Please cite as (APA 7th):
Sas, M., Snaphaan, T., Pauwels, L. J. R., Ponnet, K., & Hardyns, W. (2022). Using Systematic Social Observations to Measure Crime Prevention through Environmental Design and Disorder: In-situ Observations, Photographs, and Google Street View Imagery. *Field Methods.* https://doi.org/10.1177/1525822X221074768

Using Systematic Social Observations to Measure CPTED and Disorder: In-Situ Observations, Photographs and Google Street View Imagery

Marlies Sas[1], Thom Snaphaan[2], Lieven J. R. Pauwels[3], Koen Ponnet[4], Wim Hardyns[5]

Abstract

This study focuses on the use of systematic social observations (SSO) to measure crime prevention through environmental design (CPTED) and disorder. To improve knowledge about measurement issues in small area research, SSO is conducted by means of three different methods: in-situ, photographs, and Google Street View (GSV) imagery. By evaluating the methodological quality of the observation methods, the results of our study suggest that virtual SSO approaches have considerable promise for the reliable assessment of physical properties of small areas. We discuss challenges and provide avenues for future research to encourage the evolution of a more reliable approach to measure the physical environment.

[1] Research Group on Law Enforcement, Faculty of Law and Master of Safety Sciences, Faculty of Social Sciences, University of Antwerp, Antwerp, Belgium and Institute for International Research on Criminal Policy (IRCP), Department of Criminology, Criminal Law and Social Law, Ghent University, Ghent, Belgium.
[2] Institute for International Research on Criminal Policy (IRCP), Department of Criminology, Criminal Law and Social Law, Ghent University, Ghent, Belgium.
[3] Institute for International Research on Criminal Policy (IRCP), Department of Criminology, Criminal Law and Social Law, Ghent University, Ghent, Belgium.
[4] Department of Communication Sciences, imec-mict-Ghent University, Ghent, Belgium.
[5] Institute for International Research on Criminal Policy (IRCP), Department of Criminology, Criminal Law and Social Law, Ghent University, Ghent, Belgium and Master of Safety Sciences, Faculty of Social Sciences, University of Antwerp, Antwerp, Belgium.

Introduction

Accurate and reliable measurement of the characteristics of a neighborhood is a prerequisite to any consideration of the issues an area faces and how they might be solved. This study explores the challenges of measuring the physical properties of neighborhoods. The issue of neighborhood disorder has been extensively studied because of its impact on many aspects of daily life, including mental and physical health (Feng et al. 2010; Hill and Angel 2015), fear of crime (Kelling and Coles 1996; Perkins and Taylor 1996) and violent victimization (Morenoff et al. 2001; Sampson and Raudenbush 2001). The crime prevention through environmental design (CPTED) approach assumes that also other characteristics of the urban environment, such as surveillance (e.g., security guards, shopkeepers), access control (e.g., detection mechanisms), territoriality (e.g., fences, property signs), maintenance (e.g., the condition of building), and activity support (e.g., parks, shops, bars) are related to crime (Crowe 2000). Researchers have traditionally used administrative data (Raudenbush and Sampson 1999) or (representative) victim/self-reported delinquency surveys (Gracia et al. 2012; Ross and Mirowsky 2009) to measure neighborhood characteristics. While both methods have their strengths, they are more useful in assessing the *perceived* rather than the *actual* presence of visible neighborhood phenomena of social disorder such as public drinking or loitering, and physical disorder such as vacant buildings or dilapidation. As a consequence, neighborhood-level studies frequently use systematic social observation (SSO) by trained observers (e.g., Mastrofski 1998; Raudenbush and Sampson 1999), particularly for measuring disorder (Hinkle and Yang 2014; Hoeben et al. 2018; Johnson et al. 2016).

Using SSO eliminates well-known threats to survey validity, such as social desirability, unit nonresponse and sampling error (for an overview of validity problems in survey research, see Rossi et al. 2013; Wolf et al. 2016). While in-situ SSO can provide rich details about the presence of disorder and of CPTED measures, issues such as the cost of travel for observers or fear of crime when working in less-secure neighborhoods may arise (Griew et al. 2013; Rundle et al. 2011). These important theoretical, epistemological, and methodological concerns have led to a search for more innovative techniques with which to conduct research in neighborhood settings, such as virtual instead of in-situ SSO, often carried out using Google Street View (GSV) imagery (Ben-Joseph et al. 2013 Clarke et al. 2010; Rundle et al. 2011) or photographs taken at the location, for the purposes of research (e.g., Cannuscio et al. 2009; Yang and Pao 2015).

Previous studies have shown that virtual SSO using GSV is a reliable, time-saving, and cost-effective way of assessing environmental aspects (Ben-Joseph et al. 2013; Curtis et al. 2013; Griew et al. 2013; Odgers et al. 2012). To date, researchers have examined interrater reliability between virtual observers (Kelly et al. 2013), and the reliability between physical SSO and GSV (Ben-Joseph et al. 2013; Clarke et al. 2010; Rundle et al. 2011). In the current study, we evaluate the methodological quality of in-situ observations, photographs taken at the location, and GSV imagery in small-scale areas. Given the potential strengths and limitations associated with in-situ and virtual SSO, we used different measurement approaches to evaluate their interrater reliability, inter-modus reliability, and convergent validity.

Interrater reliability refers to the extent to which different observers give the same rating to an item. Inter-modus reliability refers to the extent to which the same results are obtained using different methods. Convergent validity is the extent to which theoretically linked measures ought to correlate. A significant disadvantage is that in-situ observations, photographs, and GSV imagery always capture a moment in time. Temporal variance is an obvious challenge, not only for measuring social disorder, but also for linking in-situ observations to virtual observations. Temporal variance should therefore be considered a potential measurement error when evaluating SSO (Oberwittler 2004; Sampson and Raudenbush 1999). Additionally, a multi-method approach is likely to have potential mode effects or measurement and selection effects. Measurement effects occur when the results of the same items differ across modes (De Leeuw 2005; Dillman et al. 2009). Selection errors occur when the observation points of individual modes differ on the variables of interest (Vannieuwenhuyze et al. 2012).

To guide our analyses and explain our expected results, we postulate six predefined hypotheses:

H1: Due to the extensive training of observers, we hypothesize that we will not find significant differences in interrater reliability between the different methods.

H2: Due to temporal incongruences, and especially because of the prevailing measures against COVID-19 at the time[1] of the in-situ observations and when the photographs were taken, we hypothesize that we will find significant differences between observations from GSV imagery and the other two methods.

H3: Due to temporal congruences, we hypothesize that we will not find significant differences between observations from photographs and in-situ observations.

H4: We hypothesize that constructs measured by using in-situ observations and photographs are stronger related than measurements considered in conjunction with GSV imagery, because these are captured at the same moment in time.

H5: We hypothesize that there will be more agreement across the different modes on the CPTED measures than on the measures regarding physical and social disorder, expecting that the CPTED measures are more stable over time.

H6: We hypothesize that there will be more agreement across the different modes on the measures of physical disorder than on the measures of social disorder, expecting that physical disorder is more stable over time.

Data and Methods

Data Collection

Research area. In the current study, SSO was applied within the neighborhood of the campuses of the University of Antwerp. The city of Antwerp consists of approximately 526,000 inhabitants and is, in terms of population, the largest city in Belgium. The university has four main campuses in four different neighborhoods. University neighborhoods serve as an interesting research setting, as they can be considered an attractive environment for motivated offenders to engage in crime. The combination of crime attractors, such as the presence of intoxicated students or insecure housing largely vacant during the day, and crime generators, such as the presence of bars and shopping venues, creates conditions conducive to opportunities for crime (Cundiff 2021). The selection of the observation points was based on three criteria, taking into account both physical and virtual characteristics:

1. Only observation points with direct access to the university campus were selected, to ensure that they were all located in the university neighborhood.
2. Only observations points that are at least 100 m apart from each other were selected, to avoid overlapping observations.
3. Only observation points of which the GSV imagery was recently captured (after May 2017) were selected, to address the limitation of temporal variance.

In total, 40 observation points, indicated by (x,y) coordinates, complied with the selection criteria. Taking into account the number of available observers, 36 observation points were randomly selected by the observers. Observers who carried out in-situ observations were instructed to carry out 360° observations from the exact (x,y) coordinates they received. From the observation point, they had to take four photographs that offered a clear overview of the area around the observation point (see Appendix, Figure 2). These photographs were later used by other observers to carry out a photograph-based analysis. The in-situ observers were free to choose a date and time to carry out the observations between October 1 and October 23, 2020. In the end, in-situ observations were carried out between 10 a.m. and 8.15 p.m. from October 3 to October 20.

Observers. Trained observers[2] (n=203) were able to subscribe into groups of four to six students. In the end, 36 groups were formed. Before they carried out the observations, they all underwent extensive training in how to use the checklist and the practical performance of the observations. They were also given theoretical information about crime prevention, disorder, and CPTED.

Each group was given an observation point with an address and specific (x,y) coordinates. Two people from each group traveled to the observation point to carry out the in-situ observations and take the photographs, and the remaining group members subsequently made their observations from the same x,y point using these photographs. The observers were free to choose which group members conducted which observations. In addition, both groups (those that observed in-situ and those that observed using the photographs) were also instructed to carry out the same observations from the same point using GSV imagery. In total, 174 females, 28 males, and one non-binary observer carried out two observations each, which led to 406 individual observations in total. Seventy-one observations were performed in situ, 132 used photographs, and 203 were virtually carried out using GSV imagery.

Materials and measures. All observations were carried out using a standardized checklist. The observation checklist consisted of 40 items to assess the presence of CPTED measures and social and physical disorder (see Appendix, Table 1). To examine CPTED, the five key principles (surveillance, access control, territoriality, maintenance, and activity support) were included in the checklist. **Surveillance** was measured by seven items (e.g., presence of formal surveillance (police patrol)). For **access control**, three items were included (e.g., there are visible campus entrances and exits). **Territoriality** was measured by three items (e.g., there are physical barriers (e.g., fences, shrubs) that separate campus from public space). All these key principles were scored on a scale from 0 (= none (=0)) to 3 (= many (>4)). Nine items were used to measure **maintenance** (e.g., in general, the observation point appears to be well maintained in terms of infrastructure; the condition of the buildings at the observation point is …). Items were scored on a scale from 1 (= totally not agree) to 4 (= totally agree) or on a scale from 1 (= very bad) to 4 (= very good). Finally, **activity support** was assessed by six items (e.g., cultural activity is taking place (e.g., a performance, a festival) at the observation point), which were scored on a scale from 0 (= none (=0)) to 3 (= many (>4)).

Drawing on previous research (Covington and Taylor 1991; Pauwels and Hardyns 2009; Steenbeek and Kreis 2015), the presence of physical and social disorder was measured by including, respectively, eight and four items in the checklist. For **physical disorder**, a distinction was made in the checklist between the cleanliness of the observation point (e.g., there are waste bins at the observation point) and the presence of damaged infrastructure (e.g., there are broken windows/doors at the observation point). Answers were scored based on a scale ranging from 0 (= none (=0)) to 3 (=many (>4)). For **social disorder**, the presence of people who might disrupt public order (e.g., there are people who are drunk at the observation point) was examined, based on a scale ranging from 0 (=none (=0)) to 3 (= many (>4)).

Analytical strategy. SPSS Statistics (version 26) was used for all analyses. The intra-class correlation (ICC) coefficient was used to assess interrater reliability. The ICC is represented in equation 1, where σ^2 represents the variance and B and W respectively stand for between and within groups (Heck et al. 2010).

$$\rho = \sigma_B^2/(\sigma_B^2 + \sigma_W^2) \quad (1)$$

Correlational analyses were used to assess both inter-modus reliability and convergent validity. To examine inter-modus reliability, the correlation per modus and per observation point was assessed for all items together. In these analyses, the dataset is transposed to the extent that the concepts related to CPTED and disorder are represented as cases. To examine convergent validity, the mean per observation point was compared over the different modes and for the different concepts related to CPTED and disorder.

Results

Interrater Reliability

ICC coefficients (one-way random, single measures) were computed for the three observation methods. The results (see Appendix, Figure 3) show that the variation in ICCs per observation point was higher, and highly similar, for the observations done using photographs (M_{photo}) and GSV imagery (M_{GSV}), compared with the in-situ observations ($M_{in\text{-}situ}$). The average ICCs are 0.969 (SD=0.036) for in-situ observations, 0.888 (SD=0.070) for observations via photographs and 0.881 (SD=0.057) for observations via GSV imagery. These averages indicate strong levels of agreement between the raters. The highest ICCs were found for $M_{in\text{-}situ}$, which are also characterized by small variations. The lowest ICCs, but still high, were obtained for M_{photo} and M_{GSV} and show larger variations.

These results suggest that the findings are in line with the first hypothesis. Although the ICCs differed by observation method, in particular regarding $M_{in\text{-}situ}$ which yielded higher interrater reliability, the average ICC of each method was high. Considering these ICCs, very similar results were obtained by different observers using the same observation method.

Inter-modus Reliability

The inter-modus reliability of the three observation methods is shown in Figure 1. This reliability measure refers to correlations between the different methods used for the observations of all 40 items that were considered in this study (see Materials and measures). Average Spearman rho correlations range from .882 to .934, indicating a strong inter-modus reliability. Low variations and the highest correlations were found between M_{photo} and M_{GSV} (mean r=.934, SD=.041). This finding reveals that there is a strong agreement between the two virtual observation methods. Similar variations and correlations were obtained between $M_{in\text{-}situ}$ and M_{GSV} (mean r= .894, SD=.045) and between $M_{in\text{-}situ}$ and M_{photo} (mean r=.882, SD=.069). Although these correlations are not as high as the correlations between M_{photo} and M_{GSV}, they still can be considered as high.

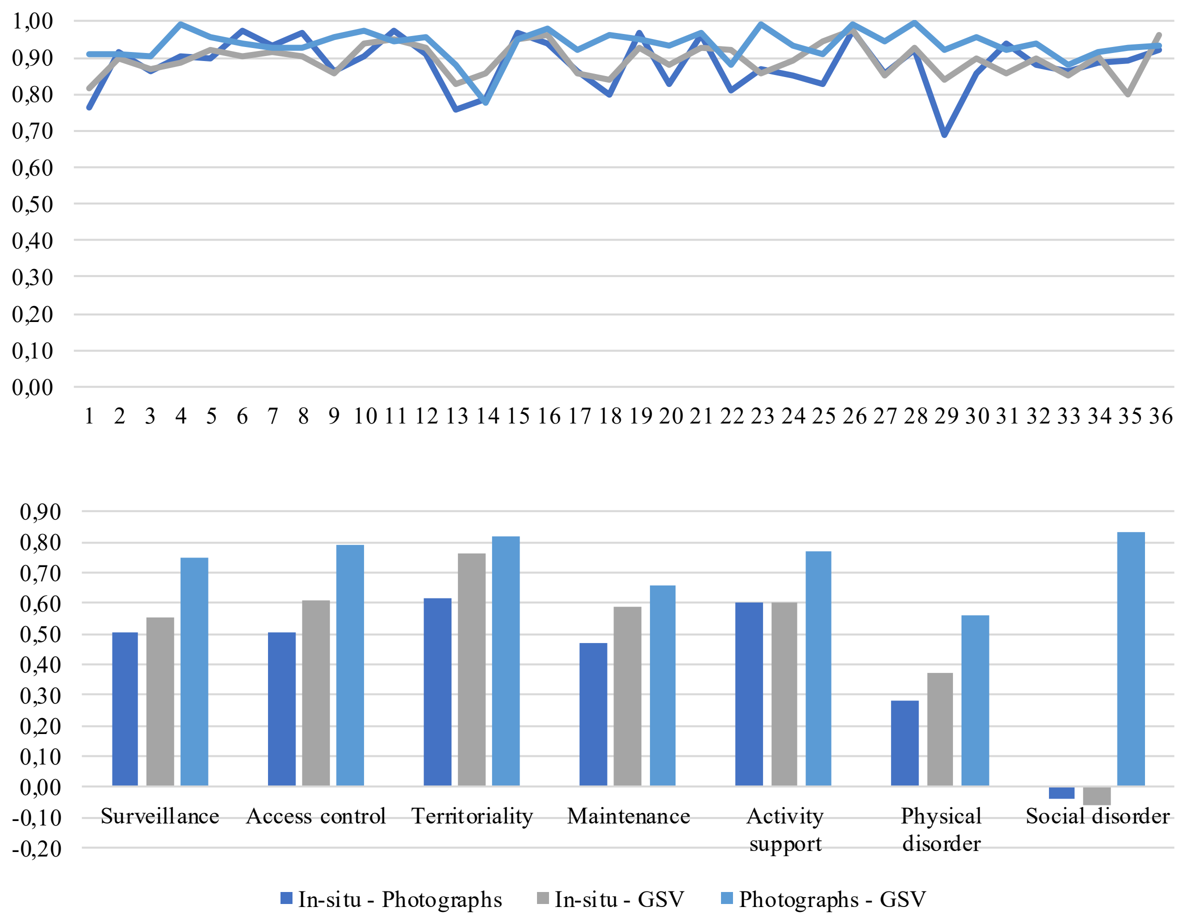

Figure 1. Above: Correlation of observations between different methods per observation point. Below: Average correlation of observations between different methods per construct.

These results show that the findings are not in line with the second hypothesis, as M_{photo} and M_{GSV} yield very high correlations. These results are not in line with the third hypothesis. The correlations between $M_{in\text{-}situ}$ and M_{photo} are high; even the lowest correlation on the location level (r=.799 for location 35) can be considered high. However, it should be noted that correlations between the two virtual methods are higher than those with the in-situ observations, which indicates that virtual observations may lead to more similarity than in-situ and virtual observations that have been taken at the same location at the same time.

Convergent Validity

The convergent validity was first investigated by comparing the average scores per observed measure over the three observation methods (see Appendix, Figure 4). The results indicate that the average scores were highest for $M_{in\text{-}situ}$. Except for the constructs maintenance and social disorder, higher average scores are found compared to M_{photo} or M_{GSV}. It should, however, be noted that although social disorder items were included in the observation checklist, they were not observed during the in-situ observations (i.e., social disorder was largely absent). Subsequently, although M_{GSV} yields higher scores for all constructs on average, the results show that the mean scores on the constructs observed via M_{GSV} and M_{photo} are comparable.

Figure 1 reveals the average correlations between the different observation methods per group of items (i.e., constructs). The results show the highest correlations between M_{photo} and M_{GSV} (mean r=0.709, SD=0.197). Lower correlations are found between $M_{in\text{-}situ}$ and M_{photo} (mean r=0.458, SD=0.239) and $M_{in\text{-}situ}$ and M_{GSV} (mean r=.535, SD=0.257). Our findings are not in line with hypothesis 4 as average correlations between the two virtual methods, M_{photo} and M_{GSV}, are higher compared to correlations between $M_{in\text{-}situ}$ and the two virtual observations, even while this comparison is not subject to temporal incongruences.

To evaluate the agreement across the different modes, we computed the standard deviations (SD) of the average correlations for the three methods per construct (see Appendix, Table 2). The results show that the SD range from .107 to .409. The highest agreement is found for the construct territoriality, while the lowest agreement is obtained for social disorder. Importantly, regarding social disorder it should be noted that three of the four items were not included, because the scores of $M_{in\text{-}situ}$ were nil.

Our findings are in line with hypothesis 5, because stronger agreement was found across the different modes for CPTED measures compared to physical and social disorder. But caution is warranted because only one of four items was included for social disorder. Our findings also align with hypothesis 6 as there is more agreement across the different modes on the measures of physical disorder compared to the measures of social disorder. However, we were not able to compare the observations across modes, as no social disorder was observed during the in-situ observations.

Discussion and Conclusion

This article has evaluated the methodological quality of three different observation methods in order to improve our knowledge of measurement issues in small area research. To assess the presence of CPTED and disorder at street segment level, observations were carried out in-situ, via photographs, and via GSV imagery. The potential strengths and weaknesses of physical and virtual SSO was assessed by evaluating their interrater reliability, inter-modus reliability, and convergent validity. In general, our findings support the use of photographs and GSV imagery as reliable and cost-effective tools for gathering information about the presence of physical cues in a neighborhood. Although there are single measures to assess, for example, scale reliability (internal consistency, by means of Cronbach's α[3]), such a single measure to evaluate the appropriateness (in terms of reliability and accuracy) of measurement methods does not exist. Therefore, studies like this one are important and highly needed for a reliable and accurate measurement of the physical environment.

High levels of interrater agreement were found, indicating that similar results were obtained by different observers for each of the three observation methods. An evaluation of inter-modus reliability showed that there was strong agreement between in-situ observations and photographs, and in-situ observations and GSV imagery. A higher correlation was obtained between the two virtual observation methods, indicating that there is stronger agreement between observations based on secondary material. This indicates that virtual observations may lead to more similarity than in-situ and virtual observations that have been taken at the same moment in time. Overall, the items measured using in-situ observations were characterized by the highest mean scores. In addition, more agreement was found between CPTED items compared to disorder, while physical disorder items showed a stronger agreement in comparison to social disorder. Based on these findings, despite high correlations and agreement, we can conclude that in-situ observations, which gave the best measurement of the actual scores, observed on average more disorder and CPTED measures. The virtual methods (photographs and GSV) missed an equal number of aspects (as well as the same aspects) of reality, due to the specific characteristics of these methods (e.g., objects obstructing the view on the image, limited view due to the specific position of observation, fixed evaluation of a static image, time spent recording).

Although our study provides valuable insights, the results must be interpreted in light of a number of limitations and challenges that should be taken into account in future research. The first challenge relates to the way the study design affects interrater reliability. Although interrater reliability for all three methods proved to be high, the variation in the number of observers may have contributed to these results: 71 observers carried out in-situ observations, while 132 and 203 observers respectively carried out their observations using photographs and GSV imagery. Therefore, the high ICCs for in-situ observations may have been influenced by the lower number of unique observations that were included in the analysis. It is recommended that future research ensures that an equal number of observers use each observation method.

Interrater reliability is a difficult parameter to control. In our study, it is possible that the coherence between the raters was artificially increased due to the different number of observers per observation point. A potential limitation arises because the in-situ observations were almost always carried out by two observers. As a result, even though the observers were instructed to perform the observation independently, they were able to discuss the measures they observed. It is recommended that future studies should ensure that individual observers cannot be influenced by others taking part in the observations. Moreover, the observers in this study were free to choose which group members would conduct the in-situ observations. It is possible that the more motivated students completed the observations on location, which may have affected the higher ICCs for this method. Randomly assigning observers is recommended for future research.

A second challenge relates to the evaluation of the inter-modus reliability. Although strong correlations were found, these results may be influenced by the task setting. Every rater was asked to perform two different observations, one of which had to be done using GSV imagery. Therefore, every rater observed every observation point twice. It is possible that the first observation they carried out may have biased their second observation. It is recommended that future research should ensure that observers are not biased by previous observations of the same point. Additionally, future research could assess these limitations quantitatively by performing cross-classified multilevel models; unfortunately, the number of observations in this study was too low to conduct this analysis.

A third challenge was that social disorder was not observed during the in-situ observations. The prevailing measures against COVID-19 at the time of the observations could have restricted the presence of social incivilities at the observation points. These external factors, which we could not control, may have influenced our results. In addition, the concept of social disorder is particularly subject to time-of-day differences, which makes it more difficult to record via observations made at one moment in time. Also, we did not include any meta-data (e.g., weather conditions, image construction issues) in our measurement approach. Future research should consider these meta-data and time-of-day effects (see also Sampson and Raudenbush 2004).

A fourth challenge concerns convergent validity, which arose because of the temporal incongruence of combining physical and virtual observations. This challenge emerges in various studies (Clarke et al. 2010; Rundle et al. 2011; Taylor et al. 2011). Both physical and social disorder are considered temporally variable items, which may lower the correlation among the measured items over the different methods used. The CPTED measures may be seen as items that are more stable over time. Remarkably, our study showed high correlations between photographs and GSV imagery, indicating that there is a stronger effect of the modus than an actual shift in levels of CPTED and disorder. In other words, the concepts studied seemed to be relatively stable over time. It is therefore recommended that future research examines this spatial stability further. Additionally, consistent with previous studies, convergent validity tended to be lower for disorder compared to CPTED measures. Clarke et al. (2010) explain this finding by stating that characteristics that require a qualitative judgment (e.g., the condition of a buildings) and items that require highly detailed observations at street level (e.g., the presence of litter) may be less easily observed, especially using virtual methods.

The fifth and final challenge refers to the characteristics of GSV imagery. The quality of the imagery is not always optimal, for different reasons, such as the quality of the initial recording, compression of the data, or resolute blurring of the data because of privacy considerations (Google n.d.). Furthermore, GSV imagery is recorded by a car from the road, which means that not all facets of the streets are mapped, due to objects blocking the view at the time of the recording. These aspects in sum mean that it is generally not easy to capture, detect and/or recognize smaller objects via GSV imagery (Aghaabbasi et al. 2018; Rzotkiewicz et al. 2018). In addition, though not applicable for this study, GSV imagery is not available for all geographical regions and/or all the micro places within these geographical regions (Bloch 2020).

Despite the number of limitations and challenges, the results of our study still suggest that virtual SSO approaches have considerable promise for the reliable assessment of neighborhood-level physical properties. Moreover, our study contributes to the few existing studies that have investigated the extent to which virtual observations are complementary to other methods. By doing so, we not only evaluated the quality of observations by means of GSV imagery, but also investigated a second virtual observation method—photographs. In future research, the challenges that our study identified should be considered to keep evolving toward a more valid and reliable measurement approach for the physical environment. Finally, future research could explore replacing human observers by computers (e.g., Kim et al. 2021) because, although the results using human observers are eminently satisfactory, observation remains a labor-intensive process. Exploratory studies have provided promising results regarding the classification of various types of urban disorder (and, by extension, more aspects in public space, such as presence of CCTV) on a micro-level scale (see Snaphaan and Hardyns 2021). In this context, it would be necessary to conduct similar research to test the reliability and validity of human and computer observers instead of different observation methods.

Notes

1. At the time of the observations, the following measures against COVID-19 that are relevant for this study applied: (1) at the University of Antwerp, only 20% of students were allowed to be physically present in the auditoriums, all other students had to follow classes online from home; (2) telework was highly recommended for staff members, several days a week; (3) all bars had to be closed at 11 p.m.; (4) nonorganized gatherings outside were limited to a maximum of four people, except for family members who lived under the same roof; (5) a maximum of three close contacts per month were allowed for everyone; (6) extra enforcement efforts had to be made ensure that the above measures were applied everywhere.
2. All observers were trained Master students in criminological sciences at Ghent University. They received course credit for their observation efforts.
3. While α-values are normally used for attitude scales, one cannot make the same demands on other types, such as behavioral or observational scales (see Sijtsma 2009).

References

Aghaabbasi, M., M. Moeinaddini, M. Z. Shah, and Z. Asadi-Shekari. 2018. Addressing issues in the use of Google tools for assessing pedestrian built environments. *Journal of Transport Geography* 73:185–98.

Ben-Joseph, E., J. S. Lee, E. K. Cromley, F. Laden, and P. J. Troped. 2013. Virtual and actual: Relative accuracy of on-site and web-based instruments in auditing the environment for physical activity. *Health & Place* 19:138–50.

Bloch, S. 2020. An on-the-ground challenge to uses of spatial big data in assessing neighborhood character. *Geographical Review* 110:210–23.

Cannuscio, C. C., E. E. Weiss, H. Fruchtman, J. Schroeder, J. Weiner, and D. A. Asch. 2009. Visual epidemiology: Photographs as tools for probing street-level etiologies. *Social Science & Medicine,* 69:553–64.

Clarke, P., J. Ailshire, R. Melendez, M. Bader, and J. Morenoff. 2010. Using Google Earth to conduct a neighborhood audit: Reliability of a virtual audit instrument. *Health Place* 16:1224–29.

Covington, J., and R. B. Taylor. 1991. Fear of crime in urban residential neighborhoods: Implications of between- and within-neighborhood sources for current models. *The Sociological Quarterly* 32:231–49.

Crowe, T. D. 2000. *Crime prevention through environmental design: Applications of architectural design and space management concepts*. Oxford: Butterworth-Heinemann.

Cundiff, K. 2021. Colleges and community crime: An analysis of campus proximity and neighborhood crime rates. *Crime & Delinquency* 67:431–48.

Curtis, J. W., A. Curtis, J. Mapes, A. B. Szell, and A. Cinderich. 2013. Using Google Street View for systematic observation of the built environment: Analysis of spatio-temporal instability of imagery dates. *International Journal of Health Geographics* 12:1–10.

De Leeuw, E. D. 2005. To mix or not to mix data collection modes in surveys. *Journal of Official Statistics* 21:233–55.

Dillman, D. A., G. Phelps, R. Tortora, K. Swift, J. Kohrell, J. Berck, and B. L. Messer. 2009. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research* 38:1–18.

Feng, J., T. A. Glass, F. C. Curriero, W. F. Stewart, and B. S. Schwartz. 2010. The built environment and obesity: A systematic review of the epidemiologic evidence, *Health Place* 16:175–90.

Google. n.d. Google-contributed Street View imagery policy. https://www.google.com/streetview/policy/ (accessed November 20, 2020).

Gracia, E., M. C. Fuentes, F. García, and M. Lila. 2012. Perceived neighborhood violence, parenting styles, and developmental outcomes among Spanish adolescents. *Journal of Community Psychology* 40:1004–21.

Griew, P., M. Hillsdon, C. Foster, E. Coombes, A. Jones, and P. Wilkinson. 2013. Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. *International Journal of Behavioural Nutrition and Physical Activity* 10:103–9.

Heck, R. H., S. L. Thomas, and L. Tabata. 2010. *Multilevel and longitudinal modeling with IBM SPSS*. New York: Routledge.

Hill, T. D., and R. J. Angel. 2005. Neighborhood disorder, psychological distress, and heavy drinking. *Social Science & Medicine* 61:965–75.

Hinkle, J., and S. M. Yang. 2014. A new look into broken windows: What shapes individuals' perceptions of social disorder? *Journal of Criminal Justice* 42:26–35.

Hoeben, E. M., W. Steenbeek, and L. J. R. Pauwels. 2018. Measuring disorder: Observer bias in systematic social observations at streets and neighborhoods. *Journal of Quantitative Criminology* 34:221–49.

Johnson, D., E. R. Maguire, S. A. Maass, and J. Hibdon. 2016. Systematic observation of disorder and other neighborhood conditions in a distressed Caribbean community. *Journal of Community Psychology* 44:729–46.

Kelling, G. L., and C. M. Coles. 1996. *Fixing broken windows: Restoring order and reducing crime in our communities*. New York: The Free Press.

Kelly, C. M., J. S. Wilson, E. A. Baker, D. K. Miller, and M. Schootman. 2013. Using Google Street View to audit the built environment: interrater reliability results. *Annals of Behavioral Medicine* 45:S108–12.

Kim, J. H., S. Lee, J. R. Hipp, and D. Ki. 2021. Decoding urban landscapes: Google Street View and measurement sensitivity. *Computers, Environment and Urban Systems* 88. https://doi.org/10.1016/j.compenvurbsys.2021.101626 (accessed June 18, 2021).

Mastrofski, S. D. 1998. *Systematic observation of public police: Applying field research methods to policy issues*. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. https://www.ojp.gov/pdffiles/172859.pdf (accessed January 20, 2021).

Morenoff, J., R. J. Sampson, and S. W. Raudenbush. 2001. Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology* 39:517–58.

Oberwittler, D. 2004. A multilevel analysis of neighborhood contextual effects on serious juvenile offending: The role of subcultural values and social disorganization. *European Journal of Criminology, 1*(2), 201–235.

Odgers, C. L., A. Caspi, C. J. Bates, R. J. Sampson, and T. E. Moffitt. 2012. Systematic social observation of children's neighborhoods using Google Street View: A reliable and cost-effective method. *Journal of Child Psychology and Psychiatry* 53:1009–17.

Pauwels, L., and W. Hardyns. 2009. Measuring community (dis) organizational processes through key informant analysis. *European Journal of Criminology* 6:401–17.

Perkins, D. D., and R. B. Taylor. 1996. Ecological assessments of community disorder: Their relationship to fear of crime and theoretical implications. *American Journal of Community Psychology* 24:107.

Raudenbush, S. W., and R. J. Sampson. 1999. Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology* 29:1–41.

Ross, C. E., and L. Mirowsky. 2009. Neighborhood disorder, subjective alienation, and distress. *Journal of Health & Social Behaviour* 50:49–64.

Rossi, P. H., J. D. Wright, and A. B. Anderson, eds. 2013. *Handbook of survey research.* Cambridge, MA: Academic Press.

Rundle, A. G., M. D. M. Bader, C. A. Richards, K. M. Neckerman, and J. O. Teitler. 2011. Using Google Street View to audit neighborhood environments. *American Journal of Preventive Medicine* 40: 94–100.

Rzotkiewicz, A., A. L. Pearson, B. V. Dougherty, A. Shortridge, and N. Wilson. 2018. Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health & Place* 52:240–46.

Sampson, R. J., and S. W. Raudenbush. 1999. Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology* 105:603–51.

Sampson, R. J., and S. W. Raudenbush. 2001. *Disorder in urban neighborhoods: Does it lead to crime?* Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Sampson, R. J., and S. W. Raudenbush. 2004. Seeing disorder: Neighborhood stigma and the social construction of "Broken Windows." *Social Psychology Quarterly* 67:319–42.

Sijtsma, K. 2009. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 74:107.

Snaphaan, T., and W. Hardyns. 2021. Utilizing geo-referenced imagery for systematic social observation of neighborhood disorder. *Computers, Environment and Urban Systems* 90:Article 101691.

Steenbeek, W., and C. Kreis. 2015. Where broken windows should be fixed: Toward identification of areas at the tipping point. *Journal of Research in Crime and Delinquency* 52:511–33.

Taylor, B. T., P. Fernando, A. E. Bauman, A. Williamson, J. C. Craig, and S. Redman. 2011. Measuring the quality of public open space using Google Earth. *American Journal of Preventive Medicine* 40:105–12.

Vannieuwenhuyze, J. T., G. Loosveldt, and G. Molenberghs. 2012. A method to evaluate mode effects on the mean and variance of a continuous variable in mixed-mode surveys. *International Statistical Review* 80:306–22.

Wolf, C., D. Joye, T. W. Smith, and Y. C. Fu, eds. 2016. *The Sage handbook of survey methodology*. New York: Sage.

Yang, S. M., and C. C. Pao. 2015. Do we "see" the same thing? An experimental look into the black box of disorder perception. *Journal of Research in Crime and Delinquency* 52:534–66.

Appendix

Table 1. Descriptive statistics of CPTED and disorder items.

Variable	Mode 1: In-situ		Mode 2: Photographs		Mode 3: Google Street View	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Number of observers	1.97	0.17	3.67	0.53	5.64	0.59
Surveillance						
Formal surveillance	0.21	0.44	0.02	0.07	0.03	0.10
Natural surveillance	2.19	0.56	1.48	0.77	1.93	0.55
CCTV	0.54	0.69	0.28	0.46	0.25	0.36
Icons indicating presence of CCTV	0.39	0.68	0.21	0.38	0.13	0.24
Lighting	2.25	0.73	1.87	0.81	2.05	0.64
Windows	2.32	0.93	2.40	0.76	2.37	0.81
Open spaces	1.14	0.97	1.11	0.82	1.11	0.84
Access control						
Entrances and exits	1.15	0.82	0.92	0.67	1.05	0.60
Icons indicating presence of entrances	0.82	0.83	0.55	0.56	0.61	0.57
Alarm systems	0.49	0.73	0.15	0.29	0.24	0.47
Territoriality						
Physical barriers	1.60	1.22	1.32	1.03	1.43	1.04
Icons indicating presence of university	1.46	0.97	0.85	0.78	0.96	0.68
Icons indicating presence of private properties	0.61	0.66	0.33	0.47	0.45	0.49
Maintenance						
Cleanliness	2.06	0.76	2.21	0.55	2.37	0.42
Infrastructure	1.94	0.67	2.26	0.58	2.21	0.45
Maintenance of buildings	3.16	0.56	3.24	0.57	3.29	0.61
Maintenance of roads	3.42	0.69	3.43	0.54	3.37	0.64
Maintenance of sidewalks	3.22	0.60	3.28	0.62	3.32	0.53
Maintenance of lighting	3.25	0.65	3.11	0.60	3.18	0.60
Maintenance of CCTV	3.11	0.51	2.88	0.68	2.91	0.76
Maintenance of alarm systems	3.15	0.37	2.95	0.90	3.10	0.71
Maintenance of vegetation	3.34	0.66	3.28	0.63	3.40	0.66
Activity support						
Bars and restaurants	0.64	0.59	0.33	0.45	0.45	0.53
Shops	0.18	0.38	0.13	0.29	0.21	0.34
Picnic tables	0.43	0.78	0.19	0.47	0.24	0.47
Cultural activities	0.11	0.40	0.03	0.17	0.04	0.17
Playgrounds	0.04	0.18	0.07	0.24	0.06	0.16
Park	0.64	0.98	0.55	0.86	0.62	0.90
Physical disorder						
Litter	0.89	0.81	0.50	0.62	0.40	0.38
Litter bins	0.65	0.76	0.55	0.60	0.72	0.74
Graffiti	0.49	0.67	0.34	0.59	0.30	0.57
Icons indicating importance of cleanliness	0.14	0.33	0.05	0.15	0.04	0.14
Car wrecks	0.01	0.08	0.01	0.04	0.00	0.03
Bicycle wrecks	0.01	0.08	0.01	0.06	0.02	0.06
Broken windows/doors	0.13	0.40	0.03	0.10	0.04	0.17
Vacant buildings	0.26	0.49	0.10	0.20	0.15	0.24
Social disorder						
Homeless people	0.00	0.00	0.01	0.11	0.01	0.10
Drunk people	0.00	0.00	0.02	0.13	0.05	0.36
People using drugs	0.03	0.17	0.02	0.13	0.05	0.36
People fighting	0.00	0.00	0.01	0.11	0.01	0.10

Item	Standard deviation
CPTED	
Surveillance	.152
Access control	.169
Territoriality	.107
Maintenance	.132
Activity support	.130
Disorder	
Physical disorder	.218
Social disorder	.409

Table 2. Standard deviations of average correlation for the three methods per construct.

Figure 2. Example of a photograph (left) and a GSV image (right) of the same observation point.

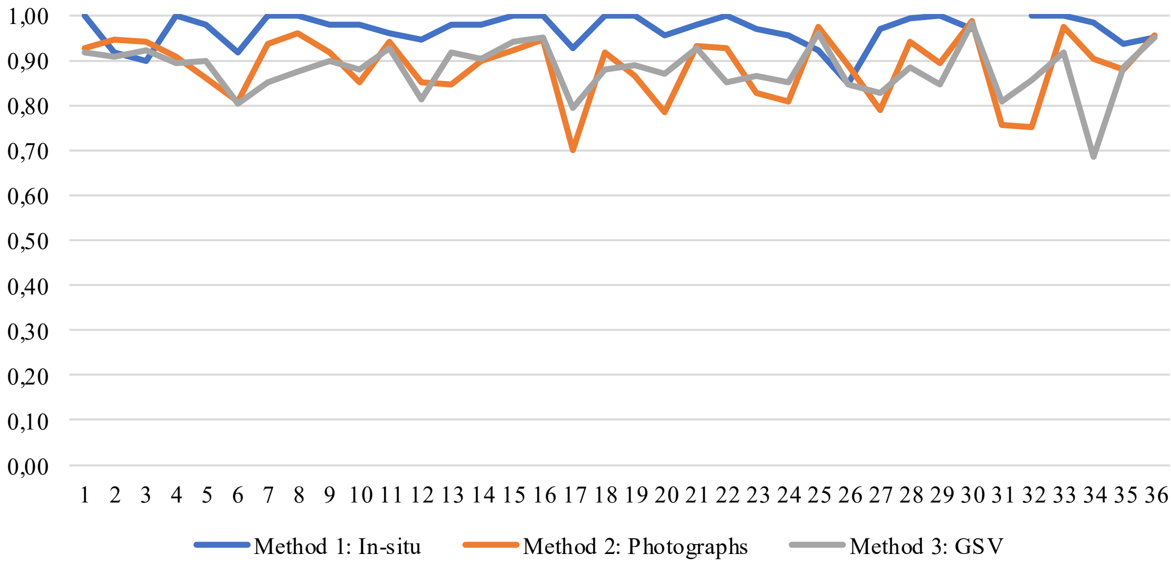

Figure 3. ICC coefficients (single measures) per modus per observation point.[6]

[6] For $M_{in\text{-}situ}$ at location point 31 an ICC coefficient cannot be determined, because there was only one observation for this location in this mode.

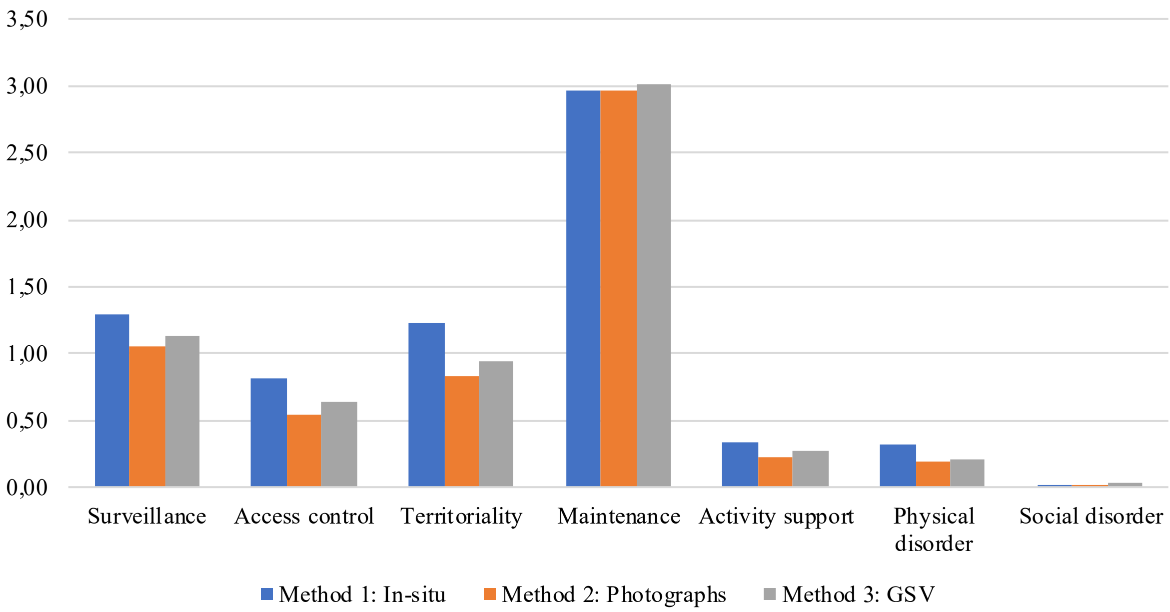

Figure 4. Mean scores for constructs related to CPTED and disorder per method.