled out as likely factors.

Though the degree to which travel mode influences audit results will naturally vary depending on the nature of the neighborhood audit, this study's findings demonstrate that travel mode is an important factor that must be taken into account during future study design and data analysis.

Table of Contents

Introduction

There has been a long standing interest in the connection between the built environment and human behavior. The importance of this topic has grown over time and spread across disciplines, becoming a focal point of much research in the fields of psychology, sociology, design, geography, urban planning, and public health, amongst others. Though the central focus of this connection tends to vary between disciplines, rising concerns over declining rates of physical activity and corresponding increases in health risks have led researchers in the fields of urban planning, public health, and design to place greater emphasis on determining how various environmental factors specifically influence physical activity. The existence of this connection may seem intrinsically obvious, yet only more recently have researchers attempted to quantify the exact relationship between these two variables, moving beyond simple macro-scale data analysis to more micro-scale, direct assessment techniques (Clifton et al, 2006; Day et al, 2006; Zenk et al, 2007; Pikora et al, 2002; Hoehner et al, 2006).

Neighborhood audits have emerged as one of the primary tools to gain street level, objective knowledge on this topic. However, in order to trust and make use of the results of these audits, researchers must understand which factors influence audit results. Though many factors have been previously tested, one factor which has yet to be tested is the influence which neighborhood auditors' travel mode has on audit results.

This study seeks to answer this prevailing question, by re-evaluating 79 road segments which were originally audited as part of the PIN III neighborhood audit. The same audit tool was used, with the only significant difference being that 2007 auditors performed the audit on foot, rather than by car. Specifically, this study is meant to answer the following questions:

1) Do audit results differ when raters conduct a neighborhood audit by car versus on foot?
2) If audit results do change according to the travel mode employed, which environmental features are perceived differently and in what manner do these differences occur?
3) More generally, what can observed differences in audit results tell us about how one perceives the built environment while using different travel modes?

In order to answer these questions, data from each audit period will be compared, with researchers looking for significant differences which might point to specific perceptions of the built environment which are most affected by the observer's travel mode. It is my hope that the results of this study will help inform the design and administration of future neighborhood audits, as well as provide more general knowledge on the influence which travel mode has on human perception.

This paper begins with a background section providing an overview of how the topic of the built environment and physical activity has evolved over time within the fields of urban planning, design, and public health. Contained within this section is a literature review of previous neighborhood audits, with emphasis placed on audit structure, administration, and results. The background section is intended to explain how neighborhood audits were formed, why they are important, and why we should be concerned with the reliability of audit results. The next section discusses the methodology used in this study, with detailed information on the study area, time frame, audit tool contents, and data collection procedure. The third section of the paper examines the actual data analysis and results of the neighborhood audit performed. Finally, a full discussion of the meaning, uses, and limitations of the study results is provided.

Background

Audit tools have served a wide range of purposes over time. Traditionally, many audit tools were very limited in their scope, evaluating only the condition of infrastructure (Handy et al, 2002) in order to identify areas in need of maintenance or repair. More recently, the purpose of many audit tools has shifted in response to disturbing trends in health and physical activity, with a number of audits focusing on features such as pathways (Craythron, 1997), parks (Bedimo-Rung et al, 2005), bicycle compatibility (Craythron, 1997) or proximity of recreational facilities (Troped et al, 2001; Bauman et al, 1999; Sallis et al, 1990). However, despite many attempts of policy makers, health officials, and educators to reverse the negative trend, reported physical activity rates show a significant decline (Brownson et al, 2005) and obesity figures have dramatically escalated (Mokdad et al, 2001). According to one national study, sixty percent of adults reported little or no leisure time physical activity (Caspersen et al, 1986-1990).

Given that walking is reported as the most common physical activity (Siegel et al, 1995), determining which environmental factors support and deter walking has become a central question in the field of public health and urban planning. Researchers concerned with this topic have focused on different types of walking activity, investigating how environmental factors influence each walking type. In the article "Understanding environmental influences on walking", Owen et al categorizes the different types of walking trips identified in past research papers as follows (Owens et al, 2004):

1. Walking for exercise or recreation
2. Total walking
3. Walking to and from specific destinations
4. Walking for pleasure/social walking

If our concern is physical health, then total walking is the most important measure. Realizing the importance which walking has for public health, researchers have utilized audit tools in order to assess the overall "walkability" of an area. The increasing importance of this issue is evident by the mere fact that the word 'walkability' has become so widely used and accepted as a legitimate term. Creating a highly walkable environment has become an important component of many modern planning initiatives, such as the New Urbanism movement, which seeks to create communities that are designed to support alternative transportation modes. Despite the pervasiveness of this term, researchers are still attempting to identify exactly which features make an area "walkable". As a result, researcher design can vary significantly from one study to another, both in terms of the environmental characteristics assessed and the data sources used to obtain information on those features.

There are three typical scales of research used to assess environmental features which influence physical activity: region/community, neighborhood, and street level. Each scale has its own advantages and disadvantages. Many earlier research efforts relied heavily on national data sources to obtain macro-scale data on a region as a whole. While national data sources have greater data consistency compared to local sources, it is often difficult to accurately generalize that data to the neighborhood or street level scale. Examples of national data sources used in studies evaluating neighborhood characteristics include the US Census of Population and Housing, American Housing Survey, Census Transportation Planning Package, National Resources Inventory, and Census TIGER/line files (Handy et al, 2002). These data sources are often used to create data on density and intensity, the mix of land uses, and street network structure (ibid). However, the usefulness of national data sources often ends there. Data sources covering a metro scale typically contain information on only a few characteristics of the built environment, requiring researchers to use supplemental data in order to obtain more detailed information.

Obtaining useful travel statistics as a means of assessing walkability represents a significant drawback to using national data sources as well. While disaggregate data can be obtained on individual or household travel data for most metro areas, little data is available on pedestrian travel, which presents a significant limitation in this field of research. Researchers have attempted to respond to this limitation by supplementing their data derived from national data sources with local data sources and self-reported information on travel activity. Local data sources relevant to assessing the built environment might include property tax records, building permit records, aerial photos, and street and sidewalk inventories (Handy et al, 2002).

As previously mentioned, many past and current studies have used self-reporting to fill in the information gaps present in national and local data sources, as well as to gather more detailed, disaggregate data on pedestrian travel and individuals' perceptions (Hoehner et al, 2006; Zenk et al, 2007; Kirtland et al, 2003; Siegel et al, 1995). In fact, a number of studies using self-reported survey data have found a strong link between the built environment and physical activity (Brownson et al, 1995-2003; Giles-Corti et al, 2002). In a study by Berigan and Troiano, the authors used the age of residents' home as a proxy measure for a dense, mixed-use, urban form with high street connectivity and evaluated how the age of homes corresponded with reported physical activity levels. Using this approach, respondents in older homes were found to walk more (Berrigan et al, 2002). In another study, Ewing et al developed and used a "sprawl indices" to evaluate the built environment and found high "sprawl" scores to be negatively associated with reported minutes walked (Ewing et al, 2003).

Yet, the accuracy of self-reported survey data has been repeatedly questioned (Sallis et al, 1990; Golledge et al, 1997; Lloyd et al, 1997; Pederson, 1997). Survey data relies on self-reporting, which has been shown to have limited reliability, compared to more objective measures. This limitation was revealed in a study by Sallis, Melbourne, and Hofstetter, which showed correlation between objective environmental measures and physical activity, but no evident relation between those same factors when using self-reporting (Sallis et al, 1990).

One method of compensating for the limitations of self-reporting that has become a standard practice when conducting research on the built environment and physical activity is to use GIS technology to augment survey data with more objective measures (Kirtland et al, 2003; Clifton et al, 2006; UNC CPC - PIN III study, 2008; Hoehner et al, 2005). A common application of GIS technology is to use it to define boundaries of neighborhoods and communities. Once these boundaries are established, specific survey questions can be tailored to the different geographic scales established. However, this approach is still subject to the limitations of self-reporting. As Kirtland et. al point out in their article, “Environmental Measures of Physical Activity Supports: Perception Versus Reality”, “the ideal distance from one’s home for the recall of environmental supports is unknown” (Kirtland et al, 2003). The defined “neighborhood” may poorly match what residents perceive as their neighborhood. Furthermore, the ideal distance for recall may vary from person to person, based on the size of their urban environment (Golledge et al, 1997; Pedersen, 1997) or other cultural, psychological, or behavior factors (Kirtland et al, 2003).

Individual’s relative experience has been shown to be highly influential on their perceptions as well. In a study by Wentworth et al., researchers found that people’s perception of park safety was strongly linked to whether or not they already used park facilities (Wentworth et al, 1976). Though the causality of this relationship is uncertain, it serves as but one example of the multitude of factors which influence people’s perceptions. Thus, there is an inherent risk in relying upon individuals’ perceptions for study data.

In light of the limitations of macro-scale data analysis and self-reporting results, direct assessment of the built environment by trained observers has become a popular method of obtaining reliable, detailed, and objective data on the built-environment. By using trained observers, researchers are able to improve the consistency of observer responses by establishing rules for the identification and classification of environmental features. The use of direct assessment in neighborhood audits typically involves a pair of trained auditors, who evaluate a defined number of street-level environmental features. In most cases, audit teams conduct the neighborhood audit on foot (Clifton et al., 2006; Day et al., 2006; Zenk et al., 2007; Pikora et al., 2002; Hoehner et al., 2006). However, some audits, such as the PIN III neighborhood audit tool, conduct a "windshield survey" instead, in which auditors evaluate features of the built environment while driving slowly down a street segment. Windshield surveys are also commonly used by municipal agencies to evaluate regions such as downtown neighborhood districts or community redevelopment areas.

The advantages of auditing by car include reduced travel time between audit segments, increased safety from crime, and the ability to audit in adverse weather. Of course, as this study seeks to discover, there may be significant consequences of not walking the actual segments while evaluating the built environment. By performing a windshield survey, auditors likely pass through the study area more quickly, are less able to move closer to observed environmental features, and experience a greater sense of separation from the built environment outside of their vehicle. Additionally, raters auditing by a given travel mode may be more aware of or more influenced by environmental features oriented to their travel mode, whether it be walking or driving. Thus, raters auditing on foot may be more aware of environmental features that are pedestrian oriented, such as sidewalks, footpaths, street trees, and certain types of signage and lighting. Similarly, raters auditing by car may be more influenced by certain road characteristics and traffic control devices, such as speed bumps, curb extensions, and billboards.

It is important to keep in mind that trained observers, such as those used in this study, do not act nor perceive the built environment in the same way as a typical pedestrian or motorist. Auditors have been trained to evaluate the built environment according to predefined rules and standards. However, if anything, observer training should reduce the amount of differences in perception that would naturally occur, making this a more stringent test of the influence which travel mode has on observers' perception of the built environment. It is perhaps due to this effect of observer training that previous research has not addressed the influence of travel mode on neighborhood audit results. Yet, to assure data quality, researchers must determine if auditor's travel mode is a significant influencing factor in audit results.

Sample Audit Tools

The following section will provide a brief literature review of several neighborhood audits using direct assessment techniques. Special attention will be paid to the audit tools' design, implementation, and the contributions it has made to the field. The following examples of neighborhood audit tools are broken up into three categories: those conducted on foot, those conducted by car, and those using both vehicular and pedestrian data collection techniques. The final group includes those studies using both techniques jointly, as well as those studies that switched from one audit method to the other. The intended purpose of this section is to provide valuable background information on neighborhood audits and provide a basis for comparison to the PIN III neighborhood audit, which was originally conducted by car.

A. Audit Tools Conducted on Foot

SPACES

The Systematic Pedestrian and Cycling Environment Scan instrument, better known simply as SPACES, is perhaps the best known street-level audit instrument. Without a doubt, the development of this systematic audit tool played a significant role in the design and function of virtually all audit instruments that followed. The SPACES instrument, developed by Pikora et al., was applied to select street segments within Perth, the capital city of Western Australia in 2000. Audited segments were selected based on a 400 meter buffer around all 1803 residences of individuals who previously submitted surveys on their physical activity levels and behavior (Pikora et al., 2002).

Eight, two person groups audited a total of 12,925 street segments on foot and manually recorded audit results (ibid). Prior to evaluating street segments, observer participants underwent a three day training program. For the purpose of this audit, 37 environmental factors were evaluated and broken into 4 categories for analysis: Functional items (such as walking/cycling surfaces, street characteristics, and traffic), Safety features (such as lighting and traffic crossings), Aesthetic features (such as tree presence, views, and architecture), and Destination features (such as the presence of parks, shops, public transportation, etc.) (Pikora et al., 2002). The SPACES tool also included a subjective assessment of the attractiveness and difficulty for walking and cycling each segment. In addition to field observations, local information sources, such as traffic and GIS data were used to assess environmental features.

The primary purpose and contribution of the SPACES instrument was that it represented the first real attempt to make a systematic, comprehensive, and reliable audit tool to assess the built environment. Both total agreement and kappa statistics were calculated for re-evaluated audit segments. Though only a total of 27 segments were re-audit for reliability purposes, the inter and intra-reliability testing performed by SPACES helped establish the importance of assuring consistency and reliability of audit instrument tools. Not only does reliability testing help assure reliable and therefore useful audit results, it also helps identify which environmental features are more difficult to assess, whether it be due to definitional ambiguity or the intrinsic nature of the feature in question.

Irvine-Minnesota Inventory

In 2004, Kristen Day et al. developed the Irvine-Minnesota Inventory, a comprehensive audit tool with 162 items, comprising over 200 measures of the built environment (Day et al., 2006). The researchers structure their audit questions to assess four domains of factors that influence physical activity: Accessibility, Pleasurability, Percieved safety from traffic, and Percieved safety from crime. Most of the 162 items included in this audit measure some function of accessibility (ability to reach destinations and traverse the built environment), or pleasurability (aesthetic features or attractions linked to one's desire to traverse a given environment).

The inventory was created using data from a literature review, focus groups, a panel of experts, and field testing within a variety of urban environments (Day et al., 2006). By incorporating focus group input, the Irvine-Minnesota Inventory was able to broaden their analysis of the built environment, beyond those features previously existing in literature. Similar to past studies, a group of two, trained observers performed field observations of street segments on foot. Audit results were recorded by audited teams using Personal Data Assistant (PDA) technology. GIS technology was also used to measure an additional five features of the built environment, such as block length, which have been previously recorded and are easily accessible.

Though the Irvine-Minnesota Inventory did not differ dramatically from past audit tools, the existence and availability of this comprehensive audit tool serves as a great asset for those who require a more extensive assessment of their study area.

Pedestrian Environmental Data Scan (PEDS)

In 2006, Clifton, Smith, and Rodriguez developed the Pedestrian Environmental Data Scan (PEDS) as a means to reliably assess a range of environmental features. At one page in length, it was designed to balance the need for detailed information with that of ease of use. In total, the PEDS audit tool has forty questions, resulting in eighty-three different measures. Audit questions are structured to assess 4 primary elements of the built environment: 1) *Environment* (land use classification, slope, etc.), 2) *Pedestrian Facility* (path type, sidewalk characteristics, buffers, etc.), 3) *Road Attributes* (road condition, # of lanes, speed limit, parking, etc.) and 4) *Walking/Cycling Environment* (lighting, amenities, cleanliness, bicycle lane, etc.). PEDS also includes a subjective assessment of the built environment, rated on a four point scale. The vast majority of audit questions use an ordinal scale, with only four simple yes/no questions (Clifton et al., 2006).

Prior to auditing, raters underwent 2 days of training, both in the field and within a classroom setting. Although the PEDS instrument was originally created in a pencil and paper format, it was adapted so that auditors could take advantage of handheld technology. Researchers believed that the use of Personal Data Assistants (PDAs) improved data collection reliability. In the PEDS study, Auditors rated a total of 995 street segments on foot within the city of College Park, MD between June and July 2004. Rating was conducted by a pair of auditors who came to joint decisions on each feature assessed. In order to create comparable segments, street segments longer than 700 feet were broken up and rated in segments (Clifton et al., 2006).

Researchers employed three methods of assessing the reliability of audit results: 1) Kappa statistics, 2) Percent agreement and 3) Concordance Correlation. Percent agreement was used for those features of low frequency or little variation and concordance correlation was used to determine the reliability of features measured on a continuous scale (Clifton et al., 2006). For all other questions, kappa statistics were used when possible.

The PEDS instrument made several important contributions to the field of environmental audits. Its training program, audit tool design, and integration with handheld technology make it an efficient, thorough, and accessible instrument for assessing the built environment. Additionally, as will be discussed in the 'Testing Audit Tools' section, the developers of PED investigated a variety of audit administration techniques in order to determine which methods obtained the most reliable audit results.

B. Audit Tools Conducted by Car

Project for Human Development in Chicago Neighborhoods (PHDCN)

In this study, Raudenbush and Sampson (1999) assessed street-level data for all streets within 196 census tracts in Chicago. Audit data was collected by slowly driving down street segments at a speed of 5 miles per hour and simultaneously videotaping and making direct observation of neighborhood characteristics (Raudenbush and Sampson, 1999). Auditors within the vehicle worked in pairs to identify and rate neighborhood characteristics. Neighborhood features assessed included indicators of both physical and social disorder, such as trash, abandoned cars, graffiti, and loitering adults (ibid).

Researchers then inputted collected data into a three-level item response model in order to estimate the degree of physical and social disorder in the built environment. By analyzing their results, Raudenbush and Sampson found significant variability in physical disorder scores at both the neighborhood and face block level. Social disorder results were only found to be reliable at the neighborhood scale (Raudenbush and Sampson, 1999).

C. Audit Tools Conducted on Foot and by Car

St. Louis Tool

Hoehner et. al recently developed two environmental audit tools designed to measure nearly 150 features of the built environment, which fall into six broad categories: land use environment, recreational facilities, transportation environment, aesthetics, signage, and the social environment. Two versions of the audit tool were created. One version consisted of a twenty seven question checklist of dichotomous choices, in which auditors answered yes/no questions or marked whether a specific feature was present or absent. The second tool is more comprehensive in nature, with primarily ordinal questions that allow a wider range of responses.

Two features make the St. Louis tool unique. First, this audit tool was tested in a “highly walkable”, area (Savannah, Georgia) and a lower-income, “low-walkable” city (St Louis, Missouri). Second, both trained observers and untrained community participants gathered data using the St. Louis tool. Similar to SPACES, street segments within a 400 meter buffer around survey respondents were audited by a pair of auditors. Prior to field observations, a telephone survey was performed, using the International Physical Activity Questionnaire (IPAQ) to obtain information on physical activity. However, rather than manually recording audit results, a PDA (Personal Data Assistant) system was used to immediately record answers into an electronic format.

Several important findings were made by the St. Louis audit. When comparing the audit results of community member participants, community members from Savannah had consistently higher agreement with researchers than community members from St. Louis did. This fact demonstrates the significant amount of variation in observer responses that can occur from using untrained auditors, and suggests that some community participants may be much more familiar with their respective urban environment, leading to more consistent audit results. When assessing the audit results as a whole, Hoehner et. al found a positive relationship between the proximity of non-residential destinations and physical activity. Yet, they found "no direct association emerged between presence of recreational facilities and meeting [physical activity] recommendations" (Hoehner et al., 2005). Assuming that reported physical activity levels are accurate, this suggests that more than proximity is needed for people to meet the recommended amount of physical activity per day.

"Broken Windows" Index

In the study ' "Broken Windows" and the Risk of Gonorrhea', Cohen et al. evaluated housing and street conditions along 55 block groups with an average population of 507 people within New Orleans (Cohen et al., 2000). The purpose of the neighborhood audit was to asses the relation between neighborhood conditions and gonorrhea rates. The premise behind this relationship is that physical signs of social disorder, such as broken windows, directly influence individuals' behavior by providing clues as to what is socially acceptable (Cohen et al., 2000).

The sample block groups were assessed by planners at the College of Urban and Public Affairs (CUPA), University of New Orleans between 1994 and 1997, in five different sections of the city (Cohen et al., 2000). To collect data on the structural condition of the built environment, CUPA planners drove through the study area, videotaping each street segment within sample blocks. The images obtained were later assessed and rated on a 4 point scale: 1, no visible damage; 2, minor cosmetic damage; 3, minor structural damage; 4, major structural damage (Cohen et al., 2000). This assessment also included land use classification.

To collect more detailed, street-level data, auditors walked the block group areas and evaluated various environmental characteristics using dichotomous variables. Neighborhood features assessed in this manner included: garbage accumulation, graffiti, abandoned cars, billboards and signs, and general upkeep of non-structures such as parks, playgrounds, vacant lots, and institutional properties" (Cohen et al., 2000). Each block group was scored based on the aggregate scores of its street segments.

The actual "Broken Windows Index" is "the sum of the percentage of homes with major structural damage, minor structural damage, or cosmetic damage; the percentage of streets with trash, abandoned cars, or graffiti; and the number of physical problems and building code violations in public high schools…" (Cohen et al., 2000). Each of these variables was valued equally. Data from the broken windows index was then compared to the location of reported gonorrhea cases, which was geocoded using GIS software.

The results of this study suggest that "physical deterioration of a neighborhood is either a marker for a risk factor for gonorrhea or itself a risk factor for gonorrhea" (Cohen et al., 2000). Thus, researchers were only able to demonstrate correlation between "broken windows" and gonorrhea rates. They obtained no proof of causation. Additionally, no information on inter-rater reliability was provided.

Caughy, O'Campo, and Patterson's Observational Instrument

The observational instrument used in this study was created by combining and adapting items included in the Project for Human Development in Chicago Neighborhoods (PHDCN) and the work of Ralph Taylor and fellow researchers. A draft protocol was pilot tested by conducting windshield surveys in two socioeconomically diverse neighborhoods. After pilot testing, the protocol was further refined, resulting in an observational instrument evaluating 45 objective items. Together, these 45 items assess "the condition of grounds and undeveloped spaces, indications of block uniformity/territoriality, type of street, presence of graffiti/litter, neighborhood resources, and presence and activities of people" (Caughy et al., 2001).

A total of 57 study neighborhoods in Baltimore City were selected for evaluation. Neighborhoods were defined as census block groups. Prior to auditing, raters had to undergo thirty hours of training over six days. The marker for when training was complete was when auditors obtained an eighty-five percent agreement on features assessed.

The audit tool was administered by block, with audit pairs evaluating both sides of the street and coming to a consensus on environmental features assessed. Initially, auditors coded neighborhood conditions by performing a windshield survey. However, early in the data collection process auditors switched to rating on foot, due to the belief that windshield assessments were received negatively by neighborhood residents. When rating on foot, data collection took approximately five to ten minutes per block. Overall, 1135 blocks were rated, with an average of twenty blocks within each neighborhood.

Three primary factors were estimated using audit data: 1) *Physical incivilities* (also known as physical disorder) – such as graffiti, trash, and vacant buildings, 2) *Territoriality* – as indicated by markers of defensible space (walls, fences, or symbolic barriers) and territorial functioning (property maintenance, symbols of protection, beautification, etc.), and 3) *Availability of play resources*- which included 17 indicators (Caughy et al., 2001). Additional measures included a composite measure of impoverishment derived from census data, neighborhood crime density, and neighborhood perceptions obtained through a series of interviews.

Audit results were fed into a three-level item response level, in which level one was the item, level two was the street, and level three was the neighborhood (Caughy et al., 2001). Based on their data, Caughy et al. correctly predicted a negative correlation between physical incivilities in a neighborhood and play resources (of -0.29). Territoriality was not found to be correlated with play resources, nor was it negatively correlated with physical incivilities, as expected (Caughy et al., 2001). At the neighborhood level, audit results had a very high level of reliability – over 90 % agreement for physical incivilities, territoriality, and play resources. However, at the street-level scale, the reliability (as in indicated by percent agreement) of physical incivilities was .74 and territoriality and play resources were only .33 and .42 (Caughy et al., 2001). Thus, researchers found that "although there is sufficient variation in territoriality and play resources to distinguish *between* neighborhoods, there is insufficient variation to distinguish between streets *within* the same neighborhood" (ibid).

Lessons

It is important to note that while there have been a number of neighborhood audits that have evaluated the built environment by car, some researchers have found it necessary to switch away from this approach (Caughy et al., 2001) or combine windshield survey data with information obtained on foot (Cohen et al., 2000) in order to gain a full knowledge of the study area. This suggests that auditing by car may not be the ideal method for some studies, depending on their focus and the nature of their audit questions. Table 1 (below) provides a more comprehensive list of audit techniques used by various studies. Although this list is by no means all inclusive, it demonstrates that most neighborhood audits are conducted on foot.

Table 1: Audit Administration Comparison Chart

	Mode of Audit Administration		
Audit Tool	**Walking**	**Driving**	**Walking and Driving**
Pin III		X	
PEDS	X		
Irvine Minnesota Inventory	X		
SLU			X
Emery Instrument	X		
PBIC Checklist	X		
SPACES	X		
NOC	X		
Pin II		X	
COP* Observational Instrument			X
PHDCN		X	
"Broken Windows" Index			X
Walkable Places Survey (WPS)	X		
PBIC Walkability Checklist	X		

PIN III: Pregnancy, Infection, and Nutrition Study; PEDS: Pedestrian Environmental Data Scan; SLU: Analytic Audit Tool – Saint Louis University; SPACES: Systematic Pedestrian and Cycling Environmental Scan; NOC: Neighborhood Observational Checklist; COP* Observational Instrument: audit tool developed by Caughy, O'Campo, and Patterson; PHDCN: Project for Human Development in Chicago Neighborhoods; PBIC Checklist: Partnership for a Walkable America

Testing Audit Tools

A number of tests have been performed on audit tools design and implementation, in order to determine how various factors influence audit results. A common and simple test used to improve reliability involves changing operational definitions or reducing the number of response options for audit questions (Zenks et al., 2007; Hoehner et al., 2005; PIN III study). For example, during development of the Neighborhood Observational Checklist (NOC), researchers changed an audit question to refer to the presence of "un-drivable" cars, rather than "abandoned" cars, which auditors found difficult to define (Zenks et al., 2007). Other tests are more involved and produce more informative results. In the St. Louis study, researchers were able to compare the audit results of trained observers and untrained community participants, in order to evaluate the influence which observer training has on auditors' perceptions and data reliability (Hoehner et al., 2005).

During development of the Pedestrian Environmental Data Scan (PEDS), different methods of administering the audit were tested in order to determine how these changes would influence reliability scores. Researchers tested the effect of having auditors rate individually versus in pairs and found auditor pairs to have higher reliability. They also tested the impact of having each auditor rate all aspects of the built environment versus having each auditor specialize on specific features included in the audit. It was found that having both auditors assess all features included in the audit resulted in greater reliability scores. Finally, Clifton et al. (2006) tested variation in the reliability of audit results by urban context (Clifton et al., 2006). The authors found that auditors generated higher reliability scores on residential segments than commercial segments. This difference was attributed to the greater complexity and variation of commercial segments (Clifton et al., 2006). Just as each of these tests helped researchers understand how various changes impact audit results, it is my hope that the results of this study will further expand that knowledge base and be taken into account during future audit design.

Methods

Study Area/Selection Criteria

As previously mentioned, this study seeks to determine if the travel mode by which an audit is conducted influences audit results. The PIN III neighborhood audit tool, created by Kelly Evenson et. al served as an ideal choice for performing this test. Not only was this audit tool originally audited by car, but my prior experience using the tool during the summer of 2006 created additional advantages, which will be discussed later in the paper. The PIN III neighborhood audit tool was created in order to expand upon the amount and type of information used in the PIN III study. The PIN III study is designed to test whether physical activity or stress are associated with preterm birth (http://www.library.unr.edu/subjects/guides/apa.html). Data collection methods include research clinic visits, telephone interviews, questionnaires, and the PIN III neighborhood audit (ibid).

The PIN III neighborhood audit tool was used during the summers of 2005 and 2006 to evaluate 39 characteristics of the built environment that evidence suggests is linked to physical activity. The original study area included a four county area, using both urban and rural street segments within Orange, Chatham, Durham, and Alamance County. Similar to previous audits, Evenson et al. used a buffer area around survey participants during the sampling process.

In order to test the influence of travel mode on audit results, a sub-set of the audit segments originally assed by car as part of the PIN III neighborhood audit were re-evaluated on foot. For the purpose of this study, re-evaluated audit segments were required to meet the following criteria: previously used for Inter-Rater Reliability (IRR), rated by a 2007 auditor during the summer of 2006, and classified as ‘urban’. These criteria were chosen to limit the number of undesired variables between this study and the original PIN III Neighborhood audit, in order to better capture the effect of conducting the audit by a different travel mode. By limiting audit segments to those I had previously rated a year before, I was able to greatly reduce the natural variation in audit results that occurs between audit teams. Rural segments were also removed from the sample list due to the danger and time requirements of rating such segments on foot. It is unclear whether audit results from rural segments would be more or less affected by the travel mode of the auditor. Since rural segments usually contain fewer items to rate per segment, one might hypothesize that auditors would be less likely to miss the presence of a given environmental feature. However, it is much harder to drive slowly on rural segments, so auditors driving might miss items or incorrectly assess items for this reason.

Finally, a few additional audit segments were removed from the study due to their isolated position in the study area, or due to complications in the field. Using this selection process, 79 urban segments within Orange, Durham, and Chatham County were re-assessed. Although the PIN III neighborhood audit does not differentiate urban and suburban segments, the vast majority of re-evaluated segments were suburban in nature.

Audit Tool Content

The PIN III neighborhood audit tool assesses 39 distinct characteristics of the built environment believed to influence physical activity. The tool is comprised of 43 questions, evaluating a range of environmental features dealing with: residential land use, non-residential land use, aesthetics, walking and bicycling amenities, and transit and road characteristics. The pencil and paper PIN III audit instrument is shown in Figure 1. In developing the PIN III audit tool, Evenson et. al sought a balance between creating an extremely detailed, comprehensive audit tool versus creating a more concise inventory with a smaller time commitment and greater ease of use. At 43 questions, the PIN III audit tool is comparable to the PEDS and SPACES instrument in length, but significantly shorter than the Irvine Minnesota Inventory.

Even with the multitude of audit tools which have been developed over the years, no consensus has been reached on exactly which features collectively create a walkable place. This fact is illustrated by the range of environmental features that have been evaluated by different audit tools (Moudon et al., 2003). One way in which the PIN III audit tool attempts to address this problem is through the inclusion of a subjective question, which asks the rater to assess the overall walkability of a given street segment. In addition to allowing auditors to more freely assess the urban area as a whole, the inclusion of a subjective assessment is one method of capturing influential variables that purposefully or unknowingly are not included in the audit tool.

Figure 1. PIN III audit instrument

NEIGHBORHOOD DATA COLLECTION TOOL

#	Item	Response Options
	Palm ID #	
	Rater ID #s	
	Time of day	
	Street segment number	
	Street name	
1	Subjective assessment: Is this street walkable	Relative to other areas you have observed, please provide a subjective assessment. ☐ Strongly agree ☐ Agree ☐ Disagree ☐ Strongly disagree
RESIDENTIAL LAND USE		
2	Number of residential units	# **(if 0, then skip entire section)**
3	Types of residential housing *(mark all that apply)*	☐ Single family - detached ☐ Multi-family/apartment/duplex ☐ Mobile homes/trailer home ☐ Housing authority/HUD projects ☐ New construction / renovation
4	Overall condition of most residential units	☐ Excellent condition ☐ Good condition ☐ Fair condition ☐ Poor/deteriorated condition ☐ Mixed condition *(extreme differences)* ☐ Cannot see because heavily wooded **(if cannot see, then skip entire section)** ☐ Cannot see for some other reason **(if cannot see, then skip entire section)**
5	Overall condition of resident-kept grounds	☐ Not applicable - no grounds ☐ Excellent condition ☐ Good condition ☐ Fair condition ☐ Poor/deteriorated condition ☐ Mixed condition *(extreme differences)*
6	Type of most front yards	☐ None ☐ Traditional lawn ☐ Landscaped ☐ Heavily wooded ☐ Mixed conditions *(extreme differences)*
7	Presence of porches	☐ None ☐ Less than half ☐ Half or more than half
8	Presence of some form of decoration	☐ None ☐ Less than half ☐ Half or more than half
9	Presence of border (fences/shrubs)	☐ None ☐ Less than half ☐ Half or more than half

10	Presence of visible security warning signs	☐ None ☐ Less than half ☐ Half or more than half
11	Any burned, boarded up, or abandoned residential units	☐ No ☐ Yes
NONRESIDENTIAL LAND USE		
12	Presence of nonresidential commercial (i.e., restaurants. cafes) land use	☐ None ☐ Yes, new construction / renovation ☐ Yes, existing ☐ Yes, new construction / renovation and existing
13	Presence of nonresidential industrial land use	☐ None ☐ Yes, new construction / renovation ☐ Yes, existing ☐ Yes, new construction / renovation and existing
14	Presence of nonresidential agricultural land	☐ No ☐ Yes, field or farm
15	Number of religious structures on segment	#
16	Overall condition of most buildings (commercial, industrial, agricultural, or religious)	☐ Not applicable ☐ Excellent condition ☐ Good condition ☐ Fair condition ☐ Poor/deteriorated condition ☐ Mixed condition *(extreme differences)*
17	Any burned, boarded up, or abandoned nonresidential units (commercial, industrial, agricultural, or religious)	☐ No ☐ Yes
18	Presence of home-based businesses (e.g., repairs, sales, etc)	☐ No ☐ Yes
PUBLIC, RESIDENTIAL AND NONRESIDENTIAL SPACE / AESTHETICS		
19	Presence of land that is vacant / underdeveloped	☐ None **(if none, skip to #21)** ☐ Less than half ☐ Half or more than half
20	Overall condition of land that is vacant / underdeveloped	☐ Excellent condition ☐ Good condition ☐ Fair condition ☐ Poor condition (illegal dumping) ☐ Mixed condition *(extreme differences)*
21	General condition of public spaces	☐ Excellent condition ☐ Good condition ☐ Fair condition ☐ Poor / deteriorated condition ☐ Mixed conditions *(extreme differences)* ☐ Not applicable (private road)

Page 1

NEIGHBORHOOD DATA COLLECTION TOOL

22	Visible people	☐ None (**skip to #24**) ☐ Children/youth only ☐ Adults only ☐ Children/youth and adults
23	Are the people being physically active	☐ No ☐ Yes, children/youth ☐ Yes, adults ☐ Yes, children/youth and adults
24	Any public or neighborhood park or playground in this segment *(mark all that apply)*	☐ No (**skip to #26**) ☐ Yes, park ☐ Yes, playground ☐ Yes, church park &/or playground
25	Overall condition of park and/or playground	☐ Excellent condition ☐ Good condition ☐ Fair condition ☐ Poor / deteriorated condition ☐ Mixed conditions *(extreme differences)*
26	Visible dogs	☐ No ☐ Yes, none are loose and out of yard ☐ Yes, at least one loose and out of yard
27	Amount of litter	☐ None (**skip to #29**) ☐ A little ☐ A moderate amount ☐ A considerable amount
28	Type of litter *(mark all that apply)*	☐ Nonalcoholic cans/bottles/paper ☐ Alcoholic cans/bottles ☐ Large items (tires, furniture, appliances, cars) ☐ Other litter
29	Amount of graffiti	☐ None ☐ A little ☐ A moderate amount ☐ A considerable amount
WALKING AND BICYCLING AMENITIES		
30	Presence of sidewalk	☐ None (**skip to #33**) ☐ One side of street, whole segment ☐ One side of street, partial segment ☐ Both sides of street, whole segment ☐ Both sides of street, partial segment
31	Sidewalk buffer	☐ Adjacent to street or curb (no buffer) ☐ Within 2 feet of street (buffer) ☐ Between 2-6 feet of street (buffer) ☐ Greater than 6 feet of street (buffer)
32	Sidewalk condition	☐ Good (very few bumps/cracks/ holes) ☐ Fair (some bumps / cracks / holes) ☐ Poor (many bumps / cracks / holes) ☐ Under repair
33	Presence of footpath along road	☐ None ☐ One side of street, entire length ☐ One side of street, partial length ☐ Both sides of street, entire length ☐ Both sides of street, partial length

34	Any trails that you can see in this segment	☐ No ☐ Yes, soft surface ☐ Yes, hard surface ☐ Yes, hard and soft surface
35	Trees shading walking area	☐ No trees along segment ☐ Yes, some trees along segment ☐ Yes, trees along entire segment
36	Public lighting	☐ None ☐ Road oriented ☐ Pedestrian oriented ☐ Both road and pedestrian oriented
TRANSIT AND ROAD CHARACTERISTICS		
37	Transit facilities (mark all that apply)	☐ None ☐ Bus stop without bench or shelter ☐ Bus stop with shelter ☐ Bus stop with bench
38	Number of lanes	Min number of lanes to cross: ____ Max number of lanes to cross: ____
39	Is road paved	☐ No ☐ Yes, paved only ☐ Both, paved and gravel
40	Highest speed limit for segment	MPH: # ______ (if unknown enter 0)
41	Presence of a shoulder or bike lane	☐ No ☐ Yes, soft surface ☐ Yes, hard surface but not a bike lane ☐ Yes, hard surface bike lane
42	On-street parking	☐ None or not allowed ☐ Allowed, but restricted ☐ Allowed, no restriction
43	Traffic control devices, crossing aids, and signs in segment (mark all that apply)	☐ None ☐ Traffic light(s) ☐ Flashing warning sign(s) ☐ Stop sign(s) ☐ Pavement marking / crosswalk(s) ☐ Yield to pedestrian paddles / signal / crossing street sign(s) ☐ "Share the road" bicycle sign ☐ Other pedestrian or bike friendly traffic signs ☐ Bicycle parking facilities ☐ Speed bumps ☐ Median / traffic island ☐ Curb extension(s) ☐ Neighborhood entrance signs (e.g. Foxcroft) ☐ Neighborhood crime watch ☐ No trespassing(s) ☐ Beware of dog / invisible fence ☐ Billboard

It is important to note that, despite the usefulness of the PIN III audit tool for evaluating features of the built environment related to physical activity, it was not designed specifically for the research purpose of this study. Thus, the ability of this tool to identify features which may have distinct influences upon observers while traveling by different modes is constrained by the existing audit structure. For example, the PIN III audit does not include any questions specifically targeting perception of safety or interaction with community members. While these factors may significantly influence walkability and likely vary according to ones travel mode, the subjective assessment question serves as the sole outlet for capturing the possible effects of these environmental factors.

Training

In order to ensure high consistency and reliability in audit results, neighborhood auditors were required to go through extensive training in the application of the audit tool. Two auditors with relevant educational backgrounds and previous experience using the PIN III audit tool were enlisted for this study. Prior to data collection, each auditor participated in a five day training session designed to familiarize participants with the study design, data collection methods, as well as each other. The original training session for the two auditors participating in this study took place either during the summer of 2005 or 2006. Training consisted of: establishing operational definitions of all terms, reviewing a slide show of neighborhood attributes, palm pilot training, establishing how to use GPS devices to determine the location of new street segments, and multiple practice sessions using the study protocol in the field. Each neighborhood walk-through was designed to help establish a collective understanding of how the protocol should be applied, with the intent of improving inter-rater reliability. A copy of the PIN III neighborhood audit training schedule is included in the appendix.

In addition to their original training sessions, 2007 auditors underwent an additional day of training, in which emphasis was placed on re-familiarizing auditors with the PIN III Neighborhood Data Collection Manual and additional practice sessions were performed out in the field, on foot. Practice segments were chosen in both residential and commercial areas of Chapel Hill in order to give auditors a chance to apply all measures included in the audit tool, in the same urban context as actual audit segments.

Data Collection

Data collection consisted of rating 39 chosen characteristics of the built environment for all 79 street segments in the study area. In contrast to the original PIN III neighborhood audit, pairs performed the neighborhood audit on foot, rather than by car. While the same audit tool was used to rate, the method by which the audit was performed differed slightly. To begin rating, auditors first identified an endpoint of an audit segment, using their audit maps and street signs as guides. Each segment, which is defined by the street section between two consecutive intersections, was given a unique 'AuditID' number. Prior to the study's start date, a series of audit maps were created in GIS, each showing a study area containing anywhere from one to twelve audit segments, identified by their street names and unique Audit IDs. When street signs could not be located or street labels were missing, auditors used their audit map or the shape of intersecting roads to assist in the identification process. Once the street was verified as an audit segment, raters walked the entirety of each segment and evaluated all properties and grounds on both sides of the street, as well as the street itself. In order to improve reliability, auditors jointly reached a consensus on each environmental characteristic evaluated in the audit tool. Only those structures facing the audit segment were included in their evaluation. Additionally, while auditors rated both public and private roads, they did not audit driveways or any road marked 'no trespassing'.

If raters discovered a new intersecting road within the segment indicated by their map, they were instructed to continue to rate the road to the end of the original road segment. This ensured that the same length of road was evaluated during each Inter Rater Reliability (IRR) session. In a few cases, audit segments were skipped and thrown out of the audit because the extent of the original road segment was unclear.

Recording Responses

Several past neighborhood audits cite the advantages of using PDA devices over pencil and paper methods, stating that by eliminating the need for subsequent data input, one reduces input error and improves data quality (Clifton et al., 2006; Zenks et al., 2007; Day et al., 2006; Hoehner et al., 2005). Due to complications with the required software, PDAs were not used for this study, despite their use during the 2005 and 2006 sessions of the PIN III neighborhood audit. However, for this study, the lack of PDA use was not believed to be a significant factor, as explained below. The benefits of using a PDA in regards to reducing subsequent input error is only true if one inputs the data immediately into the PDA device. Immediate data input is logical and appropriate for questions regarding the presence of certain features along a given segment, but, depending upon the manner of evaluation used, makes less sense for questions regarding the condition of elements of the built environment.

During the summer of 2006, neighborhood auditors found that in order to accurately determine the mean score for the condition of houses or yards along a street segment, they needed to physically or mentally tally the results from a number of units. When evaluating longer segments, keeping track of each unit's condition becomes increasingly difficult. Yet, if raters physically tally the results of each unit, they negate many of the advantages of having the PDA. In my experience as an auditor, this was often found to be the case when rating road segments as part of the PIN III neighborhood audit study. In fact, it became common practice to tally scores for all of the questions on an abbreviated audit tool form and then quickly copy all of the answers into the PDA upon completion of the segment. This practice thus had the same potential of introducing input error as using a standard pencil and paper recording method. As a result of this existing limitation, the lack of PDA use in this study is believed to be an acceptable limitation. Furthermore, the small sample size (79 segments) of this study made performing quality control on recorded results relatively quick. However, all that being said, the fact that different recording methods were used means that one cannot *conclusively* state that observed differences in audit results are due to the change in travel mode and not the change in recording methods.

Time Requirement

Using the method described above, the data collection protocol took, on average, between 8 and 14 minutes to complete for each street segment. Due to the size of the study area, a private vehicle was used to transport auditors *between* segments. Depending upon the proximity of audit segments, in some cases auditors found it more efficient to walk to the next closest segment. Auditing was performed between August 6th and August 9th 2007, from 8:30 A.M. to 4:30 P.M.

Although this study did not perform detailed comparison of time requirements using the two different travel modes, no significant difference in the amount of time it took to rate urban segments by car versus foot was detected. Of course, the same would not be true of rural segments. In the original PIN III audit sessions, raters drove as slowly as traffic permitted in order to better observe environmental features. Additionally, the inability to stop while in traffic meant that audit teams often had to make multiple trips along busy roads in order to complete their assessment. Thus, though assessing by car was likely faster for roads with little traffic, rating by foot was found to be easier and quicker for high traffic street segments.

Data Analysis

A reliability assessment was performed for all 43 questions in the PIN III neighborhood audit tool using the recorded results of re-evaluated segments included in this study. By comparing the reliability of the recorded results between the 2006 and 2007 audit sessions, I hoped to identify which environmental features were perceived differently due to the change in travel mode by which the audit was administered. Thus, I in essence performed a test-retest reliability measure and searched for any measures in which auditors obtained different results for the same segments. Though some small amount of variation is inevitable due to the year-long time delay between audit sessions and simple human error, significant differences in audit scores were attributed to the change in travel mode.

Two methods of assessing the reliability of audit results were used: percent agreement and kappa statistics (Landis and Koch, 1977). Percent agreement is simply the raw agreement between recorded scores for each audit session. Recorded scores must match exactly in order to pass this test. Past research suggests that percent agreement is a more appropriate measure for those variables with little variation, as well as for those features which are infrequent in the urban environment studied (Handy et al., 2002; Clifton et al., 2006). Overall, this study used a relatively small sample size (max = 79), with some variation in sample size between different questions. However, a larger sample size could not be obtained without sacrificing the advantages of the selection criteria used. Variation in sample size is a result of the absence of certain environmental features, such as residences or commercial structures, which precludes an assessment of environmental features dependent on their presence. Due to the small sample sized and relatively homogenous nature of audit segments used in this study, I was forced to rely more heavily upon percent agreement than many past studies.

The Kappa statistic is a chance-corrected agreement measure, which is derived by comparing the total agreement against that which might be expected by chance (Landis and Koch, 1977). Thus, where it can be calculated, Kappa statistic gives one a better measure of the deliberate agreement between observers. Kappa scores range from -1 to 1, with 1 indicating perfect agreement and any negative value representing a case in which agreement levels were lower than that which could be predicted by chance. I used SPSS Version 14 to cross-tabulate items scores and calculate kappa statistics for all nominal and ordinal variables, where possible. For ordinal variables, weighted kappa statistics were calculated to determine not simply whether auditors displayed disagreement, but to what degree they disagreed for a given environmental feature. Standard kappa statistics were calculated for all other questions.

In order for kappa statistics to be calculated, all cells within the cross-tabulation tables created from the responses of each audit session must have values. Thus, there cannot be any recorded values during one audit session that were absent from the other, for any given question. Due to the small sample size of this study, this limitation barred the use of Kappa statistics for some questions. For example, in the question regarding the overall condition of most residential units, asymmetry occurred because auditors recorded the answer 'Mixed conditions' in 2006, but not in 2007. Where possible, categorical responses were combined in order to eliminate asymmetry and allow kappa statistics to be generated. However, combining question responses was not always possible or reasonable, leaving some questions without kappa scores. For other questions, Kappa statistics were not calculated because some answer responses were so infrequent that they would lead to skewed Kappa scores. For these questions, percent agreement alone had to be used. This is a common problem for questions asking raters to mark the presence of rare environmental features, such as graffiti or vacant land.

Since there are no 'true' continuous variables contained within the PIN III neighborhood audit, those questions asking raters to record a numeric value (# homes, # churches, # lanes, speed limit) were recoded categorically so that kappa statistics could be used for these features as well. The one exception to this practice concerns question #2, which asks raters to record the number of residential units present on a segment. Due to the amount of variation in responses between audit segments (min = 0, max = 29), there was no practical way to categorize responses. As a result, percent agreement and a comparison of means were used to evaluate this question.

When using percent agreement to measure reliability, an 80% agreement standard was used to identify those question which showed significant differences between the two audit sessions, as a result of the different method of administration. Thus any audit results for an environmental feature with less than 80% total agreement was found to be significantly influenced by the change in travel mode. A similar standard was used for kappa statistics.

Landis and Koch (1977) developed the following scale for interpreting kappa statistics:

.2 → .0 = poor agreement
.4 → .2 = fair agreement
.4 → .6 = moderate agreement
.61 → .8 = substantial agreement
.80 → .99 = almost perfect agreement

Since one of the auditors in the auditing pair (myself) re-evaluated the exact same segments, one would expect higher agreement scores overall, compared to studies using different audit pairs. As a result, any question with less than 'substantial agreement' was considered to exhibit a significant difference. Using this standard, the audit results from two additional questions (#36 and portions of #43) were found to display significant differences when obtained by different travel modes.

Results

The majority of the questions included in the PIN III audit tool had high reliability scores, displaying little variation between the two audit sessions due to the change in travel mode. This was not unexpected given the nature of many of the audit questions. Yet, more importantly, 15 questions were found to have significant differences between audit sessions using the 80% percent agreement standard. Two additional questions exhibited significant differences (<.61) using the kappa standard. The results of the reliability measures are presented in the table below.

Table 2: Reliability between 2006 / 2007 audit sessions

Pin III Audit Questions		Sample size	Raw agreement	Kappa (95% CI)
1	**Is this street walkable?**	79	0.65	.50 (.35 , .66)
2	**# of residential housing**	79	0.70	-
3	**Types of residential housing****	59		
	Single family - detached (--Marked)		0.95	.79 (.57, 1)
	Multi-family/apartment/duplex (--Marked)		0.98	.95 (.87, 1)
	Mobile homes/trailer home (--Marked)		1.00	1
	Housing authority/HUD projects (Marked)		1.00	1
	New construction / renovation (--Marked)		0.97	N/A*
4	**Overall condition of most res. units**	59	0.68	0.39 (.19, .60)
5	**Condition of resident-kept grounds****	58	0.74	.66 (.48, .83)
6	**Type of most front yards****	58	0.85	N/A*
7	**Presence of porches****	58	0.95	N/A*
8	**Presence of some form of decoration****	58	0.72	N/A*
9	**Presence of border****	58	0.71	.54 (.35, .73)
10	**Presence of security warning signs****	58	0.86	.77 (.61, .92)
11	**Any abandoned residential units****	58	1.00	-
12	**Presence of commercial land use**	79	0.94	0.75 (.54, .96)
13	**Presence of industrial land use**	79	1.00	-
14	**Presence of agricultural land**	79	1.00	-
15	**# of religious structures on sgmnt**	79	98.70	N/A*
16	**Overall condition of most buildings**	79	0.90	N/A*
17	**Any abandoned nonresidential units**	79	0.99	N/A*
18	**Presence of home-based businesses**	79	0.99	N/A*

Pin III	Audit Questions	Sample size	Raw agreement	Kappa (95% CI)
19	**Presence of vacant/underdeveloped land**	79	0.96	N/A*
20	**Condition of vacant/underdeveloped land****	1	0.00	-
21	**General condition of public spaces**	79	0.70	N/A*
22	**Visible people**	79	0.60	
23	**Are the people being physically active****	18	0.78	
24	**Any public/neighborhood park/playground**	79		
	No (--Marked)		0.92	N/A*
	Yes, park (--Marked)		0.92	N/A*
	Yes, playground (--Marked)		0.96	N/A*
	Yes, Church park &/or playground (Marked)		1.00	-
25	**Condition of park &/or playground****	4	0.50	N/A*
26	**Visible dogs**	79	0.84	N/A*
27	**Amount of litter**	79	0.62	.34 (.17, .52)
28	**Type of litter****	40		
	Nonalcoholic cans/bottles/paper (Marked)		1.00	-
	Alcoholic cans/bottles (--Marked)		0.63	.28 (.08, .48)
	Large items (--Marked)		0.93	N/A*
	Other litter (--Marked)		0.80	N/A*
29	**Amount of graffiti**	79	0.95	N/A*
30	**Presence of sidewalk**	79	0.87	.79 (.67, .91)
31	**Sidewalk buffer****	31	0.84	.83 (.69, .97)
32	**Sidewalk condition****	31	0.90	N/A*
33	**Presence of footpath along road**	79	0.92	N/A*
34	**Any visible trails on segment**	79	0.92	N/A*
35	**Trees shading walking area**	79	0.70	.58 (.43, .72)
36	**Public lighting**	79	0.80	.60 (.42, .77)
37	**Transit facilities**	79		
	None (--Marked)		1.00	1.00
	Bus stop w/o bench or shelter (--Marked)		1.00	1.00
	Bus stop with shelter (--Marked)		1.00	1.00
	Bus stop with bench (--Marked)		1.00	1.00
38A	**Minimum Number of lanes to cross**	79	0.95	.73 (.49, .98)
38B	**Maximum Number of lanes to cross**	79	0.99	N/A*
	<u>Or recoded as:</u>	79	0.99	1.00
	1-2 lanes			
	3-4 lanes			
39	**Is road paved**	79	0.96	.71 (.41, 1)
40	**Highest speed limit for segment**	79	0.96	
	<u>Or recoded as:</u>	79	0.96	.94 (.87, 1)
	0 mph			
	10-25 mph			
	30-45 mph			
41	**Presence of a shoulder or bike lane**	79	0.96	N/A*
42	**On-street parking**	79	0.77	.55 (.37, .73)
43	**Traffic control devices/crossing aids/signs**	79		
	None (--Marked)		0.89	.68 (.49, .86)
	Traffic light(s) (--Marked)		0.99	.93 (.78, 1)

PIN III Audit Questions	Sample size	Raw agreement	Kappa (95% CI)
Flashing warning sign(s) (--Marked)		1.00	-
Stop sign(s) (--Marked)		0.94	.87 (.77, .98)
Pavement marking / crosswalk(s) (--Marked)		0.99	.93 (.78, 1)
Yield to ped paddles/signal/crossing sign(s) (--Marked)		0.94	.51 (.14, .88)
Share the road bicycle sign (--Marked)		1.00	-
Other ped or bike friendly traffic signs (--Marked)		0.87	.32 (0.0, .63)
Bicycle parking facilities (--Marked)		0.99	.66 (.04, 1)
Speed bumps (--Marked)		0.96	.71 (.39, 1)
Median / traffic island (--Marked)		0.97	.79 (.50, 1)
Curb extension(s) (--Marked)		1.00	-
Neighborhood entrance signs (--Marked)		0.92	N/A*
Neighborhood crime watch (--Marked)		0.95	.72 (.47, .98)
No trespassing(s) (--Marked)		0.91	.42 (.06, .77)
Beware of dog / invisible fence (--Marked)		0.94	.67 (.40, .94)
Billboard (--Marked)		0.96	N/A*

Note:
* Skewed Kappa Score (variable not present or too infrequent to compute valid Kappa score)
** Reduced sample size due to skip patterns

However, applying these two standards alone, without looking more closely at the sample size and nature of a question, can provide deceptive results. Two of the questions that demonstrate low percent agreement scores (Condition of vacant land; condition of park/playground) occur too infrequently (n =1 & n =4) to rely on percent agreement or kappa scores. Moreover, three additional audit questions (Visible people; Visible people active; Amount of litter) displaying low agreement scores have significant amounts of natural variation over time, and are of no real use to the purpose of this study. While one might expect auditors rating on foot to perceive greater amounts of litter, there was no clear directionality present between responses from the two audit sessions to suggest that differing answers for this question were of any significance. Finally, the low level of agreement (77%) auditors had when assessing on-street parking is most likely due to the operational definition used, rather than actual variation in observer responses. Because auditors mark 'no on-street parking' for street segments in which they observe *either* no cars parked on the street *or* signage indicating that on-street parking is not allowed, audit responses will vary significantly due to the presence of vehicles on the segment at any given time.

By removing those six questions from the list, 11 questions displaying significant variation between the two audit sessions remain. Six of these remaining questions were found to have 70% or lower raw agreement scores. It is not surprising that there is significant variation (65 % agreement) between the two audit sessions for the question asking auditors to rate the overall walkability of a segment. Subjective assessments will naturally vary both between and within audit groups. However, the clear directionality of the difference in audit scores is intriguing. Of the 28 segments displaying different walkability scores, 27 of the segments were rated lower by auditors rating on foot in 2007.

Three possible explanations for this occurrence were identified. It could be that auditors were in fact influenced by environmental features differently while on foot, compared to when auditing from a vehicle. Auditors might also have been influenced by environmental features that were not perceived at all while rating by car, thus affecting their subjective assessment of the segment. However, the low reliability scores for the subjective assessment of the built environment could be a result of the nature of the sampled segments. Compared to the entire sampling of audit segments included in the PIN III neighborhood audit, re-evaluated segments were fairly homogenous and predominantly in well maintained urban areas. It is possible that 2006 auditors were influenced by the character of other segments included in the audit. A segment that raters ‘agreed’ was walkable in 2007 might have caused raters in 2006 to ‘strongly agree’, when compared to less maintained segments in downtown Durham or rural portions of Chatham and Alamance Counties. Although observer training *should* significantly minimize this problem, it could still be an influencing factor.

The low reliability scores for question #2 (number of residential units) may be deceptive as well. Out of 79 total segments re-evaluated, 24 displayed disparate scores. Nineteen of the segments were off by a single residential unit. Despite the fact that 17 of the 24 recorded differences were cases in which auditors on foot recorded more residential units present, I question the value of this statistic. It is logical that auditors on foot might perceive an additional residential unit that is tucked behind another structure. However, the likelihood of this occurring in such frequency as to explain the number of differences recorded for this question is slim. As a result, I would be skeptical to make any judgment based on the low reliability scores for this question, believing instead that some type of error occurred.

Audit questions concerning the condition of public space, condition of houses, presence of alcoholic litter, and amount of shading all were found to have 70% or lower percent agreement as well. When rating the condition of public space, 16 of the 24 differences recorded were cases in which auditors on foot gave the segment a lower rating. Furthermore, although the presence of alcoholic litter would logically change significantly over time, all 15 of the recorded differences were instances in which auditors on foot noticed the presence of alcoholic litter. Such clear directionality points to a distinct difference when rating on foot rather than by car.

Similar to the 'condition of public spaces' question, auditors' assessment of the condition of residential units displayed low reliability scores (68% raw agreement). However, there was no clear directionality within the audit results. The same is true of auditors' assessment of the amount of tree shading. Yet, despite the lack of directionality in the former case, intuitively one would assume that actually walking a street segment would be a more appropriate means of assessing the amount of shade and degree of litter along a street segment. Though the data results can not prove this conclusion to be the case, they strongly suggest that the travel mode used makes a clear difference for these questions.

Auditors' assessment of the 'condition of grounds', 'presence of decoration', and 'presence of border' displayed significant differences in their percent agreement scores as well. The questions concerning the presence of decoration or a border were the only ones exhibiting directionality. For the 'presence of a border' question, 11 of the 17 differences were instances in which 2007 raters recorded fewer borders along a segment. There is no clear explanation why this would be the case. Furthermore, while 13 of the 16 differences in rated decoration were cases in which 2007 raters observed less decoration, this statistic may be deceptive.

One limiting factor in auditors' assessment of decoration and the condition of grounds was the extreme heat wave that central North Carolina experienced in August, 2007. In fact, on one of the audit days the temperature topped 104 degrees while we were out rating. This extended period of heat had a definite effect on the grass and vegetation in people's yards, as well as likely reduced the amount of landscaping and yard maintenance that people performed during this period. As a result, many yards were probably not in as good a condition and vegetative decoration (i.e. potted flowers) were likely more sparse. Although auditors were instructed to try to ignore the effects of heat on people's yards during the summer, it was difficult to give an 'excellent' rating to a yard that is dying.

Two additional audit questions (36, & portions of 43) were found to have significant differences in their audit results when applying the kappa standard (< .61). With a .60 kappa score, the auditors' assessment of 'public lighting' just fell below the standard used. There are a number of reasons why travel mode might affect observer's perception of public lighting. One possible explanation concerns the abundance of street trees in the study area. In more heavily wooded areas, auditors rating by car may find it difficult to spot road lighting while moving in traffic. Given the fact that 10 of the 16 recorded differences were instances in which 2007 auditors recorded road lighting to be present, this seems to be a plausible explanation for the discrepancy present in this study. It is also possible that, due to the design and scale of pedestrian lighting fixtures, auditors rating on foot might record more accurately the presence of pedestrian lighting. However, the data results from this study do not support this explanation.

The presence of three types of signage (‘yield to pedestrian’, ‘other pedestrian or bike friendly signs’, and ‘no trespassing’) exhibited low kappa scores. The latter two signage types both showed signs of directionality, with 80% and 71% of the differences occurring due to 2007 auditors recording the presence of these signs. Finally, it is worth noting that while question ’43-0’ does not have extremely low reliability scores, all nine of the recorded differences were cases in which auditors on foot perceived *some* form of signage or traffic control device, where 2006 auditors perceived none. Given the fact that many signs are oriented to pedestrian observers, it makes sense that auditors on foot might capture more of these types of small-scale environmental features.

Table 3 provides a summary of those audit questions whose results were found to be significantly impacted by the travel mode used to administer the audit tool. This table contains only those audit questions whose results met either the percent agreement or kappa standards and were not skewed by limiting factors, such as a small sample size or high variability over time.

Table 3. Summary Table of Significant Differences

	On foot versus by car (same rater, 2006-2007)		
Audit Question	% Agree	Kappa (95% CI)	Directionality (of recorded differences)
(#1) Subjective assess.	0.65	0.5 (0.35, 0.66)	27 of 28 rated lower on foot
(#2) # of res. Units*	0.7	n/a	17 of 24 rated higher on foot
(#4) Cond. of res. Units	0.68	0.39 (0.19, 0.6)	-
(#5) Cond. of grounds**	0.74	0.66 (0.48, 0.83)	-
(#8) Pres. of decoration	0.72	n/a	13 of 16 rated lower on foot
(#9) Presence of border	0.71	0.54 (0.35, 0.73)	11 of 17 rated lower on foot
(#21) Cond. of pub. space	0.7	n/a	16 of 24 rated lower on foot
(#28_2) Pres. of Alc. litter	0.63	0.28 (0.08, 0.48)	15 of 15 rated 'present' on foot
(# 35) Trees shading	0.7	0.58 (0.43, 0.72)	-
(#36) Public lighting	0.8	0.6 (0.42, 0.77)	10 of 16 : road lighting recorded on foot
(#43_5) Yield to ped…	0.94	0.51 (0.14, 0.88)	-
(#43_7) Other ped./bike	0.87	0.32 (0.0, 0.63)	8 of 10 rated as 'present' on foot
(#43_14) No trespassing	0.91	0.42 (0.06, 0.77)	5 of 7 rated as 'present' on foot

*possible error in data results
**recoded values used

The results of certain audit questions clearly vary over time, in some cases changing by daily or even hourly. For this reason, 4 audit questions with low percent agreement or kappa scores were previously removed from those questions included in Table 3. However, in order to be confident that low percent agreement and kappa scores are truly due to travel mode and not the 1-year time delay that took place between audit sessions in this study, an additional test is required.

The design of this study made it somewhat difficult to test the influence of time on audit question results. Since one rater stayed the same between audit sessions, the audit results from this study cannot be directly compared to previous audit sessions. Thus, in order to test the possible influence of a year long time delay, PIN III audit results from 2005 and 2006 audit sessions were compared for those questions contained in Table 3. Specifically, kappa scores from ‘same-day’ Inter Rater Reliability (IRR) testing in 2006 was compared to the 1-year IRR testing conducted using both 2005 and 2006 data. In order to confidently say that audit results changed due to the time delay between audit sessions, only those questions whose confidence intervals do not overlap when comparing same-day and 1-year data were considered significantly different. Table 4 displays the results of this testing.

Table 4. Influence of Time on Audit results

	Same Day IRR (2006)		1-year IRR (2005-2006)	
Audit Question	% Agree	Kappa	% Agree	Kappa
(#1) Subjective assess.	0.56	0.65 (0.58, 0.72)	0.69	0.42 (0.34, 0.51)
(#2) # of res. Units	-	-	-	-
(#4) Cond. of res. Units	0.73	0.77 (0.68, 0.86)	0.95	0.55 (0.45, 0.66)*
(#5) Cond. of grounds	0.73	0.59 (0.45, 0.72)	0.85	0.40 (0.28, 0.53)
(#8) Pres. of decoration	0.91	0.70 (0.53, 0.88)	0.94	0.41 (0.17, 0.66)
(#9) Presence of border	0.49	0.59 (0.49, 0.69)	0.71	0.414 (0.30, 0.53)
(#21) Cond. of pub. space	0.68	0.50 (0.30, 0.70)	0.92	0.48 (0.36, 0.60)
(#28_2) Pres. of Alc. litter	0.66	0.63 (0.45, 0.81)	0.87	-0.04 (-0.23, 0.16)
(# 35) Trees shading	0.7	0.63 (0.55, 0.72)	0.75	0.54 (0.45, 0.63)
(#36) Public lighting	0.96	0.87 (0.81, 0.93)	0.96	0.77 (0.70, 0.85)
(#43_5) Yield to ped…	0.98	0.81 (0.63, 0.99)	0.98	0.79 (0.59, 0.99)
(#43_7) Other ped./bike	0.92	0.71 (0.53, 0.89)	0.96	0.36 (0.14, 0.57)
(#43_14) No trespassing	0.96	0.85 (0.71, 0.98)	0.98	0.66 (0.48, 0.85)

*possible error in data results (for 1-yr IRR, * denotes unbalanced kappa table)

Using this method, 2 questions showed signs of being influenced by a year long time delay: ‘condition of residential units’ and ‘presence of alcoholic litter’. However, after taking a closer look, neither of these questions was conclusively found to be affected by the time delay. It was already known that alcoholic litter can change dramatically over time. The reason that this question’s audit results were still considered important was due to the strong directionality of its recorded differences. Furthermore, the kappa table from the 2005-2006 1-year data for ‘condition of residential units’ was unbalanced, and thus cannot reliably be used. Finally, it is important to note that while kappa scores were lower for the 1-year IRR data compared to same-day IRR data, percent agreement scores displayed the opposite trend for each of the questions contained in Table 3. Thus, there is no conclusive evidence that time delay, rather than the change in travel mode, played a significant role in the observed differences of this study's audit results.

Discussion

The data results reveal a number of audit questions whose low reliability scores indicate the strong influence of travel mode on observer perception. For the most part, theoretical explanations support the apparent discrepancies in audit results between the two audit sessions. The ability of this study to identify and quantify environmental features which evidence suggest may be perceived and thus evaluated differently while traveling by different modes represents a significant contribution to the continued study of the built environment and human perception. However, additional research into the effect which travel mode has on audit results is still required to better understand the strength and breadth of this connection.

This study was subject to a number of unavoidable limitations which somewhat constrain our ability to generalize the data results obtained. A number of these limitations have been addressed: the relatively small sample size, the homogeneity of the study area, and the possible influence of a heat wave. Future research efforts would benefit greatly by simply avoiding the limitations present in this study, in order to build upon its results. A more extensive study area would allow researchers to compute kappa statistics for virtually all of the audit questions, and might also capture a sufficient number of rare environmental features which may be perceived differently due to the travel mode of the observer.

Another limitation of the study design is that only one of the two auditors had previously evaluated the same audit segments. Thus, there is a small amount of variation in audit results that could be attributed to having one of the two auditors change between audit sessions. Due to the number of selection criteria in place, selecting only those audit segments which had been previously assessed by both auditors would have led to an insufficient sample size. If future research were able to use an identical audit pair to re-evaluate identical audit segments, it would lend even more credit to the results obtained in this study.

A number of past researchers cite auditor fatigue as one possible cause of low reliability scores, and stress the importance of using auditors who are capable of and enjoy walking for extended periods (Pikora et al., 2002; Hoehner et al., 2006). Since both participants had previous experience auditing, this was not a factor in this study. However, another potentially limiting factor that was not discussed in any of the articles reviewed is the possible effect of mental fatigue. By 'mental fatigue', I am referring to the tendency to grow tired of repeating the same exercise over an extended period of time. It would be interesting to test whether auditors rate more accurately early in the day, or earlier in the study period. For example, do participants become tired of using the audit tool after the first month? While the benefits of additional experience might mask the potentially negative affects of mental fatigue, it would be beneficial to determine if this factor exists none the less. If mental fatigue indeed exists, it could be an additional explanation for low reliability scores obtained in this study. Given the short time frame and experience level of 2007 raters, their audit results would have been less influenced by mental fatigue, compared to 2006 raters.

Conclusion

This study has identified a number of environmental features which data suggests are perceived differently by observers using different travel modes. Thus, despite the real and potential limitations present in the study's design, I believe this study to be a success. Though the degree to which travel mode influences audit results will naturally vary depending on the design and context of the audit, the results of this study makes it clear that the impact of travel mode is not something which can in good conscience be ignored. Much like many other aspects of neighborhood audit design and structure, the choice of travel mode by which an audit is conducted represents a tradeoff. For most audit questions, travel mode does not appear to have a significant impact. In these cases, the ease and speed of conducting a windshield survey may be preferred. However, as this study reveals, certain environmental features are clearly assessed differently when rating on foot versus by car. While the study results cannot prove that on-foot observations are more valid or accurate, the nature of recorded differences for certain audit questions suggests that this may be the case for specific environmental features.

The completion of this study represents an important step in the continued improvement of neighborhood audits and the study of the built environment. It is my hope that future audit development will take into consideration the results of this study and that additional research will be performed to more fully capture the impact which travel mode has on one's perception of the built environment. By better understanding this connection, we can gain a clearer picture of how the built environment affects physical activity and take further steps in promoting walking as a leisure activity, as well as a more viable transportation alternative.

References

Ball, K., Bauman, A., Leslie, E., Owen, N, 2001. Perceived environmental aesthetics and convenience and company are associated with walking for exercise in Australian adults. Prev. Med. 33, 434-440.

Bauman, A., Smith, B., Stoker, L., Bellew, B., Booth, M., 1999. Geographical influences upon physical activity participation: evidence of a "coastal effect." Aust N Z J. Public Health. 23, 322-324.

Bedimo-Rung, A., Mowen, A., Cohen, D., 2005. The significance of parks to physical activity and public health. A conceptual model, *Am J Prev Med*. **28** (2S), 159–168.

Berrigan, D., Troiano, P., 2002. The association between urban form and physical activity in U.S. adults. Am. J. Prev. Med. 23 (2S).

Brownson, R., Baker, E., Houseman, R., Brennan, L., Bacak, S., 2001. Environmental and policy determinants of physical activity in the United States. Am. J. Pub. Health. 91, 1995-2003.

Brownson, R.C., Boehmer, T.K, Luke, D.A., 2005. Declining rates of physical activity in the United States: What are the contributors? Annu. Rev. Pub. Health. 26, 421-443.

Caspersen, C., and Merritt, R., 1995. Physical activity trends among 26 states, 1986–1990, *Med Sci Sports Exerc*. 27, 713–720.

Caughy, M., O'Campo, P., and Patterson, J., 2001. A brief observational measure for urban neighborhoods. Health & Place. 7 (3), 225-236.

Clifton, K., Smith, A., Rodriguez, D., 2006. The development and testing of an audit for the pedestrian environment. Landscape and Urban Planning. 80, 95-110.

Cohen, D., Spear, S., Scribner, R., Kissinger, P., Mason, K., and Wildgen, J., 2000. "Broken Windows" and the risk of gonorrhea. Am. J. Pub. Health. 90 (2), 230-236.

Crane, R., 1996. On form versus function: will the new urbanism reduce traffic, or increase it? J. Plan. Educ. Res. 15, 117-126.

Craythron, E., 1997. Checklist for footpaths, walkways and cyclepaths. Center for Disease Prevention and Health Promotion, New South Wale Health Department.

Day, K., Boarnet, M., Alfonzo, M., Forsyth, A., 2006. The Irvine-Minnesota Inventory to Measure Built Environments. Am. J. Prev. Med. 30 (2), 144-152.

Emery, J., Crump, C., Bors, P., 2003. Reliability and validity of two instruments designed to assess the walking and bicycling suitability of sidewalks and roads. Am. J. Health Promot. 18 (1).

Ewing, R., Schmid, T.L., Killingsworth, R., Zlot, A., Raudenbush, S., 2003. Relationship between urban sprawl and physical activity, obesity, and morbidity. Am. J. Health Promot. 18, 47-57.

Handy, Susan L., Boarnet, Marlon G., Ewing, Reid, Killingworth, Richard E., 2002. How the built environment affects physical activity. Am. J. Prev. Med. 23, (2S), 64-79.

Hoehner, C., Ivy, A., Ramirez, L., Meriwether, B., Brownson, R., 2006. How reliably do community members audit the neighborhood environment for its support of physical activity? Implications for participatory research. Journal of Public Health Management and Practice. 12 (3), 270-277.

Hoehner, C., Ramirez, L., Elliot, M., Handy, S., Brownson, R., 2005. Perceived and objective environmental measures and physical activity among urban adults. Am. J. Prev. Med. 28 (2) (2S), 105-116.

Giles-Corti, B., Donovan, R., 2002. The relative influence of individual, social and physical environment determinants of physical activity. Soc. Sci. Med. 54, 1793-1812.

Golledge, R., Simson, R., 1997. Spatial behavior: a geographic perspective. New York: Guilford Press.

Greenwald, M., Boarnet, M., 2002. The built environment as a determinant of walking behavior analyzing non-work pedestrian travel in Portland, Oregon, *Transp Res Rec.* 1780, pp. 33–42.

Kirtland, K., Porter, D., Addy, C., Neet, M., Williams, J., Sharpe, P., Neff, L., Kimsey, D., Ainsworth, B., 2003. Environmental Measures of Physical Activity Supports: Perception Versus Reality. Am. J. Prev. Med. 24 (4), 323-330.

Krizek, K., 2003. Operationalizing neighborhood accessibility for land use-travel behavior research and regional modeling. *J Plann Educ Res* **22**, pp. 270–287.

Landis, J., Koch, G., 1977. The measurement of observer agreement for categorical data. Biometrics. 33, 159-174.

Lloyd, R., 1997. Spatial cognition: geographic environments. Norwell MA: Kluwer Academic Publishers.

Mokdad, A.H., Ford, E.S., Bowman, B.A., 2001. Prevalence of obesity, diabetes, and obesity-related health risk factors. JAMA. 289, 76-79.

Owen, N., Humpel, N., Leslie, E., Bauman, A., Sallis, J., 2004. Understanding environmental influences on walking. Am. J. Prev. Med. 27 (1), 67-76.

Pedersen, D., 1997. Effects of size of home town on environmental perception. Perceptual Motor Skills. 45, 955-966.

Pikora, T., Bull, F., Jamrozik, K., Kniuman, M., Giles-Corti, B., Donovan, R., 2002. Developing a reliable audit instrument to measure the physical environment for physical activity. Am. J. Prev. Med. 23 (3), 187-194.

Raudenbush, S.W., and Sampson, R.J., 1999. Ecometrics: toward a science of assessing ecological settings, with applications to the systematic observation of neighborhoods. Sociologiacl Methodology 29, 1-41.

Sallis, J., Hovell, M., Hofstetter, C., et al., 1990. Distance between homes and exercise facilities related to frequency of exercise among San Diego residents. Pub. Health Rep. 105, 179-86.

Siegel, P., Brackbill, R., Heath, G.W., 1995. The epidemiology of walking for exercise: implications for promoting activity among sedentary groups. Am. J. Pub. Health. 85, 706-710.

Troped, P., Saunders, R., Pate, R., Reininger, B., Ureda, J., Thompson, S., 2001. Associations between self-reported and objective physical environmental factors and use of a community rail-trail. Am. J. Prev. Med. 32, 191-200.

UNC Carolina Population Center (2008). PIN III Study. Retrieved March 12th, 2008 from the World Wide Web: http://www.cpc.unc.edu/projects/pin/design_pin3.html

Wentworth, W., Nager, A., 1976. Bryant Park: a comprehensive evaluation of its image and use. New York: City University of New York Environmental Psychology Program.

Zenk, S., Schulz, A., Mentz, G., House, J., Clarence, G., Miranda, P., Miller, P., Kannan, S., 2007. Inter-rater and test-retest reliability: Methods and results for the neighborhood observational checklist. Health & Place 13 (2), 452-465.