Journal of Experimental Criminology
https://doi.org/10.1007/s11292-023-09599-1

The effect of checklists on evidence collection during initial investigations: a randomized controlled trial in virtual reality

Cory P. Haberman[1] · Ming Tang[2] · Clay Driscoll[1] · Bradley J. O'Guinn[1] · Calvin Proffit[1] · J.C. Barnes[1]

Accepted: 20 November 2023

Abstract

Objective To examine the impact of an investigative checklist on evidence collection by police officers responding to a routine burglary investigation.

Methods A randomized control trial was conducted in virtual reality to test the effectiveness of an investigative checklist. Officers in the randomly assigned treatment group (n = 25) were provided with a checklist during the simulated investigation. Officers in the control group (n = 26) did not have access to the checklist at any time. The checklist included five evidence items commonly associated with burglary investigations.

Results Officers who were randomly provided with an investigative checklist were significantly more likely to collect two evidence items located *outside* of the virtual victim's home. Both treatment and control officers were about equally as likely to collect three evidence items located *inside* the residence.

Conclusions Investigative checklists represent a promising new tool officers can use to improve evidence collection during routine investigations. More research is needed, however, to determine whether checklists improve evidence collection or case clearances in real-life settings. Virtual reality simulations provide a promising tool for collecting data in otherwise difficult or complex situations to simulate.

Keywords Investigations · Burglary · Checklists · Policing · Experiment · Randomized controlled trial

✉ Cory P. Haberman
cory.haberman@uc.edu

[1] School of Criminal Justice, University of Cincinnati, PO Box 210389, 660H Teachers-Dyer, Cincinnati, OH 45221-0389, USA

[2] School of Architecture and Interior Design, University of Cincinnati, 7215 College of DAAP, Cincinnati, OH 45221-0389, USA

Introduction

According to the FBI's Uniform Crime Report, the U.S. burglary clearance rate has been between 13 and 14% for every year between 1995 and 2018. Other crime types have similar consistently low clearance rates over time. These low clearance rates continue despite recent advances in evidence collection and analysis technologies, such as DNA (Eck & Rossmo, 2019). Improving police clearance rates for the purpose of improving justice and crime control remains an important yet undeveloped area needing research (Braga & MacDonald, 2019).

Research from fields ranging from medicine to aviation suggests that a simple checklist can improve decision-making during complex and stressful tasks by helping guard against errors and oversights. Checklists help by disaggregating complex situations into discrete, simple, yet necessary steps (Gawande, 2010). Ultimately, a checklist can improve decision-making and produce better outcomes for even the most complex tasks (de Vries et al., 2010; Pronovost et al., 2006; Weiser et al., 2010).

We juxtapose the idea of using checklists to improve outcomes of complex tasks against the historically low clearance rates of burglary investigations in the USA. Doing so presented the following hypothesis: police officers who have a checklist will be more likely to collect evidence from a crime scene than police officers who do not have a checklist. To test this hypothesis, we conducted a randomized controlled trial (RCT). The RCT was conducted using virtual reality (VR) because it allowed us to study a potentially effective yet unproven practices with minimal intrusion and burden on any police agency (i.e., expense) while standardizing the burglary scenario across participants which would be impossible in practice. After reviewing the literature on police investigations, checklists, and virtual reality as a research method, we detail the RCT, present the study results, and discuss the study's policy implications and limitations.

Literature review

Police clearance rates

Although police investigation studies started at least 40 years ago, research remains scarce (Eck & Rossmo, 2019). Nonetheless, the process in which crimes are cleared remains relatively simple. Generally, crimes are cleared by the initial responding officer at the scene because the crime is in progress or the offender is still at the scene (Greenwood, 1970; Greenwood & Petersilia, 1975). Crimes are also frequently cleared simply because the victims and/or witnesses at the scene can provide the name, address, or other identifying information about the offender to the responding officer (Eck, 1983; Greenwood et al., 1975). In those cases, clearances require relatively minimal "investigative work" beyond tracking down the accused suspect (Greenwood et al., 1975). When no initial offender identification is made, crimes are then "investigated" and most likely to be cleared based on evidence collected by the initial responding officer (Eck, 1983; Greenwood & Petersilia, 1975).

When crimes are investigated, research suggests evidence that might lead to higher case clearance includes: (1) identifying item observed/found at scene, such as vehicle or license plate description; (2) tracking a stolen vehicle; (3) fingerprints from the scene used to identify or confirm a suspect; (4) footprints from the scene used to confirm a suspect; (5) DNA from the scene used to identify or confirm a suspect; (6) photo lineups shown to witnesses used to identify or confirm a suspect; (6) tips from informants or law enforcement colleagues on potential suspects; (8) neighborhood canvasses to find witnesses; (9) tool marks from a scene used to confirm a suspect; (10) recovered stolen property that can be traced back to suspects; (11) information from interrogations that can be used to identify or confirm suspects as well as connect similar crimes to a suspect; or (12) computer records checks of past crimes or modus operandi to identify potential suspects (see Chaiken et al., 1977; Donnellan & Ariel, 2019; Eck, 1979, 1983; Greenwood, 1970; Greenwood & Petersilia, 1975).

It is important to point out that the burden of the case clearance process initially falls on the responding officer(s), and how this process is initiated will often affect the entire investigation. In most police agencies, the initial investigation is key in determining if a case will receive a follow-up investigation by a detective with roughly half of all cases screened out of the investigatory process after the initial investigation (Eck, 1979, 1983). In sum, the initial responding officer either makes the arrest or records the information that leads to case clearances during the detectives' follow-up (Chaiken et al., 1977; Donnellan & Ariel, 2019; Eck, 1979, 1983;Greenwood, 1970 ; Greenwood & Petersilia, 1975).

Officers exercise considerable discretion in all facets of police work, including investigations (President's Commission on Law Enforcement and Justice, 1967; Walker, 1993). Initial investigation quality will, therefore, vary by officer, organizational, and perhaps other factors (Greer, 2014; Horvath et al., 2001). Officers start with different characteristics, training, and experience. Moreover, initial investigations are difficult in practice. Officers experience everyday factors that affect job performance (e.g., lack of sleep or personal stressors). Officers are also under pressure to resolve calls quickly due to high workloads (Greenwood, 1970; Greenwood & Petersilia, 1975). It is common for officers to have a long queue of serious calls that require their attention (see: Moskos, 2008). Likewise, organizational policy, goals, and workload may impact officer discretion. All of these factors can ultimately impact the quality of initial investigations (Eck, 1983; Peterson, 1974; President's Commission on Law Enforcement and Justice, 1967).

Given the police have relatively little control over whether the offender is still present, or victims observe the offender during a crime (Greenwood et al., 1975), methods to improve evidence collection and processing by frontline officers are the most fruitful path for improving police investigative outcomes (also see: Eck, 1983; Greenwood & Petersilia, 1975). These points lead to an important question: Is it possible to systematically improve the quality of initial investigations?

Checklists

In fields with highly skilled practitioners, such as medicine or aviation, there have been countless instances where important outcomes, such as keeping people alive, were not achieved due to the ***difficulty of applying core knowledge in routine yet complex situations*** *(*Gawande*,* 2010*)*. Sometimes experts simply forget about key steps or information under stress or the presence of competing demands for attention. Other times experts may miss steps in a process as tasks become routine (Gawande, 2010). Fortunately, however, these fields have derived a solution: checklists. Checklists guard against simple decision-making errors by disaggregating complex situations into discrete, simple, yet necessary events requiring a yes/no or completed/not completed response (Gawande, 2010). While simple tools, checklists improve decision-making and produce better outcomes (e.g., catheter infections, surgical complications, and mortality risks) for complex medical procedures (de Vries et al., 2010; Pronovost et al., 2006; Weiser et al., 2010).

Perhaps, checklists can be translated to policing (Ratcliffe, 2018; Sidebottom et al., 2012). Greer (2014) was the first researcher to propose and test checklists for initial investigations.[1] In that study, a checklist for burglary and theft from motor vehicle crimes was created after a robust literature review on solvability factors and input from line officers in the Morristown (NJ) Police Department. Officers were then trained on and used the investigative checklists. After implementation, case files were coded to identify which pieces of evidence were collected for cases during the 4-month implementation period and the 4-month pre-implementation period. Greer (2014) found the collection of evidentiary items per 1000 cases increased for six categories (previous victimization, fingerprints, offenders observed, suspect descriptions, suspect names, stolen property recovered), decreased for six categories (DNA collection, footprint/shoe impressions, articles left by offenders, vehicles stolen, witnesses located, canvasses), and remained unchanged for three categories (in-progress crime, offender vehicles observed, stores checked for stolen items) when the pre- and post-period cases were compared. Evaluating real-world evidence collection, however, comes with limitations. Primarily, the presence of a particular evidentiary item is outside the control of police, and there may be non-random factors that influence the presence of evidentiary items between cases and across time. As Eck and Rossmo (2019: 606) observed for murders, "Not all murders can be solved; police have no control over the killer's mistakes, location of the crime, presence of forensic evidence, existence of witnesses, or the nature of the victim–offender relationship (Wellford & Cronin, 2000)."

[1] While discussing preliminary results from this study and during the review process, we heard anecdotal reports that some police agencies previously adopted and informally tested investigative checklists with varying levels of success. However, our literature search did not surface any official reports/studies that can be cited to inform the current study.

Data and method

As such, the present study will assess the impact of investigative checklists used by initially responding officers on evidence collection success using a randomized control trial (RCT) in virtual reality. Virtual reality (VR) provides the unique opportunity to create a realistic burglary response experience that can be standardized across all investigating officers. In the physical world, it would be both expensive to implement and evaluate an investigative checklist program as well as impossible to ensure that all officers are investigating similar types of cases.

Study Site

The study took place at a large, mid-western municipal police department. The agency has roughly 1000 sworn officers to police a city with more than 300,000 residents. According to the National Incident-Based Reporting System (NIBRS), in 2021, the agency had a burglary clearance rate of roughly 13% which was just slightly higher than the national burglary clearance rate of 11.3%.

Virtual reality burglary simulation

A "standard" burglary call for service (CFS) VR experience was developed by a team of student developers and one of the authors (Professor Ming Tang). The experience was staged during civil twilight as burglaries commonly occur during the day and are reported when people come home from work (Rengert & Wasilchick, 1989). The scenarios started with participants exiting a patrol vehicle in front of a house where a burglary victim was standing in the garage of their home.

The victim then provided the following automated script to the officer:

> *Hi officer. My name is Jane Doe. This is my house. I called as soon as I noticed things were off. When I came home, I found my house was broken into. I went to open my door, and noticed it was unlocked. I thought it was weird, but went inside. I noticed a lot of our stuff was tossed around and a few things broken. I looked around and noticed our master bedroom was a mess. I think they broke in through the bedroom window. It is broken and my jewelry is missing. I didn't see anything out of place upstairs. When I realized someone broke in, I called 911 immediately…. So, I haven't talked to any neighbors. I didn't see anyone lurking around or anything suspicious since I've been home. Please go in and investigate if you want. I'll be out here if you need anything.*

The script was written so that the victim had minimal information to provide (due to not being home during the incident), and the officer would need to investigate the scene on his or her own to collect information.

After the victim provided their automated description of the incident, the participating officer was able to start their preliminary investigation of the burglary based on the instructions provided in the study's protocols/scripts (described in detail below).

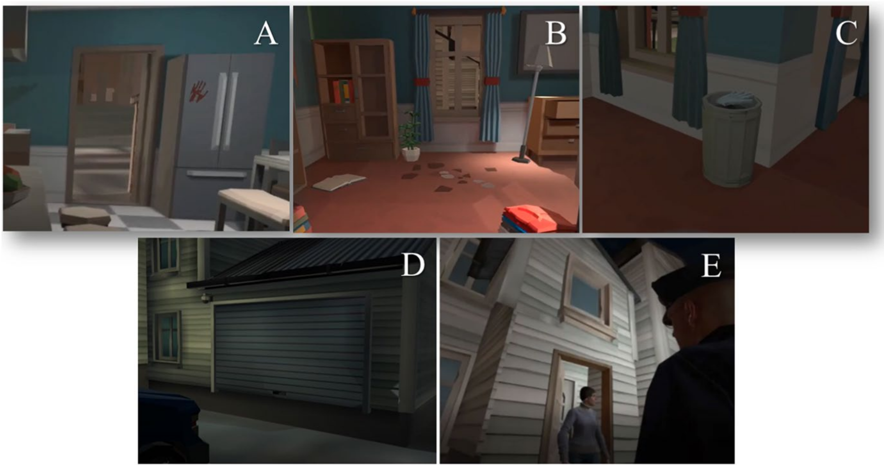

Fig. 1 Evidentiary items available for collection in the virtual reality simulation. **A** DNA/blood stain found on the refrigerator door. **B** Footprint mold found in the first-floor bedroom. **C** Item left at the scene (i.e., white glove) by the perpetrator. **D** Private security camera located on the neighbors house. **E** Neighbor who witnessed portions of the burglary.

In general, officers were able to move through the inside of the victim's burglarized home, on all sides of the burglarized home, and along the burglarized home's street block. The victim's home consisted of six rooms: (1) a garage, (2) a kitchen/dining area, (3) a living room, (4) a foyer, and (5) a bedroom with an (6) attached bathroom. The simulated house suggested it had a second floor, but it was programmatically inaccessible during the simulation. Participants were informed of this limitation during the simulation if they attempted to go upstairs.

As participating officers completed their initial investigation, they were able to collect five different pieces of evidence by moving to the vicinity of the evidence and reaching out to interact with the evidentiary item. When an evidence item was collected, a prompt would show on the screen indicating as much. Asking officers to "touch" evidence during the experience was a break from standard police practice and is a limitation of the experience's presence/realism, but participating officers were alerted of this issue by study's protocols/scripts (see below). It is emphasized that participants in both groups were not told how many or which evidentiary items were available for collection.

Figure 1 shows a visual representation of the five evidentiary items. We do note, however, that Fig. 1 (or any still images of a VR simulation) does not fully depict the realism commonly experienced by VR users.[2] Additional details on each evidence item are provided below:

[2] The evidentiary item images are provided for reference, but do not adequately capture the realism and presence users typically experienced in VR. Honestly, one of the worst ways to sell the realism of VR is to show 2D images and/or video. We remain confident that all users could equally collect/experience all five evidentiary items regardless of assignment group.

1. ***DNA/blood stain:*** A bloody handprint was located on the refrigerator door inside the home (see Fig. 1A). Note that in a side bedroom, there was a broken window indicating that the offender entered the home in that manner and hence the blood stain.
2. ***Footprint mold:*** After entering the home, there was a first-floor bedroom just past the foyer and standard living room. The broken bedroom window was located inside this bedroom. Upon investigating the broken window, officers could find clear, muddy footprints on the carpet and choose to collect the evidence as a footprint mold (see Fig. 1B).[3]
3. ***Item left at the scene:*** A trashcan was staged inside the front door within the entry foyer, and a black glove sat on top of the trashcan in a relatively obvious manner. Officers could collect the glove as a piece of evidence (see Fig. 1C).
4. ***Private security camera video from adjacent neighboring house:*** By investigating the outside of the burglarized home, near the broken window used to enter the home, security cameras on the immediately adjacent neighbors' home would have been visible to participants. In fact, the camera that was placed onto the home was larger and less technologically advanced than many modern home surveillance camera systems, which presumably made the cameras more obvious to participants. Participants could have then approached the neighboring home to knock on the door, at which point a cut scene with pre-determined dialogue would start, and the neighbor would offer to make a copy of the camera's recording for the officer (see Fig. 1D).
5. ***Eyewitness statement about offender description from across the street neighbor:*** Across the street and a few doors down the block from the burglarized home, another residence had its porch lights on with a car in the driveway giving the appearance that the resident was home. The other homes on the block were dark and without cars present. If the participant approached the home's front door to knock, a cut scene would initiate once the participant was within a step of the home's porch (i.e., a proximity trigger). The resident, via an automated dialog, then provided a description of the burglar, a getaway car, and the vehicle's license plate number (see Fig. 1E).

These five evidentiary items were based on the literature review of case clearances and solvability factors conducted for the study. Evidentiary items previously correlated with case clearances (see above) that the development team believed they could relatively easily implement within the virtual environment were selected.[4] Participants could engage with several other items in the experience (e.g., a neighbor

[3] The initial plan was to have the footprint in the mud outside the broken window, but it proved difficult to achieve for the developers. Placing the footprint inside was more feasible.

[4] In response to a reviewer, we note the choice to include five evidentiary items was arbitrary but provided a logically round number that balanced having enough items to make the investigation interesting but not so many that the experience would be burdensome for participants. Future studies should vary the number of evidentiary items to determine if it has any impact on the results.

Fig. 2 Checklist vs. no checklist virtual reality experience examples

without any information to contribute and standard household items), but they possessed no evidentiary value and thus did not represent "collected evidence."

Randomized control trial design

Within this general VR simulation framework, a ***randomized controlled trial*** *(RCT)* was conducted to evaluate the effectiveness of investigative checklists for collecting evidence during initial investigations. Participants in the randomly assigned ***treatment group*** were provided with an investigative checklist to use during the simulation and, as described in more detail below, were provided instructions to use the checklist during the pre-simulation instructions. Specifically, participants in the treatment group had an investigative checklist "attached" to their wrist, much like a watch, during their simulation. The checklist included a bulleted list of the following items: (1) DNA, (2) item left at the scene, (3) footprint mold, (4) private security camera video, and (5) neighborhood canvas (see Fig. 2, treatment panel). The checklist could be viewed at any time throughout the simulation. Participants in the control group simply did not have access to the checklist at any time (or know that other participants did; see Fig. 2, control panel). From a technical standpoint, separate experiences were developed where the only difference between the experiences was that one included the checklist for use and the other did not.

Data were collected on-site at two of the participating police department's district buildings. Data collection occurred over seven multi-hour sessions. Data collection sessions occurred near roll call times, and an introduction to the study was provided by a police supervisor and the research team. The protocol instructed officers: "The purpose of this research study is to utilize fully immersive 3D VR simulation—in coordination with other sensing technologies—to develop, study, and enhance the next generation of evidence collection training for Ohio law enforcement." Officers could volunteer to participate by simply showing up at a table outside a room in the police district where data collection took place. A police supervisor helped manage the flow of participants by recruiting officers who were not engaged in official business and/or instructing officers to return at a later time when a backlog of potential participants occurred.

Once an officer approached the study enrollment table, they were informed about their human subject research rights, asked questions about their potential participation in the study, and ultimately volunteered to participate. All volunteers completed the 16-item Simulator Sickness Questionnaire (Kennedy et al., 1993) to identify any possibilities of having an adverse experience with VR. Any potential participant indicating they had moderate or severe symptoms to one of 16 ailments would have been screened out of the study, but that did not occur. Next, officers were assigned an anonymous unique identifier and completed a pre-experience survey that collected demographic information and prior experience with virtual reality. After completing the pre-survey, participants' treatment status was determined from a pre-randomized list and the VR component of the study started.

Participants stood in the center of the room in a pre-marked square and were given an overview of the VR headset and controllers prior to putting on the VR headset. Next, a research team member started reviewing an introductory script. The first part of the script provided training on how to move, investigate objects, and interact with the experience. After the VR orientation was completed, participants were instructed: "Please use your new skills to interact with the environment and investigate the burglary in accordance with your training and experience." Participants assigned to the treatment group then received the additional instructions:

> We have provided you with an evidence checklist to assist you during your investigation. Once you enter the simulation, the checklist will be activated. To look at the checklist, turn your left wrist as if you are checking the time on a watch.

Participants in the control group did not receive this additional instruction or have access to checklist feature at all.

The logic of the treatment is that having the checklist will guide treatment officers' discretion during the initial investigation (Walker, 1993). The investigative checklist will encourage officers to conduct their initial investigation systematically and thoroughly and ultimately collect more evidence (Gawande, 2010). It is important to note, however, that the treatment did not represent an extensive training program, but rather more of a nudge towards desired behavior (National Academies of Sciences, Engineering, 2023; Thaler & Sunstein, 2009).

Participant enrollment

The study sought to enroll roughly 50 participants. A total of 28 treatment participants and 26 control participants were ultimately enrolled in the study. However, three participants were ultimately dropped from the study because the officer(s) did not (1) complete the simulation and/or follow-up survey as designed.[5] This left 26 treatment participants and 25 control participants, for an analytical sample of 51.

[5] One officer was called to a crime in progress before completing the simulation and post-survey. Another officer spent most of their time comparing another VR experience they were familiar with rather than completing the task at hand. Finally, one officer failed to complete the post-survey.

Table 1 Sample characteristics by treatment and control group

Variables	Full sample (N= 51)		Treatment (N = 26)		Control (N = 25)		Test statistic[1]	p value
	Mean	SD	Mean	SD	Mean	SD		
Age[2]	36.16	9.96	36.92	9.87	35.36	10.21	-0.556	0.581
White	0.82	0.39	0.85	0.37	0.80	0.41	1.366	0.726
Male	0.82	0.39	0.81	0.40	0.84	0.37	0.804	1.000
Bachelors or higher	0.65	0.48	0.65	0.49	0.64	0.49	1.061	1.000
LE experience[2]	10.14	8.76	11.03	9.22	9.22	8.34	-0.734	0.466
Weekly video game[3]	3.22	5.18	2.73	4.71	3.75	5.70	0.686	0.496
Prior VR experience	0.57	0.50	0.65	0.49	0.48	0.51	2.017	0.264
Own VR system	0.20	0.40	0.15	0.37	0.24	0.44	0.582	0.499

[1]Independent sample t tests were conducted for continuous measures and Fisher's exact tests were conducted for nominal measures

[2]Measure was not captured at the precision of two decimal places, but two decimal places were used to maintain consistency across the table

[3]One participant did not disclose the amount of time spent playing video games each week and was not included in the t test analysis for this particular item

To assess if the randomization process created balanced treatment and control groups as intended, eight items from the pre-survey were analyzed. Four measures captured basic demographics. First, officers provided their birth year, and their ages were computed as a continuous measure of years by differencing the year of data collection and their birth year.[6] Second, officers provided their race/ethnicity, which was binarized into officers identifying as white (coded "1") versus non-white (coded "0") due to limited variation in the measure across the other response categories. Third, officers identified themselves as either male (coded "1") or female (coded "0"). Fourth, officers' education levels were captured through a binary variable that measured whether an individual completed a post-secondary degree (2-year college degree or greater coded as "1") versus those who did not complete college (coded as "0"). Four more measures captured officers' police experience and VR/video games familiarity. Officers provided the specific number of years they have worked in law enforcement. Next, participants provided the average number of hours per week spent playing video games with a controller or keyboard over the last year. Prior VR usage was captured with a binary variable indicating Never (coded "0") versus using VR Once or more (coded "1"). Finally, VR ownership captured whether or not the participant or someone in their house owned a VR system (Yes = "1" versus No = "0").

Table 1 shows randomization resulted in completely balanced treatment and control groups. Officers in the treatment group were an average age of 36.92 years old compared to 35.36 years old in the control group. Both the treatment and control

[6] Data collection took place in the late 2022 and early 2023.

group officers were mostly male (81% vs. 84%). Participants in both groups also predominantly self-identified as white (treatment group = 85% and control group = 80%). Education levels were also similar in the treatment and control group with about 65% and 64% of each group respectively reporting having a high school/GED degree without any additional university-level education. Generally, the sample was experienced with the mean number of years on the job equaling 11.03 for the treatment group and 9.22 for the control group (different of 1.81 years). In terms of video game and VR usage, the treatment and control groups both had at least some experience with videogames and VR: (1) treatment group participants (MEAN = 2.73 hours) averaged about an hour less of video game/VR playing per week than control group participants (MEAN = 3.75 hours); (2) about 65% of the treatment group had prior VR experience compared to 48% of the control group; yet (3) 15% of the treatment group reported they or someone in their house owned a VR system relative to 24% of the control group. Statistically significant differences between the treatment and control group on the above items were assessed using Fisher's exact test for binary variables and *t*-tests for the continuous measures. Overall, none of the minor differences between the treatment and control group achieved statistical significance on any item described above (see Table 1).

Outcomes and analytic plan

Five outcomes were examined. The five outcomes corresponded to the five evidence items that were available to be collected during the initial investigation: (1) item left at the scene; (2) DNA/blood stain; (3) footprint mold; (4) private security camera video from adjacent neighboring house; (5) eyewitness statement about offender description from across the street neighbor. The experience automatically tracked when a participant collected the items. Participants were instructed to "return to their police car" at the end of their investigation. At that point, the experience was over and displayed a "scoreboard" indicating which pieces of evidence were collected by the participant.[7] Once the participant removed the VR headset, the evidence items collected by the participants were recorded in a database by a member of the research team. A binary measure of collected (coded "1") versus not collected (coded "0") was used for each of the five evidence items.

Contingency tables comparing evidence collection versus non-collection between the treatment and control groups were produced for each item. Statistical significance in the differences in evidence collection between the treatment and control groups was computed using Fisher's exact test given some contingency table cells contained fewer than five cases. The results were computed using the *stats* package in R version 4.2.1

[7] The participants' view within the VR experience was simultaneously cast to a computer monitor in the data collection room and recorded as a backup plan as well, but it was not necessary to view the recordings to collect the outcome data.

Table 2 Evidence collection by item for treatment vs. control group participants

Evidence found?	Treatment ($N = 26$)	Control ($N = 25$)	OR	p value
Item left at scene				
Yes ($N = 42$)	22 (85%)	20 (80%)	1.366	0.726485
No ($N = 9$)	4 (15%)	5 (20%)		
DNA evidence				
Yes ($N = 46$)	23 (88%)	23 (92%)	0.672	1.000000
No ($N = 5$)	3 (12%)	2 (8%)		
Footprint mold				
Yes ($N = 47$)	24 (92%)	23 (92%)	1.043	1.000000
No ($N = 4$)	2 (8%)	2 (8%)		
Security camera				
Yes ($N = 27$)	22 (85%)	5 (20%)	20.147	0.000004
No ($N = 24$)	4 (15%)	20 (80%)		
Eyewitness statement				
Yes ($N = 25$)	19 (73%)	6 (24%)	8.170	0.000702
No ($N = 26$)	7 (27%)	19 (76%)		

Percentages are calculated by column for each item

Results

Table 2 shows the contingency tables for each item and the results of the Fisher's exact tests. Recall the first three pieces of evidence were available to be collected from within the victim's home. The "item left at the scene," a glove sitting atop a trash can inside the front door, was collected by about 85% of the 26 treatment group participants and about 80% of the 25 control group participants. Next, roughly 88% of the 26 treatment group participants and 92% of the 25 control group participants collected the DNA evidence from the crime scene. Finally, roughly 92% of both the 26 treatment group participants and 25 control group participants respectively collected the footprint mold. Unsurprisingly, the Fisher's exact tests confirmed that these differences were not statistically significant. Overall, the conclusion that can be drawn is that most participants, regardless of whether they used the investigative checklist, found the three pieces of evidence inside the virtual victim's home.

The last two evidence items analyzed in Table 2 were collected from outside the virtual victim's home. For the security camera footage collected from the next-door neighbor's home, about 85% of the 26 treatment group participants collected the evidence item whereas about 20% of the 25 control group participants collected the evidence item. According to the Fisher's exact test, this difference is statistically significant ($p = 0.000004$). Likewise, about 73% of the 26 treatment group participants collected the eyewitness statement from the across-the-street neighbor, but only 24% of the 25 control group participants collected the evidence item. Again, a Fisher's exact test showed the difference between the treatment and control groups' collection of the eyewitness statement was statistically significant ($p = 0.000702$).

Discussion

This study sought to test if providing patrol officers with an investigative checklist during an initial burglary investigation within a VR experience would improve evidence collection. The results showed that officers who were randomly provided with an investigative checklist were more likely to collect security camera footage from an adjacent neighbor and an eyewitness account from a neighbor across the street, but both the treatment and control officers were just as (highly) likely to collect an item left at the scene (i.e., glove), DNA, and a footprint impression.

The value of checklists is that they breakdown complex tasks into simpler steps and prompt users to complete all steps of a routine process, and their use has improved outcomes across a variety of fields (Gawande, 2010). As first hypothesized by Greer (2014), the present results suggest investigative checklists for initial investigations are a promising innovation. It is noted that the core difference between the treatment and control officers in the study was the collection of evidence items that were "more difficult" to collect — evidentiary items located outside the home. In other words, both experimental groups collected the items inside the home that were relatively obvious after entering the home and completing a simple walk around the burglary scene. As such, the present findings might be interpreted as particularly promising given the checklists seemed to put a "check" on evidentiary items that might be most likely to be missed when officers are busy, under stress, or just "going through the motions."

It is also noteworthy that the two evidentiary items (i.e., eyewitness account and CCTV camera footage) where the difference between treatment and control officers was found have been shown to be key solvability factors in the literature (Donnellan & Ariel, 2019; Eck, 1983). If these important solvability factors are commonly "missed" in practice despite being available, then checklists are especially promising and will likely ultimately improve clearance rates — but that is an empirical question not tested in this study. Given Eck and Rossmo's (2019: 609) assertion that there has been little innovation in the investigative process, these results are a promising divergence and show that it is possible to make positive changes to the investigative process.

It is important to note, however, that this study may have limited external validity, and these results may not hold in real-world settings. For example, checklists may hold value in situations that are stressful, where there are competing demands, behavior becomes routinized, and so on. While the officers in our study were still under pressure to go out on the street to start or continue their patrol shift, they were not necessarily under the same stress and pressures officers might face in live-action scenarios or routine patrol work, especially since they were given permission by their superiors to participate in the study. Alternatively, the effects of checklists could decay over time as their use becomes routine for officers as well. Of course, external validity is always an empirical question that must be investigated with further research, and it is certainly possible that the limitation of checklists could be addressed with clever programmatic changes, such as reviewing the body-worn cameras of officers using checklists to ensure they continue to use them over time (also see Greer, 2014). In the future, using qualitative methods to understand how officers perceive and use checklists in practice would be particularly instructive. For example, it would be insightful to know why some officers in the treatment group did not collect all evidentiary items despite having the checklist.[8]

Also, it is an empirical question as to whether or not increased evidence collection ultimately leads to changes in case clearances. Other intervening variables, such as detective workload, laboratory evidence processing, and so on, may mean that investigations still face challenges even when initially responding officers use checklists and collect evidence. Direct tests of the impact of investigative checklists on case clearances are paramount for future research.

It is also worth noting that using investigative checklists could unintentionally impact the length of time it takes patrol officers to clear a call for service. Presumably, the total length of time officers spend on calls for service would increase if officers were spending more time searching for evidence. But it is also possible that checklists could streamline initial investigations, thus leading to reduced call-for-service lengths. If investigative checklists and subsequent evidence collection do not improve investigative outcomes, then this additional time may not be a good investment of scarce police personnel. Again, this is an empirical question that cannot be answered with the present study but will be important for future research.

Finally, we offer a quick note on VR policing experiments that may be informative for future studies. Our study shows that it is possible to use VR as a research method to test novel ideas in policing; whereas, we suspect many studies will use VR as a tool for treatment (e.g., see Martaindale et al., 2023). However, we want to stress that VR experiments do not remove many of the logistical and political challenges that face policing field studies. For example, the participating policing agency, like all agencies, has been facing recruiting and retention challenges; thus, there were administrative concerns about taking officers off the street to participate. This led to a compromise of limiting data collection to a pilot study size sample of roughly 50 participants. We anticipate researchers will face similar and other challenges when executing VR studies in policing agencies in the future.

Overall, the present study supports the potential for using VR as a data collection methodology. In this study, we enrolled participants, got them acclimated to navigating the VR experience, and ultimately through the simulation in a reasonable period of time. No officers reported any adverse health effects (i.e., nauseous, dizzy) to the authors' knowledge, which is commonly feared (and still possible in future studies) when using VR. Overall, this is promising for the potential of VR in the future. Not only might VR be used to deliver training interventions in policing, but researchers can also use it to put officers through experiences that otherwise might be too dangerous, impossible, counterproductive, or expensive in real-world settings (Bailenson, 2018). In effect, VR can potentially unlock a range of new research studies.

[8] The authors thank one of the anonymous reviewers for raising this important point.

References

Bailenson, J. (2018). *Experience on demand: what virtual reality is, how it works, and what it can do*. Norton.

Braga, A. A., & MacDonald, J. (2019). Improving police effectiveness in ensuring justice. *Criminology & Public Policy, 18*(3), 511–523.

Chaiken, J. M., Greenwood, P. W., & Petersilia, J. (1977). The criminal investigation process: a summary report. *Policy Analysis, 3*(2), 187–217.

de Vries, E. N., Prins, H. A., Crolla, R. M. P. H., den Outer, A. J., van Andel, G., van Helden, S. H., Schlack, W. S., van Putten, M. A., Gouma, D. J., & Dijkgraaf, M. G. W. (2010). Effect of a comprehensive surgical safety system on patient outcomes. *New England Journal of Medicine, 363*(20), 1928–1937.

Donnellan, P. G., & Ariel, B. (2019). Assessing solvability factors in greater manchester, england: the case of residential burglaries. In T. Coupe, B. Ariel, & K. Mueller-Johnson (Eds.), *Crime Solvability Factors* (pp. 35–51). Springer International Publishing.

Eck, J. E. (1979). *Managing case assignments: the burglary investigation decision model replication*. Police Executive Research Forum.

Eck, J. E. (1983). *Solving crimes: the investigation of burglary and robbery*. Police Executive Research Forum.

Eck, J. E., & Rossmo, D. K. (2019). The new detective: rethinking criminal investigations. *Criminology & Public Policy, 18*(3), 601–622.

Gawande, A. (2010). *The checklist manifesto*. Penguin Books.

Greenwood, P. W. (1970). *An analysis of the spprehension activities of the New York City Police Department*. Rand Corporation.

Greenwood, P. W., Chaiken, J. M., Petersilia, J., & Prusoff, L. (1975). *The criminal investigation process volume III: observation and analysis*. Rand Corporation.

Greenwood, P. W., & Petersilia, J. (1975). *The criminal investigation process volume III: summary and policy implications*. Rand Corporation.

Greer, S. (2014). *Improving police investigations in burglary and theft from motor vehicle cases using checklists: a phase-one implementation trial in a medium-sized police agency*. Cambridge University.

Horvath, F., Meesig, R. T., & Lee, H. Y. (2001). *National survey of police policies and practices regarding the criminal investigations process: twenty-five years after Rand*. National Institute of Justice.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203–220.

Martaindale, M. H., Sandel, W. L., Duron, A., & McAllister, M. J. (2023). Can a virtual reality training scenario elicit similar stress response as a realistic scenario-based training scenario? *Police Quarterly.*

Moskos, P. (2008). *Cop in the hood: my year policing baltimore's eastern district*. Princeton University Press.

National Academies of Sciences, Engineering, and M. (2023). *Behavioral economics: policy impact and future directions*. National Academies Press.

Peterson, J. L. (1974). *The utilization of criminalistics services by the police: an analysis of the physical evidence recovery process*. National Institute of Justice.

President's Commission on Law Enforcement and Justice. (1967). *The challenges of crime in a free society* (Vol. 4). United States Government Printing Office.

Pronovost, P., Needham, D., Berenholtz, S., Sinopoli, D., Chu, H., Cosgrove, S., Sexton, B., Hyzy, R., Welsh, R., & Roth, G. (2006). An intervention to decrease catheter-related bloodstream infections in the ICU. *New England Journal of Medicine, 355*(26), 2725–2732.

Ratcliffe, J. H. (2018). *Reducing crime: a companion for police leaders*. Routledge.

Rengert, G. F., & Wasilchick, J. (1989). *Space, time, and crime: ethnographic insights into residential burglary*. National Institute of Justice.

Sidebottom, A., Tilley, N., & Eck, J. E. (2012). Towards checklists to reduce common sources of problem-solving failure. *Policing, 6*(2), 194–209.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: improving decisions about health, wealth, and happiness*. Penguin Books.

Walker, S. (1993). *Taming the system: the control of discretion in criminal justice 1950-1990*. Oxford University Press.

Weiser, T. G., Haynes, A. B., Dziekan, G., Berry, W. R., Lipsitz, S. R., & Gawande, A. A. (2010). Effect of a 19-item surgical safety checklist during urgent operations in a global patient population. *Annals of Surgery, 251*(5), 976–980.

Wellford, C., & Cronin, J. (2000). Clearing up homicide clearance rates. *National Institute of Justice Journal, 243*, 1–7.

Cory P. Haberman Ph.D. is an Associate Professor in the University of Cincinnati's School of Criminal Justice where he also directs the Institute of Crime Science. Dr. Haberman's work focuses on crime and place, crime science, police effectiveness, and the application of virtual reality to all of those topics.

Ming Tang Registered Architect, is a Professor at the School of Architecture and Interior Design, College of Design, Architecture, Art, and Planning, University of Cincinnati. He is the Director of UC XR-Lab and leads research on Extended Reality, AI, Metaverse, and BIM-based Digital Twin. He is the founding partner of TYA Design. His multi-disciplinary research includes Virtual reality & Augmented reality, BIM-Based Digital twins, Gamebased learning, Computational design, Digital fabrication, Generative and performance-driven design, Human behavior analysis, and simulation. He also serves on the board of the UC Institute for Research in Sensing (IRiS) committee of the Industry 4.0/5.0 Institute.

Clay Driscoll is currently a Ph.D. candidate at the University of Cincinnati. His work includes a variety of topics across policing, crime prevention, and public safety. His current primary research focuses are on the use of virtual reality as a training device and as a tool for research. He also examines various topics, such as driver behaviors and substance use treatment. More information about him can be found at claydriscoll.com.

Bradley J. O'Guinn is a Ph.D. Candidate in the School of Criminal Justice at the University of Cincinnati. His research focuses on police effectiveness, problem-solving, and police organizations. His recent work has appeared in Crime and Delinquency and Police Practice and Research: An International Journal.

Calvin Proffit is a Ph.D. student in the School of Criminal Justice at the University of Cincinnati. His research focuses on issues related to race/ethnicity and crime, immigration, crime prevention, and policing.

J.C. Barnes Ph.D. is Professor and Director in the School of Criminal Justice at the University of Cincinnati where he teaches research design and advanced statistics.