Participatory Serious Game Design for Societal Intervention

Exploring the Effectiveness of Participative Game Design Processes Between Communities and Police in the United States as a means for Reduced Police Bias

Master Thesis

a partial requirement for a
Master of Science in Engineering and Policy Analysis
to be defended publicly on Wednesday 29th of September 2021

by

Caroline Cullinan

Faculty of Technology, Policy and Management,
Delft University of Technology, Delft

Thesis Committee:

Prof Dr. Ir. A Verbraeck	Delft University of Technology, Chair
Dr. L. J. Kortmann	Delft University of Technology, First Supervisor
Dr. C. N. van der Wal	Delft University of Technology, Second Supervisor
Dr. R. Kenter	Center for Policing Equity, External Supervisor

Keywords: societal intervention, serious game design, participatory systems, community-oriented policing, police bias

ACKNOWLEDGEMENTS

This project was made possible with the help and support of countless individuals. I would first like to thank my Thesis Committee. To my First Supervisor, Dr. Rens Kortmann, thank you for your continued guidance and critical feedback throughout this project. Your personal interest in the work and your passion for serious game design was incredibly motivating. To my Second Supervisor, Dr. Natalie van der Wal, thank you for providing me with very valuable insight from the perspective of social psychology and group dynamics. Your feedback was critical in shaping the intervention design in which served as the foundation of this research. To my Chair Person, Dr. Alexander Verbraeck, thank you for your constructive feedback and your incredibly swift email replies. In the very rare instances that I felt both lost in my work and pressured for time, it was incredibly comforting to know that you were just a quick email away. To my External Supervisor, Dr. Rob Kenter, thank you for your incredible insight and knowledge into the world of US policing. Your support, guidance, and feedback were fundamental in shaping this work as something that is not only scientifically relevant, but also, hopefully, socially relevant as well.

I would also like to thank the *Center for Policing Equity (CPE*) for their support and collaboration on this work. Specifically, I would like to thank Dominique Johnson for her support early on in situating this work internally within the organization and engaging community interest. I would also like to thank Price Nyland, who not only helped with community insight, logistics, and scheduling, but who also dedicated countless hours to the actual running and facilitation of workshops. Price, you were fundamental in this work. Finally, I would also like to thank Dr. Tracie Keesee. In November of 2020 I made a very bold attempt to contact *CPE* through a LinkedIn direct message sent to the Co-Founder and Senior Vice President of Justice Initiative at *CPE*. To my utter shock, Dr. Tracie Keesee not only replied to my message, but she was also interested in pursuing a collaboration. Thank you Dr. Tracie Keesee for not only making time for me, but for also taking a chance on me. Without your time and patience, this project would have never been brough to light.

Caroline Cullinan
Rome, IT, August 2021

PREFACE

In considering my own personal reflections in completing this project, I believe that this study provided an excellent demonstration of the value of mixed methods approaches to research. Although some results did emerge through the quantitative analysis put forth in this project, small sample sizes made it difficult to rely on quantitative output alone with full certainty. As such, the qualitative component to this work, instead, succeeded to provide nuanced context and greater detail to this analysis.

This project also provided me with an excellent introduction to working at the intersection of community engagement and research. In particular, the opportunity to exercise my community organizing skills in collaboration with the *Center for Policing Equity (CPE)* provided invaluable experience in working with multiple actors across multiple time zones, cultures, and political orientations - all within a scientific context. Through this experience, I not only developed my skills as a researcher, but I also gained skills in project management and stress management, as participant circumstances constantly impacted the direction and design of this project.

While the "working with human subjects" aspect to this project brought on a series of challenges and complications, I believe that it greatly impacted my own commitment and passion for this project. As stated by one research participant, "*real* people want *real* solutions to *real* problems", and as such, I cannot imagine completing a community-oriented study through less *real* and non-participatory means.

As such, I highly recommend future MSc thesis students that are interested in community-oriented studies to consider taking participatory approaches to their work. In taking such approach, not only can future students have a more meaningful research experience, but their research can also be more societally impactful.

Caroline Cullinan
Rome, IT, August 2021

EXECUTIVE SUMMARY

Throughout history, complex societal problems have plagued societies with their ever-changing dynamic natures and sheer societal consequences. In efforts to address such complex problems, many actors have turned to participatory methods as a means to incorporate community level knowledge and produce community-oriented solutions while addressing societal concerns. Simultaneously, in the field of serious game design, a large body of research has studied the positive effects of using serious game play for societal intervention. However, fewer efforts have been allocated to exploring the effectiveness of participatory serious game design as a means for societal intervention. In other words, *can participatory serious game design act as an effective method for societal intervention?*

Such inquiry is the basis of this research, where the complex societal problem of *biased US policing* serves as the application of this study. A mixed methods approach to this research was implemented to explore how to develop interventions where serious game design concepts are to be created between US communities and police. Through a mixed methods approach involving quantitative elements of inferential and descriptive statistical analysis and qualitative elements of content analysis, sentiment analysis, and micro-interlocutor analysis, group brainstorming data, workshop observation data, and semi-structured group interview data was explored in an effort to understand if and how participatory serious game design can be leveraged effectively as a means to societal intervention in the context of US policing and beyond.

The outcomes of this research are two-fold. First, with regard to improving police-community relations through societal intervention, results from this research indicate that current relations between US police and BIPOC communities, specifically black communities, are in a dire state of mistrust. As such, any attempt at intervention between these two groups will likely be received with caution and scepticism. However, in demonstrating that intervention efforts are long-term oriented and not just "throw away" events, more willing engagement could be facilitated. In addition, this research has also demonstrated that police accountability in intervention efforts can exhibit to communities that such efforts to improve police-community relations are serious. In maintaining police accountability throughout interventions, communities may be more willing to welcome police efforts in engagement and relation building.

Second, in considering participatory game design as an intervention study, results from this study indicate that participatory game design has the potential to be an effective intervention method if it is implemented in a way that appropriately caters to the desired audience. In using participatory game design as a means to societal intervention, participant "buy-in" must be facilitated early on to ensure effective engagement. This study also alludes to the fact that participatory game design could be particularly effective as an intervention method when the topic of intervention is taboo, discomforting, or difficult to talk about, as the game-like nature of such intervention can provide an adequate amount of abstraction from reality that helps facilitate less anxiety-inducing dialogue. In a similar vein, the reality-abstracted game-like nature of participatory game design interventions could also be effective at fostering safe and inclusive spaces where all participants can feel able to engage and contribute to discussion, regardless of their backgrounds. Finally, with regard to using participatory game design as an intervention study, this research has exhibited that in ensuring a diversity of participant backgrounds and perspectives, echo chambers, polarity, and groupthink can be avoided in participatory game design-based interventions. Similarly, more meaningful intervention outcomes and creative problem solving has the potential to arise if a healthy amount of group conflict and pressure is managed appropriately within participatory game design interventions.

To the best of our knowledge, this research represents the first mixed methods study aimed at investigating the effectiveness of participative serious game design as a societal intervention method for biased US policing. Therefore, this study provides several potential scientific contributions to the fields of intervention science and serious game design, and it also has several potential implications in the context of society and public policy.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANCOVA	Analysis of Covariance
BIPOC	Black, Indigenous, People of Color
COP	Community-Oriented Policing
COPS	Community Oriented Policing Services
CPE	Center for Policing Equity
DPE	Design, Play, and Experience
IAT	Implicit Association Test
KAP	Knowledge, Attitude, and Practice
KMO	Kaiser-Meyer-Olkin Measure of Sampling Adequacy
LAPD	Los Angeles Police Department
MC-SDS	Marlowe-Crowne Social Desirability Scale
NLTK	Natural Language Toolkit
PCA	Principle Component Analysis
P-III	Play-Centered Iterative, Interdisciplinary, and Integrated
SARA	Scanning, Analysis, Response, and Assessment
SGD	Serious Game Design
TGD	Triadic Game Design
US	United States
VADER	Valence Aware Dictionary for Sentiment Reasoning

1

INTRODUCTION

Today, societies are plagued with a plethora of complex problems. These complex societal problems are complicated by their ever-changing dynamic natures, multiple actors with changing and conflicting views, and the sheer societal consequences associated with such problems (DeTombe, 2001). While human action is often the root cause of these problems, further societal intervention is necessary to address these issues (DeTombe, 2001).

In dire need of societal intervention is the case of United States (US) policing. Oppressive institutions and systematic discrimination within the US have fueled not only hesitance and opposition towards the police but also violent and bias policing (Gilbert & Ray, 2016; Bryant-Davis et al., 2017; Dukes & Kahn, 2017; Kahn & Martin, 2016). As demonstrated by countless protests which erupted across the country following the May 25, 2020 murder of an unarmed Black man, George Floyd, by Minneapolis police, many US citizens feel neither protected nor served by institutions that govern public safety (Dreyer et al., 2020; De Soto, 2018). While the US criminal justice system is supposed to be just for all, instances of inequality still persist regardless of the requirements of the U.S. Constitution (Robinson, 2002). Both implicit and explicit bias are determined to be major contributing causes for such inequality in US policing (Spencer, Charbonneau, & Glasser, 2016).

As illustrated by Duke, serious games can be used as intervention tools to address such complex societal problems (1974; 1980, as cited in DeTombe, 2001). In the instance of policing, serious games have been used for training purposes involving gameplay for crime scene investigation, investigative interviews, communication, and terrorism (Akhgar, Redhead, & Saunders, 2019; Lukosch, van Ruijven, & Verbraeck, 2012). While serious gameplay has been used as a means for societal intervention in policing, in a 1971 study by Gamson it was concluded that games for educational purposes are most successful when participants partake in the design and development of the game instead of gameplay. In the context of policing, this conclusion has led to the following question: Can participatory approaches to game design, involving both US law enforcement officers and community citizens, benefit biased US policing? With biased US policing serving as an application study, a curiosity to explore participatory serious game design as a societal intervention motivates this study.

In the following chapters, issues with United States policing and existing attempts at intervention are brought to light. Bias is defined, effective community-oriented policing (COP) methods are explored, and appropriate serious game design (SGD) methods for participative serious game design as a societal intervention method for biased US policing are identified. Methods and the experimental design to this intervention are outlined before both quantitative and qualitative results from police-community serious game design workshops are put forth. Finally, improving police-community relations through societal intervention is discussed before conclusions about the effectiveness of participatory game design intervention are drawn.

1.1 Societal Intervention

1.1.1 The Case of United States Policing

Considered by many to be a public health issue, biased policing and the resulting psychological and physical violence in the US is a grand societal problem (Helander & McNeill Brown, 2020; Cooper et al., 2004; Feldman, 2015; McGregor, 2016; Obasogie & Newman, 2017). A 2012 study by DeGue, Fowler, & Calkins concluded that the fatality rate for Black people in use of lethal force by law enforcement situations was 2.8 times greater than that for white people. A 2018 study by Edwards, Lee, & Esposito also determined that Black, Indigenous, and Latinx, people were more at-risk to death by police use of force than their white counterparts. As noted by Jee-Jyn García et al., "racism is a social determinant of health", and this notion is especially true in the context of policing and use of force situations (as cited in Obasogie & Newman, 2017). As a result of such reality, bias associated with US policing practices poses a major problem to US public health.

Diving deeper into the repercussions of biased policing, public health consequences associated with the trauma and anxiety of being subjected to police violence can result in community fragmentation (Gomez, 2016 as cited in Obasogie & Newman, 2017). In fear of being targeted by biased police practices, fewer residents may feel as if they can safely move freely across their communities (Gomez, 2016). In turn, limited "free-moving street life" has negative economic impacts for local shops, markets, and restaurants, as fewer patrons venture outside to support their local businesses (Gomez, 2016). Further highlighting economic and social consequences, a 2019 study by Insler, McMurrey, & McQuoid concluded that household charitable donations towards community initiatives were greatly reduced, especially amongst Black households, as a result of police militarization and the resulting violent use of force associated with it (Insler, McMurrey, & McQuoid, 2019).

Furthermore, biased policing can lead to unwarranted interactions and arrests which further exacerbate issues related to both social cohesion and economy. According to the Pew Research Center, of the US public, Black adults are about 5 times more likely than white adults to believe that their police encounters were unwarranted (DeSilver, Lipka, & Fahmy, 2020). When such unwarranted police interactions result in arrests, the obtained criminal record, regardless of a criminal conviction or not, greatly impacts an individual's access to education, employment, and housing (Curtis, Garlington, & Schottenfeld, 2013; Leisure, 2019; Evans, Szkola, & St. John, 2019; Pager, 2003; Evans & Porter, 2015; Uggen et al., 2014). According to a 2010 study by Weissman et al., about 55% of institutions of higher education collect and use criminal record information in their admissions decisions (as cited in Evans et al., 2019). Even more striking, a 2014 study by Pierce et al. concluded that amongst the higher education institutions surveyed, 35% of institutions flat out deny admission to anyone with a criminal record (as cited in Evans et al., 2019). Access to higher education facilitates further opportunity for employment, however criminal histories compound negative ramifications for those whose interactions with police have led to arrests (Evans et al., 2019). As demonstrated by countless studies, evidence suggests that criminal records pose as a major barrier to employment (Agan & Starr, 2017; Uggen et al., 2014; Pager, 2007; Solomon, 2012). US anti-discrimination laws to protect those with criminal records are not only inexistant in the context of employment, but also in the context of housing (Evans et al., 2019; Decker et al., 2015; Pager, 2003; Uggen et al., 2014; Evans & Porter, 2015). As a result, people with criminal history are even more at-risk to not receiving the assistance in which they need.

All of these compounding factors undermine the goal of public policy to equitably support all members of society (Curtis et al., 2013). As such, it should come as no surprise that the US government and law enforcement agencies have identified a need to limit the number of unwarranted police encounters, arrests, and convictions. In tackling this goal, law enforcement agencies have identified bias in policing as a major point for intervention. In limiting biased police practices, many unnecessary police-citizen interactions can be avoided, and further societal damage can be mitigated.

However, in limiting biased police practices and unnecessary police-citizen interactions, institutional challenges exist. Specifically, the American system of federalism, where central authorities and local authorities have different powers and neither authority can interfere with the other, poses organizational issues for improving US policing as a whole (Boettke, Palagashvili, & Piano, 2017). As a result of such system, the federal government has little control over state and local law enforcement practices (Boettke et al., 2017). Subsequently, and according to the US Department of Justice, 18,000 different police agencies operate at the local, county, state, and federal levels, and they also determine their own practices and policies (Banks et al., 2016). With such little federal oversight and hyper-localized control, it is incredibly challenging to monitor police activity at the agency level, much less at

the officer level. As such, police officer accountability is a growing and valid concern amongst the US public (Walker, 2001). To address such concerns, individual law enforcement agencies have adopted policies and training programs for improved practices.

1.1.2 Intervention for US Policing

Although many US law enforcement agencies have adapted policies and training programs with top-down approaches in an attempt to address bias and often violent policing, effective hierarchical intervention methods are still unknown (Bayley, 2008; Spencer et al., 2016). While a decentralized approach to police policy has not been incorporated fully in US policing institutions, community-oriented policing (COP) has been effective in parts of Europe, Africa, Central America, and Asia (Norges Miljo-Og Biovitenskaplige Universitetet, 2015; Kocak, 2018; Rake, 1987). With such bottom-up approach to policing, participatory design is embraced as citizens participate in decision-making processes, voice their needs, and influence organizational change (Geva & Shem-Tov, 2002).

However, while proactive community-based approaches to policing are often implemented with the intent of bringing such positive change, studies by Weisburd et al. (2019), Gordon (2020), and Ruteere & Pommerolle (2003) have indicated that reinforcement of oppressive institutions and inequitable police practices can result from poorly implemented COP. Notably, officer racial bias can often impact local policing strategy, thus resulting in racial disparities in community-focused police practices (Weisburd et al., 2019; Gordon, 2020). Interestingly, in many participatory approaches to police reform, citizens are arguably underrepresented in the decision-making process. For example, in a 2008 study by Toch, most of the participatory emphasis to the design and implementation of police reform was centered around rank-and-file officers. Studies by Govender (2020) and Schneider, Agee, & Chronopoulos (2020) also indicate that many previous attempts to involve citizens in police reform with participatory approaches were interpreted as performative. Some citizens have been dissatisfied with the level of community representation and the intensity of possible opportunities for participation in policing policy. Much of the intervention efforts fail to properly implement opportunities for citizen participation in ways that satisfy citizen needs for non-bias representation in policing.

1.1.3 Intervention via Serious Games

In addition to community participation-based interventions, serious gameplay has also been used as an intervention method for improved police practices. As defined by Abt (1970), "serious" games are those in which have been designed for purposes which are not solely for amusement (as cited in Djaouti et al., 2011; Laamarti, Eid, & El Saddik, 2014). They are designed and guided by institutional goals (Bogost, 2007). With these games, skills, knowledge, and competency can be acquired by players (Ritterfield, Cody, & Vorderer, 2009).

In the context of simulation games for law enforcement training, games have been designed for training purposes involving crime scene investigation, investigative interviews, communication, and terrorism (Akhgar, Redhead, & Saunders, 2019; Lukosch, van Ruijven, & Verbraeck, 2012). In the instance of a 2018 study by Sorace et al., European citizens were involved in serious games for police reform in order to try and increase citizen awareness of the potential values associated with citizen-police relationships. While this study investigated shifts in citizen perspective in COP from serious gaming, it did not emphasize shifts in law enforcement officer perspectives. Overall, serious games in the context of policing have been used more as tools for law enforcement officer training purposes via gameplay. More radical and innovative approaches to COP and serious gaming with police forces are necessary in order to elicit organizational learning and organizational change amongst law enforcement officers.

Radical and innovative, research suggests that participatory approaches to game design can be more effective to eliciting organizational learning and organizational change (Gamson, 1971; Druckman & Ebner, 2007; Akcaoglu, 2014; Vos, van der Meijden, & Denessen, 2011; Ke, 2014). According to Gamson (1971, p. 307) "playing a game may be a more active experience than listening to a lecture, but developing a game is more active still". In his 1971 study, Gamson concluded that games for educational purposes are most successful when participants partake in the design and development of the game instead of gameplay. In addition to cognitive benefits, motivational benefits also exist for those who participate in the design of serious games (Akcaoglu, 2014). A 2011 study by Vos, van der Meijden, and Denessen concluded that levels of intrinsic motivation and the use of deep learning strategies were more apparent in participants that were involved in constructing a game instead of playing a game. In their 2014 study about design-based learning and mathematical computer games, Ke concluded that participants who aided in the design of mathematical learning games developed more positive attitudes

towards mathematics as a subject. These results suggest that the process of game design can aid in positively changing the disposition that design participants have towards the particular subject of the game's interest (Ke, 2014). Such cognitive and motivational benefits have the potential to be particularly useful for the case of biased US policing and COP initiatives. Therefore, participatory game design activities should be explored as a means for intervention in such contexts.

1.2 Knowledge Gap & Research Questions

While existing research indicates that serious game design may be a more effective means to societal intervention than serious gameplay, it is unknown as to how participatory serious game design can be used as an effective societal intervention method in the context of community-based US policing. Furthermore, current academic research does not provide enough insight into how serious games can be effectively co-designed for such purpose. To address this knowledge gap, the following main research question has been posed:

Main Research Question

How can a participatory serious game design process for reduced police bias, involving both communities and police, be developed in the United States as a means for societal intervention?

Sub-Research Question

SQ1: *How can police bias be defined?*

SQ2: *Based on theory and practice, what are successful approaches to involving community members in interventions for reduced police bias that can inform a participatory serious game approach to intervention?*

SQ3: *Which serious game design methods are suitable for participatory serious game design processes between communities and police forces in the United States?*

SQ4: *What methods are suitable for measuring the effectiveness of a police-community serious game design process for reduced police bias?*

SQ5: *What can the results from this study conclude about reducing police bias and improving police-community relations through participatory serious game design-based intervention?*

SQ6: *What can the outcomes from this study deduce about participatory serious game design as a means for societal intervention in general?*

The objective of this research is to explore the effectiveness of participatory serious game design as a societal intervention method for biased US policing. The scientific significance of this research question is borne in the fact that different serious game design methods have the potential to affect the associated learning that different participants experience via involvement in participative serious game design processes. The societal significance of this research question stems from the fact that current efforts to mitigate against police bias are failing, and the participatory design of a serious game to address police bias could be explored as a potential solution.

1.3 Research Approach & Study Design

1.3.1 Mixed Methods Research Approach

In order to investigate the main research question, a mixed methods approach to the research, involving both quantitative and qualitative approaches, has been chosen for this research.

The effectiveness of a police-community serious game design process as a means for police intervention can be measured both quantitatively and qualitatively. Therefore, a mixed methods approach has the potential to provide a more holistic account to the research. With such approach, both quantitative and qualitative methods can be integrated in order to develop even further understanding of the research problem (Pardede, 2019). In considering both quantitative and qualitative methods to this research, the limitations of each type of data can be better balanced (Green et al., 2016). Such mixed methods

approach could also be helpful in understanding contradictions that may arise between both result types (Creswell & Clark, 2007 as cited in Green et al., 2016). Ultimately, with a mixed methods approach to the research, it is possible to elaborate, enhance, and clarify results from one method (e.g., quantitative) with the results from the other method (e.g., qualitative) (Schoonenboom & Johnson, 2017).

1.3.2 Before-After (Pre-Post) Intervention Study Design

In exploring the effectiveness of participative serious game design as a means for societal intervention for the case of biased policing, a before-after (pre-post) interventional study design can be implemented. According to Aggarwal & Ranganathan, "in intervention studies, the researcher actively interferes with nature - by performing an intervention in some or all study participants - to determine the effect of exposure to the intervention of the natural course of events" (2019, p. 137). In using an intervention study design for this research, the effect that participative serious game design workshops have on police officers and community members can be measured. In particular and seeing as this research is focused on understanding how the intervention workshops affect participants, a before-after (pre-post) interventional study design can be implemented. With such design, conclusions can be drawn on the basis of time in relation to the measurements of the intervention (Aggarwal & Ranganathan, 2019). With such study design, limitations do exist in the fact that the outcome of the intervention could be related to changes unrelated to the experimental design of the intervention itself (Aggarwal & Ranganathan, 2019). For example, in participating in a serious game design processes to address biased policing, participants may become more interested in seeking information on race and discrimination. In such instance, the knowledge gained through participant self-education would not be directly related to the intervention workshops, however it could still influence observed outcomes from the study. However, simultaneously, such instance would result in overall greater awareness of racial biases, and while such results would not be a direct result of the intervention workshops themselves, they would be an indirect and positive consequence of the workshops that may not have occurred without an intervention in the first place. Therefore, for the sake of this research, such study design is considered appropriate.

1.4 Report Structure

This report is structured into four parts: I: Preparation, II: Data Collection, III: Data Analysis, and IV: Synthesis. Table 1 below outlines this report structure in relation to report chapter content. A research flow diagram also outlines this report structure in Figure 1.

Table 1: Report Structure Outline

Part	Chapter
I. Preparation II. Data Collection	1. Introduction 2. Literature Review 3. Method & Experimental Design
III. Data Analysis	4. Quantitative Results 5. Qualitative Results
IV. Synthesis	6. Discussion 7. Conclusion

Part I of this report introduces the research problem. Chapter 1: Introduction defines the problem and provides context. Chapter 2: Literature Review, provides the theoretical underpinnings to this research and acts as input to the design of this research experiment. With input from the Literature Review presented in Chapter 2, the research method and experimental set up of the research are put forth in Chapter 3: Method & Experimental Design. Furthermore, Part II of this report highlights data collection via participatory game design workshops and associated surveys, observations, and semi-structured interviews with Chapter 3: Method & Experimental Design.

In Part III, analysis techniques are put forth for both quantitative and qualitative perspectives. More specifically, Chapter 4: Quantitative Results and Chapter 5: Qualitative Results outline the findings which result from such analysis.

Finally, Part IV of this report synthesizes the research. Results are discussed in Chapter 6: Discussion, and research strengths, limitations, recommendations, and implications are identified. The main research question is addressed in Chapter 7: Conclusion, and a summary of the research findings is outlined.

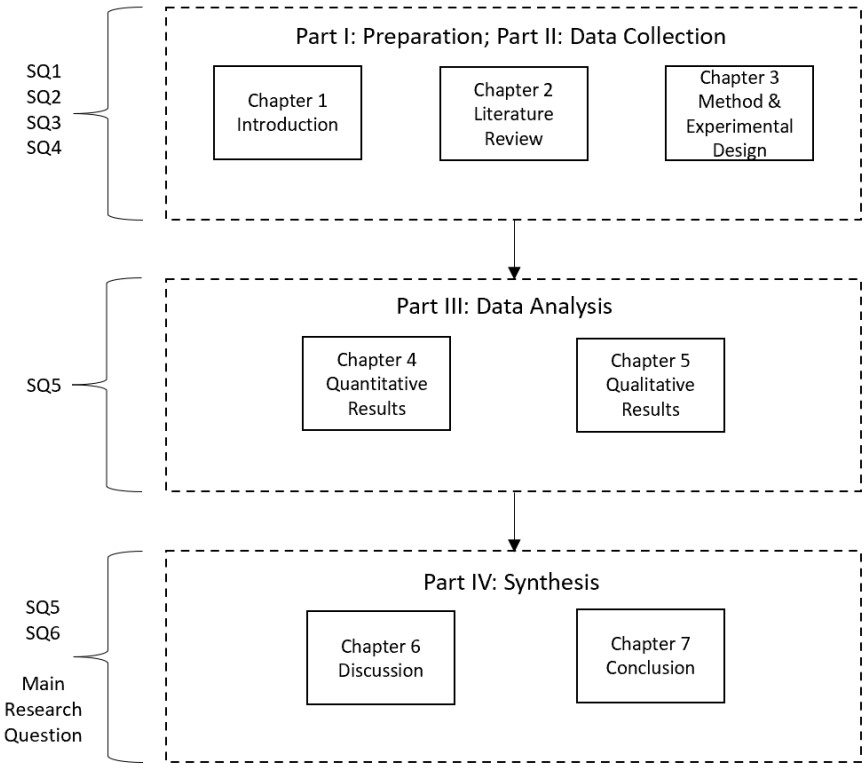


Figure 1: Research Flow Diagram

2

LITERATURE REVIEW

This chapter aims to provide the theoretical underpinnings to this research and act as input to the design of this research experiment. As such, background and foundation will be established in this chapter. First, in answering *SQ1*, the concept of bias will be explored and defined for the purpose of this study in section 2.1. Then, in order to address *SQ2*, successful approaches to traditional community member involvement in interventions for improved policing and reduced police bias will be discussed in section 2.2. In doing so, existing COP methods will be analyzed and criteria for effective COP will be determined. Serious games as a means for societal intervention will then be explored in section 2.3, and existing serious game design methods will also be analyzed. Given commonalities in effective COP methods and existing serious game design methods, serious game design methods that may be suitable to serve as input to the participatory serious game design process will be identified in order to address *SQ3*. Finally, chapter conclusions will be drawn in section 2.4.

2.1 Defining Bias

According to the Oxford Dictionary, bias is defined as an "inclination or prejudice for or against one person or group, especially in a way considered to be unfair" ("Bias," 2021). Concerning social issues of discrimination, bias can be held on the basis of factors such as gender, sexuality, age, race, and skin-tone (Greenwald, McGhee, & Schwartz, 1998). As individuals' identities are varied, nuanced, and complex, the intersectionality of many different identity factors (i.e., gender, sexuality, age, race, and skin-tone) contributes to compounding bias which further puts a person at-risk to marginalization (Hall & Carlson, 2016). For example, it is possible for a person to experience bias on the basis of being Black and on the basis of being a woman. The resulting compounding bias that is associated with identifying with both of these marginalized groups further exacerbates issues of inequality for such individual (Hall & Carlson, 2016). As briefly demonstrated by this example, bias and its compounding nature is complex and multi-faceted. However, this study focuses on bias in the context of race, and subsequent mentions of "bias" in this report are in reference to racial-based bias. Such bias can operate both implicitly and explicitly, and it can also be found embedded in the context of US policing.

2.1.1 Implicit Bias

According to Daumeyer et al., "implicit biases are associations and reactions that emerge automatically and often without awareness upon encountering a relevant stimulus" (Gawronski & Bodenhausen, 2006; Greenwald & Banaji, 1995 as cited in Daumeyer et al., 2019). This specific type of bias is unintentional, and the perpetrator lacks awareness of the fact that they possess discriminatory sentiments (Bargh, 1999; Devine, 1989; Gaertner & Dovidio, 1986 as cited in Devine et al., 2012). Despite individuals' conscious efforts to uphold attitudes and intentions that are non-prejudice, implicit biases can dictate behavior which perpetuates discrimination (Devine et al., 2012).

Notably, the resulting stereotypes that emerge from implicit biases can lead to outcomes that range from "seemingly mundane" to life threatening (Devine et al., 2012). For example, in a 2001 study by McConnell & Leibold, implicit bias resulted in poorer quality interactions between study participants and Black experimenters when compared with interactions between study participants and white experimenters. Contrasting in gravity, in a 2018 study by Hester & Gray which analyzed New York City stop-and-frisk police data from 2006-2013, young Black men, especially when perceived as tall, were consistently stereotyped as "threatening", and as a result of such perception they were also victim to a disproportionate amount of police attention. Considering that Black men are faced with about a 1 in 1,000 chance of being killed by police during interactions, such unwarranted police attention has the potential to be particularly dangerous for those subjected to implicitly biased interpretations (Edwards et al., 2019).

Seeing as implicit biases can lead to severe outcomes for those subjected to such covert prejudice, much research has been conducted with the aim of exploring unconscious prejudice. Most notably, the

Implicit Association Test (IAT) was developed as a means to measure sub-conscious attitudes and prejudice that test-takers have towards different groups (Greenwald et al., 1998; McConnell & Leibold, 2001). In taking the IAT, participants' attitude towards two different target concepts (e.g., white people and Black people) are measured via association with an attribute (e.g., desirable or undesirable) (Greenwald et al., 1998). During the computer-based test, participants are asked to rapidly categorize stimulus words with each target concept via different target concept and attribute combinations (Greenwald et al., 1998; McConnell & Leibold, 2001). The user's response times to categorize "white people" and "desirable", "Black people" and "desirable", "white people" and "undesirable", and "Black people" and "undesirable", are used to calculate the different associations that the user has towards each group (Greenwald et al., 1998). In measuring the performance difference for such categorization, the implicit association that a user may have for each group is quantified (Greenwald et al., 1998). The rapid-response type nature of the IAT eliminates social desirability, a tendency for respondents to give socially desirable responses as opposed to responses that reflect their true sentiments (Greenwald et al., 1998; Grimm, 2010). As a result, much research has emphasized a need to address implicit bias, as opposed to explicit bias, in order to address discrimination. However, in doing so, more acts of discrimination have become categorized as implicitly motivated, thus warranting less concern and more lenient punishment in addressing such discrimination (Daumeyer et al., 2019; Selmi, 2018). As a result, many researchers are calling for an emphasis on addressing explicit bias as opposed to implicit bias in order to address issues of inequality.

2.1.2 Explicit Bias

In contrast to implicit bias, explicit bias involves consciousness of the preferences that one may have towards different groups (Daumeyer et al., 2019). Such bias is not automatic, and the holder of such preferences is aware of their sentiment, and when willing, they can identify and communicate such disposition to others (Daumeyer et al., 2019). Explicit bias is reflected by both beliefs and attitudes (Cantone, 2021). Beliefs, informed by knowledge acquired via culture, education, and experience, guide attitudes (Seitz & Angel, 2020; Fishbein, 1966). When operationalized, the explicit bias, which is informed by both beliefs and attitudes, leads to intentionally prejudiced behaviour or practice (Leitner et al., 2016). Explicit bias results in overt racism, which can be expressed as purposeful stereotyping, hate speech, and discrimination as well as other prejudiced actions.

Seeing as society has deemed purposeful stereotyping, hate speech, and discrimination as socially undesirable, many critics claim that self-reported explicit bias is difficult to truly quantify as a result of issues with social desirability (Chae at al., 2017). In order to conform to socially acceptable standards, respondents' answers to questions about racial attitudes and bias may not truly express actual respondent sentiment. To circumvent such issue, many self-reported explicit racial attitude studies have developed self-reporting measures that are more indirect in nature (Sears, 1988; McConahay, 1986; Jacobson, 195 as cited in Axt, 2017). However, results from a 2017 study by Axt indicate that more indirect assessments of racial attitudes weaken measurement quality, as construct irrelevant attitudes contributed to responses. Instead, results from this study suggest that the best way to quantify someone's explicit racial bias is to simply ask them directly (Axt, 2017). Results from this study also suggest that more direct assessments of explicit racial bias were stronger indicators of implicit racial bias amongst participants (Axt, 2017). In highlighting more direct approaches to quantify explicit bias, implicit bias can also be indirectly addressed. In doing so, issues related to lenient repercussions for acts of discrimination that have been categorized as implicitly motivated can also be mitigated, as explicit bias remains at the forefront of concern when addressing prejudice attitude.

In their 2011 work, Iyengar, Messing, & Hahn, measure explicit racial bias with survey indices related to overt racism and racial resentment. Additionally, in their work, self-reported thermometer ratings for warm/cold feelings towards Black people and white people also provide insight into the explicit racial attitude that respondents have (Iyengar et al., 2011). The overt racism index in their study is based on respondent responses with regard to trait association ratings for Black people and white people (Iyengar et al., 2011). Adapting the work of Kinder & Sanders (1996), Iyengar et al. (2011) also use a variety of agree-disagree type questions about respondent beliefs related to racial equality, minority groups, and individualist culture, to define a racial resentment index.

2.1.3 Police Bias

As defined by the *Center for Policing Equity (CPE)*, police bias can be conservatively estimated with race-based disparate outcomes in police contact (Goff et al., 2016). In identifying racial disparities, *CPE* recognizes that crime, poverty, resistance, drug use, alcohol use, and officer perception all play a role in

police interactions and subsequent outcomes (Goff et al., 2016). As such, *CPE* acknowledges that unequal treatment of different demographic groups by the police is not always on the basis of racial bias alone. Unfortunately, however, racial resentment, defined by Kinder & Sears (1981) as "a blend of anti-Black affect and the kind of traditional American moral values embodied in the Protestant ethic", can still greatly impact officers' beliefs, attitudes, and subsequent policing tactics and interactions (as cited in Samson, 2015, p. 164).

For example, as stated by former Los Angeles Police Department (LAPD) chief of police, Bernard Parks, "It's not the fault of the police when they stop minority males or put them in jail. It's the fault of the minority males for committing the crime. In my mind, it is not a great revelation that, if officers are looking for criminal activity, they're going to look at the kind of people who are listed on crime reports" (Glaser, 2014, p. 96 as cited in Goff et al., 2016). As illustrated with this quote, while a plethora of unintentional race-based disparities may lead to unequal outcomes in terms of police contact, explicit bias is still a factor in disparate police contact. As such, explicit bias in the context of US policing is the focus of this study. Furthermore, police bias can be defined as explicit bias in the context of policing, where purposeful stereotyping is used to understand community dynamics and potential criminal activity. As indicated by Goff et al. (2016), existing literature is limited in how such bias can be assessed with metrics. Conceptual issues exist in measuring explicit police bias, as direct observation of police encounters is needed to understand the true dynamics which produced the interaction outcomes (Goff et al., 2016). As such, in order to operationalize this definition for the scope of this study, police bias can be "measured" qualitatively in observing the content of subsequently designed serious game concepts and their inclusion of purposeful stereotyping to understand community dynamics and potential criminal activity.

2.2 Community-Oriented Policing (COP)

Community-Oriented Policing (COP) emphasizes police and community member partnerships to address underlying issues of crime (Ungar & Desmond Arias, 2012; Gill et al., 2014). With such decentralized approach, crime prevention is prioritized over crime response (Ungar & Desmond Arias, 2012; Gill et al., 2014). Broad in its definition, in its essence COP refers to any policing program or strategy that depends on citizens' involvement as co-producers to crime prevention (Thurman, 1995 as cited in Kerley & Benson, 2000). While different types of COP methods exist such as foot patrolling, policy councils, and youth services, COP can really be described as a philosophy to policing as opposed to a prescription of methods (Goldstein, 1987 as Kerley & Benson, 2000; Rosenbaum & Lurigio, 1994). With such philosophy, advocates of COP insist that reductions in crime, disorder, and fear can result (Kerley & Benson, 2000). Research has also demonstrated that COP methods can also result in increased community satisfaction and faith in law enforcement (Gill et al., 2014). For this reason, police operations across the country are increasingly more interested in a COP approach to enforcing law (Gaines, 1993; Wycoff, 1991 as cited in Gianakis & Davis, 1998; Burruss & Giblin, 2014; Dario & Crichlow, 2020).

However, research also suggests that the vagueness associated with COP has also resulted in poor and haphazard implementation of the policing method (Wycoff, 1991 as cited in Gianakis & Davis, 1998; Eck, 2006; Gaines, 1993; Dario & Crichlow, 2020). As a result of such vagueness, issues with COP also arise on the basis of law enforcement organization (Gianakis & Davis, 1998; Gaines, 1993). Institutional level organization does not always evolve with changes to officer level learning (Gianakis & Davis, 1998). As a result, the institutions that govern law enforcement do not have the necessary infrastructure to support police officers with new innovative COP perspectives (Gianakis & Davis, 1998; Gaines, 1993). Consequently, many critics argue that COP is simply a linguistic charade where COP is mentioned in police department mission statements without any characterization of *how* such method is actually implemented in practice (Dario & Crichlow, 2020). While challenges do exist, literature does provide some baseline for analysis of best practices for involving community members in policing.

2.2.1 A Framework for COP Methods

Different COP methods have been employed with various levels of success. In order to determine which methods for serious game design (SGD) could be most appropriate for a police-community serious game design process, different COP methods must be analyzed. In doing so, commonalities in elements between successful COP methods and SGD methods can be identified. Given the commonalities in effective COP methods and existing SGD methods, SGD methods that may be best suited to serve as input to the participatory serious game design process are identified. More so, according to Mildner & Mueller (2016), the examination of traditional interventions to address the identified "serious" subject

of interest can greatly aid in the design of a serious game for such issue, for "the creative process of turning an abstract theory into a practical intervention or into a game shares many similarities" (p. 65). The results of such analysis and framework ultimately assist in informing a choice of SGD methods to serve as the foundations for the police-community design processes. The results will also help establish criteria for successful COP measures, thus providing a benchmark of comparison for quantifying the effectiveness of the participatory serious game design processes for reduced police bias.

Both peer-reviewed literature and grey literature have been assessed in order to identify the following COP methods. For the literature review process, a search plan was conducted. First, an analysis of the research topic was conducted via identification of key concepts of "COP", "Community-oriented policing", and "police". Then, alternative search terms for each concept were identified and mapped. Searches were conducted in Scopus. Once relevant literature was identified, further searching was conducted with the "snowball method" and the "citation searching" method. With the "snowball method", highly relevant publications were looked into further. Of each highly relevant publication, keywords were noted and used in other searches, and references were further investigated. The other works of authors' with highly relevant publications were also looked into. With the "citation searching" approach, citations of key publications were investigated in order to see how other authors have built upon the scientific research. Trial and error were used during this search plan, and steps of the process were iterated over many times. In iterating through the search plan and narrowing down the initial results, the following COP strategies were determined to be the most relevant literature to review with regard to successful COP methods.

After reviewing the literature and identifying four methods for successful COP intervention, contents of each approach have been summarized in Table 2 below. Within this table, each analyzed COP method and its key take-aways of emphasis have been highlighted.

Table 2: Summary of model emphasis for successful COP methods

COP Method	Method Emphasis						
	Setting the Stage	Problem Identification	Systems Analysis	Alternative Strategy Formulation	Continuous Improvement	Multi-Stakeholder Involvement	Long-Term Orientation
Gaines, 1993	X	X		X	X	X	X
Cordner, 1999		X	X		X	X	X
Rau et al., 2020		X			X	X	X
COPS, 2014	X	X			X	X	X

2.2.1.1 Gaines, 1993

With the intent of outlining a standard manner for COP implementation, Gaines (1993) developed an eight step strategy for successful community focused police practices. In their COP method, Gaines (1993) highlights the importance of steps: 1. Performance Gap; 2. Recognition of a Need for Change; 3. Creating a Proper Climate for Change; 4. Diagnosing the Problem; 5. Identifying Alternative Strategies; 6. Select the Strategy; 7. Determine and Develop an Implementation Strategy; and 8. Evaluating and Modifying the Strategy.

1. Performance Gap

In the first step of Gaines' (1993) framework, a need to assess the gap in actual performance and outlined organizational expectations is required. In first acknowledging that actual policing outcomes may not meet the expectations outlined by the organization, it can be ensured that subsequent COP efforts are not performative and that instead, they work towards compensating the measurable difference in performance (Gaines, 1993).

2. Recognition of a Need for Change

The second step of Gaines' (1993) framework highlights a need for police organizations to recognize a need for change and to shift towards an emphasis of effective problem-solving. As a part of effective problem-solving, police organizations must be more preventative and solve issues before systems have deteriorated past a point of reasonable repair (Gaines, 1993).

3. Creating a Proper Climate for Change
In Step 3, change to current policing systems must be encouraged by law enforcement personnel, community members, and various external organizations (Gaines, 1993). However, in order to do so, the perceptions and values of both law enforcement officers and community members must be positively influenced.

4. Diagnosing the Problem
The fourth step of Gaines (1993) COP framework requires proper community problem identification. During this step, Gaines emphasizes the importance of effective management structures for identification of such community issues (Gaines, 1993).

5. Identifying Alternative Strategies
Step 5 of this COP framework requires that a situational approach to problem-solving is implemented with non-traditional methods (Gaines, 1993). In addition, an interdisciplinary approach to problem solving is encouraged as law enforcement officers are encouraged to engage with other governmental organizations, the public, and private organizations in order to successfully address issues of concern (Gaines, 1993).

6. Select the Strategy
Step 6 of Gaines' (1993) COP method calls for the delegation of roles and responsibilities within a system. While this method emphasizes the need for decentralized approaches to problem-solving, more centralized approaches to strategy development are important for successful COP decision-making (Gaines, 1993).

7. Determine and Develop an Implementation for Strategy
In Step 7, Gaines (1993) recognizes that a balance between strategy flexibility and guidance must exist. COP tactics must be determined and developed with a certain degree of freedom and flexibility, however, simultaneously, adequate guidance must also guide COP activities (Gaines, 1993).

8. Evaluating and Modifying the Strategy
In Step 8 of Gaines' (1993) COP framework, the importance of iterative strategy evaluation and modification in a COP problem-solving context is put forward.

2.2.1.2 Cordner, 1999
In an attempt to strategize COP as a method for policing, Cordner (1999) created a four-dimensional framework for COP. In their method, twelve elements of COP are grouped into dimensions that are of: I. philosophical, II. strategic, III. tactical, and IV. organizational importance (Cordner, 1999 as cited in Dario & Crichlow, 2020; Coquilhat, 2008). In order for a COP method to be successful, each element of each dimension of Cordner's framework must be upheld to specific criteria (Coquilhat, 2008). In collaboration with Cordner, New Zealand Police have further adapted this COP strategy for better realization of COP benefits (Coquilhat, 2008).

I. Philosophical Dimension
The Philosophical Dimension of Cordner's 1999 framework emphasizes the importance of understanding community values and how police functionality is to serve such values (Dario & Crichlow, 2020; Coquilhat, 2008). Within this dimension, three core community policing elements of *Citizen Input*, *Broad Function*, and *Personal Service*, are necessary for ensuring success in COP operations from a philosophical standpoint (Coquilhat, 2008). The element of *Citizen Input* refers to police engagement with community and extensive community ascertainment of problem solutions (Coquilhat, 2008). *Broad Function* highlights a need for continuous organizing amongst the police, public, and other external agencies (Coquilhat, 2008). Finally, the element of *Personal Service* in the Philosophical Dimension of Cordner's framework emphasizes a need for a customer-service approach to policing where police are perceived as accessible and specific officer long-term involvement in specific communities is maintained (Coquilhat, 2008).

II. Strategic Dimension

The Strategic Dimension of Cordner's (1999) framework focuses on translating philosophical elements into strategy in order to increase community trust in police systems (Coquilhat, 2008; Dario & Crichlow, 2020). Within this dimension, elements of *Re-Oriented Operations*, *Prevention Emphasis*, and *Geographical Focus* exist (Coquilhat, 2008). With the element of *Re-Oriented Operations*, tools are developed for crime prevention, and interactive long-term solutions for community well-being are established (Coquilhat, 2008). *Prevention Emphasis* places further merit on proactively preventing crime from occuring as opposed to reacting to its instances (Coquilhat, 2008). Finally, *Geographical Focus* actualizes the philosophy of officer specific long-term involvement in specific communities (Coquilhat, 2008). With this notion, locally based officers improve police communication, responsibility, and accountability as further community forms as a result of officers being linked to a specific neighborhood (i.e., foot patrolling) and being able to flexibly meet the specific needs of such communities (Coquilhat, 2008).

III. Tactical Dimension

The Tactical Dimension of Cordern's (1999) framework ensures that philosophical and strategic outlooks are actualised (Coquilhat, 2008). Elements within this dimension include *Positive Interaction*, *Partnerships*, and *Problem Solving* (Coquilhat, 2008). With the *Positive Interaction* element, trust, knowledge, and problem-solving skills are developed via positive police-community interactions (Coquilhat, 2008). Similarly, the element of *Partnerships* highlights the importance of not only positive police-community interactions, but also police-community collaboration (Coquilhat, 2008). With such collaboration, a wide range of issues must be addressed via targeted responses to each specific issue (Coquilhat, 2008). Interactive and iterative *Problem Solving* further emphasizes the importance of community involvement in identifying and addressing community issues alongside the police (Coquilhat, 2008).

IV. Organizational Importance Dimension

The Organizational Importance Dimension of the Cordern (1999) COP framework highlights a need for organizational change within police systems in order to encourage decentralized and participatory problem solving (Dario & Crichlow, 2020; Coquilhat, 2008). Elements within this dimension include *Structure*, *Management*, and *Information* (Coquilhat, 2008). The element of *Structure* refers to the employment of long-term strategies, training, and support to promote organizational goals aligned with community support (Coquilhat, 2008). *Management* highlights the importance of executive leadership within police systems to support communities in collaborative problem-solving (Coquilhat, 2008). Finally, *Information* emphasizes the importance of using qualitative (instead of quantitative) measurements in order to gauge the performance of COP systems (Coquilhat, 2008).

2.2.1.3 Rau et al., 2020

While not specifically labeled as COP, Rau et al. (2020) from *CPE* and *PolicyLink* developed a toolkit for equitable public safety practices which outlines the importance of external accountability and civilian oversight in policing. According to Rau et al. (2020), civilian oversight entities must first be established through either local government or ballot initiative. Typically, these entities are made up of professional staff with key expertise and community volunteers (Raul et al., 2020). Following entity formation, the scope of the entity's functions must be defined (Rau et al., 2020). In their functioning as oversight entities, civilian oversight groups must identify patterns of misconduct in police practices and iteratively work with the department for continuous improvement (Rau et al., 2020). Above all, it is crucial for civilian oversight groups to develop a professional and positive working relationship with law enforcement, so to ensure that community values are prioritized, and change is achieved (Rau et al., 2020). While Rau et al. (2020) acknowledge that there is not a single correct size and model for civilian oversight committee formatting, the organization must be well balanced in terms of responsibilities, resources, and community needs in order to act effectively.

2.2.1.4 COPS, 2014

The Community Oriented Policing Services (COPS) of the US Department of Justice has identified a framework for COP that comprises three main components: 1. Community Partnerships; 2. Organizational Transformation; and 3. Problem Solving (COPS, 2014). According to COPS (2014), in order to effectively support communities, increase public safety, and reduce social disorder all three of these elements must be appropriately balanced.

1. Community Partnerships

The Community Partnerships component of COPS (2014) framework highlights how collaborative partnerships between police and community can lead to increased public trust in police systems. Additionally, through active and interdisciplinary partnerships with other government agencies, community members and groups, non-profit organizations, service providers, private businesses, and media, law enforcement agencies can promote collaborative problem-solving (COPS, 2014). With such multi-stakeholder involvement, a variety of different perspectives and resources can be mobilized effectively (COPS, 2014).

2. Organizational Transformation

In shifting policing practices towards a more community-driven model, COPS (2014) recognizes the need for organizational transformation in aligning law enforcement management, structure, and information systems to best support the public. While philosophical shifts in approaches to policing are often associated with COP efforts, changes to law enforcement infrastructure must also ensue in order to encourage proactive problem solving (COPS, 2014). In doing so, long-term oriented structural changes must be made to leadership style, labor relations, decision making, strategic planning, and overall policing procedures (COPS, 2014).

3. Problem Solving

Finally, COPS' (2014) framework highlights a need to ensure engaging and proactive problem solving comprised of problem identification, systematic examination, response development, and response evaluation. With such approach, solutions to underlying issues are addressed before negative externalities occur (COPS, 2014). Innovative problem solving drives decision making within this component (COPS, 2014). The SARA (scanning, analysis, response, and assessment) model has been identified by COPS (2014) as an efficient means for eliciting structured problem solving in COP systems. With the SARA model, first problem identification (scanning) must ensue before research into the issue can commence (analysis) (COPS, 2014). Then, solutions for long-term situational improvement are developed (response) (COPS, 2014). Finally, the success of the intervention is evaluated (assessment) (COPS, 2014).

2.3 Serious Game Design (SGD)

Serious games provide a risk-free space, abstracted from reality, where complex societal problems can be explored (Abt, 1987). Research has demonstrated that serious games are an effective and innovative means for eliciting both organizational learning and organizational change when they are designed explicitly and carefully (Ribeiro et al., 2012; Laamarti et al., 2014). For this reason, serious gameplay has been employed in institutions related to healthcare, government, military, emergency response, and environment (Bogost, 2007).

For example, a 2015 study by Sorbring, Bolin, & Ryding investigated the use of game-based intervention to combat adolescent dating violence. In their study, Sorbring et al. (2015) examined Swedish adolescents' experience with playing an online serious game as an intervention method to address the sensitive issue of violence in teen dating. Results of this study indicated that young people perceived the experience as positive and found the game-based intervention to be interesting. However, the adolescent participants of this study did also provide feedback which indicated that the game was "not a 'real' game, but still okay" (Sorbring et al., 2015, p. 130). In this context, participants expressed some elements of boredom and frustration with the game content. As demonstrated by Ampatzidou & Gugerell (2017), one means to address issues with domain content, local context embeddedness, and playability is to use participatory approaches to serious game design. In involving different actors in the design process of a serious game, perceptions, goals, and interests of future players can be better reflected in the game, thus ensuring more meaningful learning outcomes for future players (Ampatzidou & Gugerell, 2017). Simultaneously, associated learning also exists for participants in the game design process (Ampatzidou & Gugerell, 2017).

Given the associated learning that participants experience during the game design process, research suggests that serious game design may be more suitable than serious game play as an intervention method. In his 1971 study, Gamson concluded that games for educational purposes are most successful when participants partake in the design and development of the game instead of gameplay. Gamson argued that as designers, participants can use game concepts and rules as resources that can be experimented with for different desired outcomes (Gamson, 1971). Similarly, a 2007 study by Druckman and Ebner also explored the educational benefits associated with game design and game

play. In the study which compared the relative benefits associated with participation via design or participation via role-play in a simulation game, design participants demonstrated better concept learning than play participants (Druckman & Ebner, 2007). Design participants also demonstrated a better understanding of concept relatedness when compared to play participants (Druckman & Ebner, 2007). A 2014 study by Akcaoglu also concluded that game design activities provide participants with the opportunity to learn, practice, and develop skills in problem-solving. As such, serious game design has the potential to be more effective than serious game play in eliciting positive change via intervention. However, while it is known that serious game design may be an effective means for intervention, the particular method to serious game design that may be most effective for intervention is unknown. Thus, in order to investigate the theoretical problem of the main research question to this study, different serious game design methods must be considered.

2.3.1 A Framework for SGD Methods

In achieving the benefits associated with participatory game design processes, different serious game design (SGD) methods can be considered. For this study, I have analyzed multiple SGD methods for how well they support participation in design for serious game development. The reviewed SGD methods were chosen on the basis of their relevance and innovation. Below, each analyzed SGD method and its key take-aways of emphasis have been highlighted in Table 3. Subsequently, the results of this analysis and framework ultimately inform a choice of SGD methods to serve as the foundations for the police-community design processes.

Table 3: Summary of method emphasis for analyzed SGD methods

Method Emphasis	SGD Method				
	Duke & Geurts, 2004	Kortmann & Harteveld, 2009	Fullerton, 2008	Winn, 2009	Abeele et al., 2012
Setting the Stage	X	X	X		
Problem Identification	X	X		X	X
Systems Analysis	X	X		X	X
Alternative Strategy Formulation	X	X			
Continuous Improvement	X	X	X	X	X
Multi-Stakeholder Involvement	X	X	X	X	X
Interdisciplinary Approach				X	X
Storytelling Emphasis				X	X
Design Approach	Linear	Iterative	Iterative	Linear	Iterative
Technology Platform	Analog & Digital	Analog & Digital	Digital	Analog & Digital	Analog & Digital

2.3.1.1 Duke & Geurts, 2004

First conceived in 1974, Duke's methodology for serious game design has been further adapted in collaboration with Geurts to support serious gaming as a means to solve problems, improve communication, and elicit organizational change (Peters & Van de Westelaken, 2014; Duke & Geurts, 2004). In their "standard" method, Duke & Geurts put forth five phases consisting of 21 total steps, for successfully organizing the design of a game (Duke & Geurts, 2004; Kortmann & Harteveld, 2009). As demonstrated by Figure 2, the phases of Duke & Geurts (2004) sequential game design method include: I. Setting the Stage for the Project; II. Clarifying the Problem; III. Designing the Policy Exercise; IV. Developing the Exercise; and V. Implementation.

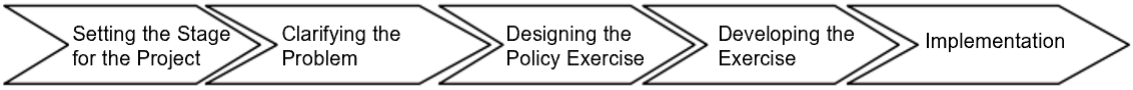


Figure 2: Phases to Duke & Geurts (2004) "standard" game design methodology adapted from Kortmann & Harteveld, 2009

I. Setting the Stage for the Project

Phase I of this method includes five steps (1. Administrative Set-Up, 2. Define the Macro-Problem, 3. Establish Goals and Objectives for the Project, 4. Project Objectives/Methods Employed Matrix, and 5. Specifications for Game Design) which are to be completed iteratively (Duke & Geurts, 2004). In this phase, administrative details are addressed to ensure successful guidance of the game design phases that follow (Duke & Geurts, 2004). First, a clear definition of the client with delineated roles and responsibilities is established before proper authorization of the project is granted (Duke & Geurts, 2004). Then, a well-defined problem statement, which outlines the issue that the game aims to address through the eyes of the stakeholders, is established (Duke & Geurts, 2004). Specifications to this problem statement are also outlined (Duke & Geurts, 2004). Following problem statement and specification delineation, the goals and objectives of the project are defined (Duke & Geurts, 2004). During this process, goals are broken down into specific sets of objectives that best address the concerns of participants, and these objectives are used to benchmark progress throughout the project process (Duke & Geurts, 2004). A project-objectives/methods-employed matrix is then used to review a range of different methods for addressing macro-problems, so to ensure that a game-based approach is appropriate (Duke & Geurts, 2004). The pros and cons of using a game to address the issue are then considered (Duke & Geurts, 2004). Finally, technical specifications for the game design are put forth (Duke & Geurts, 2004). These specifications pragmatically outline constraints and expectations for the project and phases to come (Duke & Geurts, 2004).

II. Clarifying the Problem

During Phase II (steps 6. Defining the System, 7. Displaying the System, and 8. Negotiating Focus and Scope with the Client), literature reviews, interviews, and schematics are used to cognitively map the project process (Duke & Geurts, 2004). Objectives are further clarified, and the problem statement is further developed (Duke & Geurts, 2004). First, a systems analysis is conducted (Duke & Geurts, 2004). This systems analysis is then displayed with an iteratively constructed schematic representation of the problem (Duke & Geurts, 2004). Finally, only key factors of the complete problem are selected from the schematic for incorporation into the final design of the game (Duke & Geurts, 2004).

III. Designing the Policy Exercise

Phase III (steps 9. Systems Components/Gaming Elements Matrix, 10. Definition of Gaming Elements, 11. Repertoire of Techniques, 12. Select a Format for the Game, and 13. Concept Report) of this method involves creating a blueprint for the game exercise (Duke & Geurts, 2004). First, the schematic is transformed to model a game as opposed to the defined problem (Duke & Geurts, 2004). System components are transcribed into game elements with a systems components/gaming elements matrix (Duke & Geurts, 2004). With all participants' participation in building the matrix, a storyline and all gaming elements are outlined through iterative and emergent brainstorming (Duke & Geurts, 2004). Then, the outlined game elements are developed into complete definitions (Duke & Geurts, 2004). In concretely defining these game elements, the character, purpose, and output of each game element is identified (Duke & Geurts, 2004). These elements must then be integrated with one another, so to ensure logic and consistency in the game (Duke & Geurts, 2004). Following this process, preliminary game material mockups are constructed (Duke & Geurts, 2004). These mock-up materials are then translated into more specific gaming techniques before the game format is finalized (Duke & Geurts, 2004). Finally, an interactive process with both designers and external stakeholders results in the drafting of a concept report (Duke & Geurts, 2004). This report outlines the design decisions and the game design plan for the project (Duke & Geurts, 2004).

IV. Developing the Exercise

During Phase IV (steps 14. Build, Test, and Modify the Prototype, 15. Technical Evaluation, and 16. Graphic Design and Printing) of Duke & Geurts (2004) method, the game is developed into a more tangible form. Game elements are designed and integrated with one another in iterative runs for game calibration (Duke & Geurts, 2004). In doing so, a first initial prototype is built, tested, and modified for improvement (Duke & Geurts, 2004). Then, with a game objectives/game elements matrix, a systematic comparison between the prototyped game and the initially proposed project objectives and specifications is made in order to evaluate whether or not the game design meets the pre proposed criteria for the project (Duke & Geurts, 2004). Finally, the prototype is transformed into a finished presentation of higher professional quality (Duke & Geurts, 2004).

V. Implementation

In the final and fifth phase of the Duke & Geurts (2004) game design method, the game becomes the responsibility of the client. In Phase V (steps 17. Integrate the Game into the Client's Environment, 18. Facilitating the Exercise, 19. Dissemination, 20. Ethical and Legal Concerns, and 21. Reporting to the client), the game must be properly implemented in order to achieve the initial objectives of the intervention (Duke & Geurts, 2004). First, it must be ensured that the game exercise is properly integrated into the client's environment (Duke & Geurts, 2004). In doing so, appropriate training and education must be provided to the organization for proper game facilitation (Duke & Geurts, 2004). Once properly prepared, the game is then actually implemented and played under facilitation in the organization's controlled environment (Duke & Geurts, 2004). In completing Phase V, it must be ensured that continued use of the game is possible for the organization (Duke & Geurts, 2004). Similarly, it must be ensured that participants are protected throughout the process (Duke & Geurts, 2004). In addressing Phase V of this game design method, proper closure for the project is provided as the game documentation and results are made available to the organization for future evaluation (Duke & Geurts, 2004).

Detailed descriptions pertaining to all phases and steps for this game design method can be found in "Chapter 8 - Designing the Policy Exercise" of *Policy Games for Strategic Management* (Duke & Geurts, 2004).

2.3.1.2 Kortmann & Harteveld, 2009

Originally conceptualized by Harteveld (2011), Kortmann & Harteveld built upon the methodology of Triadic Game Design (TGD) for developing serious games to reflect a more dynamic process inspired by software engineering (Kortmann & Harteveld, 2009). In their method, an emphasis on balancing components of "reality", "meaning", and "play" is executed with a nonlinear and agile approach throughout iterative development of scope, design, and built artefact (Kortmann & Harteveld, 2009). In this method, all three components must be developed equally (Harteveld, 2011). Simultaneously, each component must also be developed fully, so to ensure that the specific evaluation criteria for each component are met (Harteveld, 2011). Given that different components have different evaluation criteria for determining a "good" game, the TGD process can be considered multidisciplinary (Harteveld, 2011). As demonstrated by Figure 3 and in accordance with Harteveld's (2011) original method, "reality", "meaning", and "play" must also be developed in parallel (Kortmann & Harteveld, 2009). Notably, within this SGD method, "Scope" coincides with "I. Setting the Stage for the Project" and "II. Clarifying the Problem" from the Duke & Geurts SGD method (Kortmann & Harteveld, 2009). "Design" corresponds with "III. Designing the Policy Exercise", and "Build" corresponds with "IV. Developing the Exercise" from Duke & Geurts SGD method (Kortmann & Harteveld, 2009). Finally, the fourth Kortmann & Harteveld (2009) TGD phase of "Test" coincides with Duke & Geurts phase of "IV. Implementation".

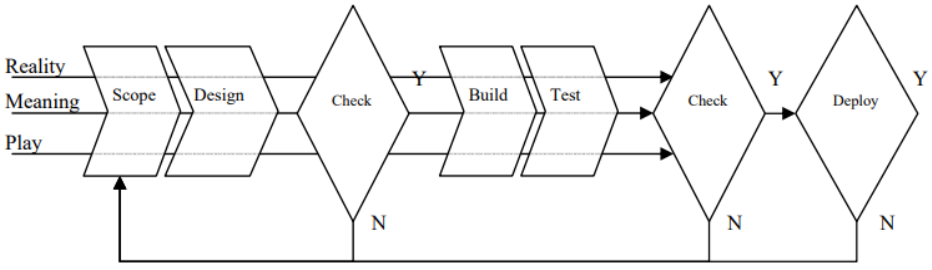


Figure 3: The Triadic Game Design (TGD) method with an agile approach by Kortmann & Harteveld, 2009

Reality

The component of "reality" refers to a representation of the real world in the game (Harteveld, 2011; Harteveld et al., 2010). Such representation can be facilitated via game players (with domain knowledge and opinions) or through actual real-world representation within the game world (Harteveld, 2011; Kortmann & Harteveld, 2009). Such "reality" is grounded in subject-specific disciplines, and it determines game variables and definitions (Harteveld, 2011; Harteveld et al., 2010; Kortmann &

Harteveld, 2009). With this approach to game design, a game can be evaluated on the basis of "reality" with criteria related to fidelity, realism, and validity (Harteveld et al., 2010; Kortmann & Harteveld, 2009).

Meaning

"Meaning" is in reference to how purposeful impact and important significance can result from the game experience (Harteveld, 2011; Harteveld et al., 2010; Kortmann & Harteveld, 2009). With elements of learning sciences, psychology, and semiotics, such "meaning" can be developed (Harteveld, 2011; Harteveld et al., 2010; Kortmann & Harteveld, 2009). Communication and learning are crucial in establishing "meaning" to the game (Harteveld, 2011; Harteveld et al., 2010; Kortmann & Harteveld, 2009). Criteria for evaluating a game on the basis of "meaning" include reflection, transfer, and relevance (Harteveld et al., 2010; Kortmann & Harteveld, 2009).

Play

The "play" component refers to how interactive, engaging, and immersive a game experience is (Harteveld, 2011; Harteveld et al., 2010; Kortmann & Harteveld, 2009). While games are to be designed with purpose and meaning, they must also be enjoyable to play (Harteveld, 2011: Harteveld et al., 2010; Kortmann & Harteveld, 2009). With the "play" component, actors, rules, resources, challenges, competition, and other game elements are selected (Harteveld, 2011; Harteveld et al., 2010; Kortmann & Harteveld, 2009). "Play" criteria elements to consider when evaluating a game include engagement, fun, and immersion (Harteveld et al., 2010; Kortmann & Harteveld, 2009).

2.3.1.3 Fullerton, 2008

Fullerton (2008) developed a methodology for game design that takes a playcentric approach. In her iterative design method, Fullerton (2008) advocates for early player feedback throughout all phases of game development. In doing so, players help build game mechanics from the ground up to ensure that player experience is central in the design of the game (Fullerton, 2008). The seven steps of Fullerton's (2008) playcentric approach include: 1. Brainstorming; 2. Physical Prototype; 3. Presentation (Optional); 4. Software Prototypes(s); 5. Design Documentation; 6. Production; and 7. Quality Assurance.

Within each specific step of the playcentric approach to game design, iterations of idea generation, idea formalization, idea testing, and result evaluations take place (Fullerton, 2008). For the iterative process, player experience goals must first be set (Fullerton, 2008). Then, an idea is generated and formalized (Fullerton, 2008). The formalized idea is tested against the set player experience goals and the results are evaluated (Fullerton, 2008). The outcome of the evaluated results dictate whether or not the idea generation process must be recommenced (e.g., negative evaluation outcome), modifications to the current idea must be implemented (e.g., neutral evaluation outcome pointing towards minor improvements), or if the iterative process can be deemed complete (e.g., positive evaluation outcomes). Figure 4 outlines this iterative process which should be carried out in each step (i.e., 1. Brainstorming; 2. Physical Prototype; 3. Presentation (Optional); 4. Software Prototypes(s); 5. Design Documentation; 6. Production; and 7. Quality Assurance) of Fullerton's playcentric approach.

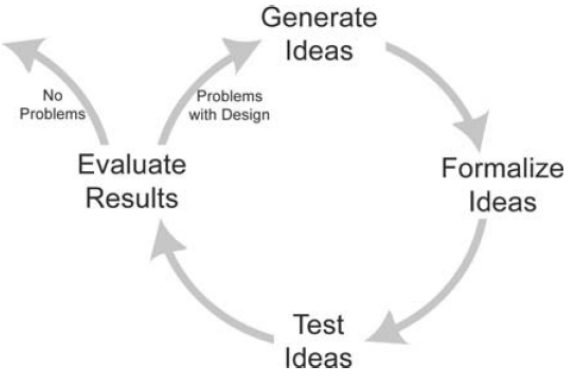


Figure 4: The Playcentric Approach to Game Design iterative process diagram by Fullerton, 2008

1. Brainstorming
During the first step of the playcentric approach to game design, core design concepts are brainstormed iteratively (Fullerton, 2008). First, player experience goals are set and game concepts that have the possibility to achieve such goals are outlined (Fullerton, 2008). A shortlist for the game concepts is made and more detailed descriptions of each concept is put forth (Fullerton, 2008). These written concepts are then tested with potential game players (Fullerton, 2008).

2. Physical Prototype
Step 2 of this approach requires a playable prototype to be crafted so that it can then be play tested (Fullerton, 2008). This process must be iterated until the prototype achieves the set player experience goals (Fullerton, 2008).

3. Presentation (Optional)
While optional, Step 3 involves curating a presentation of the game concept with the intent of securing funding for game actualization (Fullerton, 2008). This step also provides game designers with an opportunity to receive feedback about the game concept so that adjustments for success can be made accordingly (Fullerton, 2008).

4. Software Prototype(s)
During Step 4, multiple rough computer models of the game are generated (Fullerton, 2008). These software prototypes are then playtested until they achieve the set player experience goals (Fullerton, 2008).

5. Design Documentation
Step 5 of this approach involves drafting the design documentation which outlines all game aspects and functionality (Fullerton, 2008).

6. Production
Step 6 of Fullerton's (2008) approach to game design requires team member collaboration to ensure that the game design put forth is not only achievable but that it is also clearly outlined in the design documentation. As a result, sufficient direction exists for team members to initiate artwork creation and game programming (Fullerton, 2008). Throughout these processes, artwork and gameplay are still play tested iteratively (Fullerton, 2008).

7. Quality Assurance
The final step of the playcentric approach to game design, continued playtesting, with an emphasis on usability, is required to gauge the quality of the game (Fullerton, 2008).

2.3.1.4 Winn, 2009
Originally conceptualized by Winn (2009) as the Design, Play, and Experience (DPE) Framework for serious games, this work was later adapted into The Art of Serious Game Design method for participatory game design processes. With this method, facilitated workshops guide interdisciplinary teams of participants towards serious game conceptualization and design with brainstorming and prototyping activities (Winn, 2009). This method highlights the importance of balancing four interconnected and equally important elements: 1. Learning; 2. Storytelling; 3. Gameplay; and 4. User Experience (Winn, 2009). Within each of these three elements (i.e., 1. Learning; 2. Storytelling; 3. Gameplay; and 4. User Experience), symbolic meaning is attributed to designer/player concepts of: 1. Design Story; 2. Play; and 3. Experience. Further explanation of each method element and designer/player concept can be found in Table 4. A visual representation of this SGD design methodology and all of its contributing elements is also depicted in Figure 5. In actually implementing this method in practice, interdisciplinary workshop teams take on steps of: 1. Brainstorming 1; 2. Brainstorming 2; and 3. Paper Prototyping.

Table 4: The Art of Serious Game Design method element explanations adapted from Winn, 2009

Concept	Explanation
Learning	"refers to the content to be learned by players through the game with specific and measurable learning outcomes" (p. 7)
Storytelling	"refers to the background story of the game and includes a description of the character(s), the setting, and the ultimate goal of the game" (p. 7)
Gameplay	"refers to the way in which the player interacts with the game, or with other players (if a multiplayer game). It encapsulates the type of activity (e.g., puzzle, trivia, etc.) found in the game" (p. 7)
User Experience	"refers to the player's emotions and attitudes while playing the game, as well as how the player interacts with the game" (p. 7)
Design	"symbolizes the designer's "story," or all the elements that the designer introduces to the game that will allow the player to play the game" (p. 7)
Play	"symbolizes the "mediated experience" between the player and design input through play, and is influenced by the design and the player's background brought into the play experience" (p. 7)
Experience	"symbolizes the varying play experiences that players can have depending on their backgrounds, as well as the choices and actions made during game play" (p. 7)

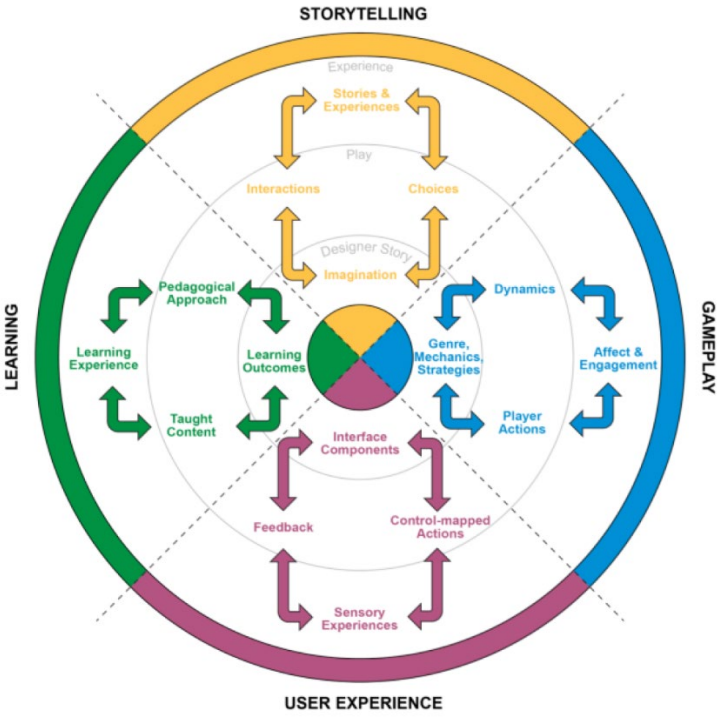


Figure 5: The Art of Serious Game Design methodology circle as adapted from the Design, Play, and Experience (DPE) Framework by Winn, 2009

1. Brainstorming 1

During the first step of Winn's (2009) approach, workshop participants that are deemed as "subject matter experts" are guided through a brainstorming session with the use of ideation cards to determine gameplay elements. During this first session, the goal is to have the participants answer as many ideation card questions as possible, so to jumpstart the serious game concept (Winn, 2009). During this brainstorming process, one team member is tasked with the responsibility of drawing each ideation card from the deck and reading its content aloud to the rest of the team (Winn, 2009). As a team, the participants then discuss their answers to the ideation card as a group until they have agreed upon a final response (Winn, 2009). Final responses are then marked on a colour coordinated sticky note and mapped onto the methodology circle (Figure 5).

2. Brainstorming 2

Similar to step 1 of this method, step 2 involves further game concept brainstorming (Winn, 2009). However, during this session, additional participants join the "subject matter experts" in game ideation (Winn, 2009). During this session, the "subject matter experts" first brief the new participants about the outcome of their Brainstorming 1 session (Winn, 2009). Then, a new set of ideation cards are used

in a similar fashion to that of step 1, and team responses are ultimately mapped once more to the methodology circle (Figure 5). Ultimately, Brainstorming 2 is an iteration of Brainstorming 1, and completion of this step results in a concrete game concept (Winn, 2009).

3. Paper Prototyping

After the concrete game concept has been determined, technological considerations for the actual developed game are first discussed before paper prototyping commences (Winn, 2009). During paper prototyping, workshop participants come together to draw out the game concept, as it would appear in a developed game interface (Winn, 2009). Group debriefing follows this process (Winn, 2009).

2.3.1.5 Abeele et al., 2012

Abeele et al. (2012) developed the Play-Centered Iterative, Interdisciplinary, and Integrated (P-III) Framework for game design to support participatory game design processes (Mildner & Mueller, 2012). In this framework, Abeele et al. (2012) highlight the importance of including both players and domain experts in the serious game design process (Mildner & Mueller, 2016). In the P-III Framework, the design of serious games is built on four pillars related to: 1. Play-Centric Design; 2. Iterative Development; 3. Interdisciplinary Teamwork; and 4. Integration of Play & Learning. While these pillars are intertwined throughout the entirety of the game design process, this framework is broken down into three phases which include 1. Concept Design; 2. Game Design; and 3. Game Development. Within each phase, different steps must be followed. The three phases and associated design steps for the P-III Framework are illustrated below in Figure 6.

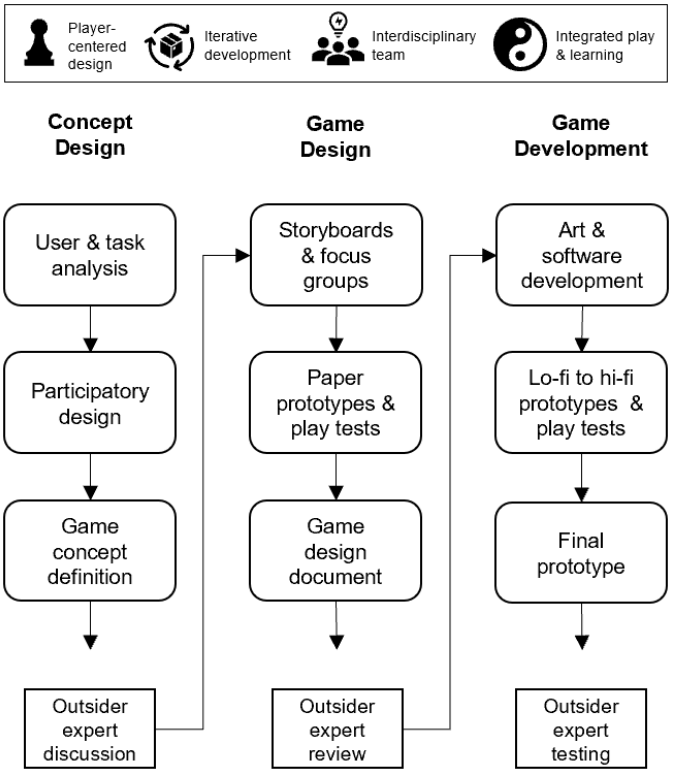


Figure 6: The Play-Centered Iterative, Interdisciplinary, and Integrated (P-III) Framework for game design process diagram adapted from Abeele et al., 2012

1. Play-Centric Design

The first pillar of the P-III Framework for SGD focuses on providing future game players with the creative ability to contribute to game input as designers (Abeele et al., 2012). While several methods can be used to involve the target audience in the design phase of the game, Abeele et al., (2012) highlight the use of participatory design sessions as a means to ensure that game concepts remain relevant to players.

2. Iterative Development

The second pillar of this framework emphasizes the importance of iterative and incremental processes during serious game design and development (Abeele et al., 2012). Such processes are fundamental in tackling phases of 1. Concept Design; 2. Game Design; and 3. Game Development. Throughout phase 1. Concept Design, a first iteration of the game concept is defined based on problem understanding and target audience (Abeele et al., 2012). Then, through a second iteration in phase 2. Game Design, this concept is expanded upon incrementally so that it is transformed into a detailed game (Abeele et al., 2012). Finally, in phase 3. Game Development, the game is developed to completion in an incremental and iterative process (Abeele et al., 2012). As such, iterative and incremental improvement is not only fundamental to the P-III Framework as a whole, but it is also key to the success of each phase within the process.

3. Interdisciplinary Teamwork

Within the P-III Framework, the pillar of Interdisciplinary Teamwork highlights the importance of involving a heterogeneous mix of participants during the serious game design process (Abeele et al., 2012). As such, all team members, including the participating target audience, must participate in all aspects of the serious game design process to encourage mutual learning and ensure an appropriate balance in expertise (Abeele et al., 2012).

4. Integration of Play & Learning

Within the final pillar of the P-III Framework, Abeele et al. (2012) advocate that "play and learning need to be integrated as close as possible" (p. 84). In order to ensure a seamless integration of both game mechanics and core learning principles, game designers must work alongside domain experts (Abeele et al., 2012). In doing so, serious objectives can be better aligned with desired emotional responses that are aimed to be drawn out by different game mechanics (Abeele et al., 2012).

2.4 Conclusions of the Literature Review

In concluding Chapter 2 - Literature Review, the theoretical underpinnings to this research have been explored. Additionally, study background and foundation have been established, and answers to *SQ1*, *SQ2*, and *SQ3* have been drawn. In establishing a foundation to this study and answering *SQ1*, *SQ2*, and *SQ3*, input for the method and design of experimental procedures have been identified.

First, in addressing *SQ1* in section 2.1, bias in this study refers to an "inclination or prejudice for or against one person or group, especially in a way considered to be unfair" on the basis of race ("Bias," 2021). While such bias can be both implicit and explicit, this study focuses on racial-based bias that is explicit in nature where such biased sentiment is neither automatic nor subconscious. Similar to the 2011 work of Iyengar, Messing, & Hahn, explicit racial bias in this study is measured with survey indices related to overt racism and racial resentment. Operationalizing this notion of explicit bias in the context of US policing, police bias is then defined as explicit bias in the context of policing, where purposeful stereotyping is used to understand community dynamics and potential criminal activity. Given the existing limitations identified by Goff et al. (2016) in section 2.1.3 relevant to measuring this type of police bias, this study "measures" police bias qualitatively by observing the content of subsequently designed serious game concepts and their inclusion of purposeful stereotyping to understand community dynamics and potential criminal activity. In doing so, this research recognizes that while unequal treatment of different demographic groups by the police is not always on the basis of explicit racial bias alone, racial resentment can still impact officers' beliefs, attitudes, and subsequent policing tactics and interactions.

In addressing *SQ2*, successful approaches to traditional community member involvement in interventions for improved policing and reduced police bias were addressed in section 2.2. While a review of the literature pointed to COP being more of a philosophy to policing as opposed to a series of prescribed methods, delineated COP approaches from Gaines (1993), Cordner (1999), Rau et al. (2020), and COPS (2012) were analyzed to identify successful approach criteria relevant to "Setting the Stage", "Problem Identification", "Systems Analysis", "Alternative Strategy Formulation", "Continuous Improvement", "Multi-Stakeholder Involvement", "Interdisciplinary Approach", and "Long-Term Orientation". Notably, all reviewed methods involved approach criteria of "Problem Identification", "Continuous Improvement", "Multi-Stakeholder Involvement", and "Interdisciplinary Approach". Therefore, for further analysis, these approach criteria are considered to be most relevant for determining an appropriate SGD method. Only Gaines' (1993) and COPS' (2014) methods included approach criteria of "Setting the Stage". Gaines (1993) was also the only method that included the

approach of "Alternative Strategy Formulation". Only Cordner's (1999) method included the approach criteria of "Long-Term Orientation". Interestingly, while the methods of Cordner (1999), Rau et al., (2020), and COPS (2014) all highlighted "Long-Term Orientation" in their method, Gaines' (1993) method did not. Results from this analysis have been summarized in section 2.2 Table 2. In analyzing the COP methods outlined by Gaines (1993), Cordner (1999), Rau et al. (2020), and COPS (2014) insight into expected results for successful COP implementation was also identified. As such, successful COP can result in reductions in community fear, increased community satisfaction, and increased community faith in law enforcement.

Finally, in addressing *SQ3*, different SGD methods have been analyzed according to their commonalities with effective COP methods in section 2.3. In considering different SGD methods, the work of Duke & Geurts (2004), Kortmann & Harteveld (2009), Fullerton (2008), Winn (2009), and Abeele et al., (2012) have been considered on the basis of "Setting the Stage", "Problem Identification", "Systems Analysis", "Alternative Strategy Formulation", "Continuous Improvement", "Multi-Stakeholder Involvement", "Interdisciplinary Approach", "Storytelling Emphasis", "Design Approach", and "Technology Platform". Results from this analysis have been summarized in section 2.3 Table 3. Given the commonalities present between these analyzed SGD methods and effective COP methods, SGD methods of Winn (2009) and Abeele et al. (2012) have been identified as most suitable for design processes between communities and police officers in the United States. Below, Table 5 supports such decision and provides visual clarity as to how the Winn (2009) and Abeele et al. (2012) SGD methods are most in line with the analyzed successful COP methods. As aforementioned, a comparison of SGD methods and COP methods have only been made on the basis of criteria related to the most prevalent COP method emphasis of "Problem Identification", "Continuous Improvement", "Multi-Stakeholder Involvement", and "Interdisciplinary Approach". While Table 5 illustrates that Duke & Geurts (2004), Kortmann & Harteveld (2009), Winn (2009), and Abeele et al. (2012) all fit criteria related to "Problem Identification", "Continuous Improvement", and "Multi-Stakeholder Involvement", the methods of Winn (2009) and Abeele et al. (2012) also fit criteria relevant to "Interdisciplinary Approach". As such, only SGD methods of Winn (2009) and Abeele et al. (2012) then fulfill all COP method emphasis criteria when compared to the other analyzed SGD methods. Furthermore, both SGD methods of Winn (2009) and Abeele et al. (2012) also have a "Storytelling Emphasis" associated with them. According to a study by Kim & Li (2021), storytelling is not only an engaging means of learning, but it also improves motivation, enhances creativity, assists identity development, and helps facilitate connection amongst peers. Therefore, for this additional reason, SGD methods of Winn (2009) and Abeele et al. (2012) have been identified as the most appropriate SGD methods to serve as potential foundations for participatory design processes. Finally, on the basis of technological platforms, both Winn (2009) and Abeele et al. (2012) SGD methods further demonstrate their goodness of fit with a police-community serious game design process. Seeing as both SGD methods allow for the conceptualization of either analog or digital games, it can be ensured that participants' creativity is not stifled during the police-community design processes.

Table 5: Summary of method emphasis comparison for analyzed COP methods & SGD methods where: Winn (2009) and Abeele et al. (2012) SGD methods share 4/4 commonalities with COP method emphasis, the Duke & Geurts (2004) and Kortmann & Harteveld (2009) SGD methods share 3/4 commonalities with COP method emphasis; and the Fullerton (2008)

Method Emphasis	SGD Method					COP Method			
	Duke & Geurts, 2004	Kortmann & Harteveld, 2009	Fullerton, 2008	Winn, 2009	Abeele et al., 2012	Gaines, 1993	Cordner, 1999	Rau et al., 2020	COPS, 2014
Setting the Stage	X	X	X			X			X
Problem Identification	X	X		X	X	X	X	X	X
Systems Analysis	X	X		X	X		X		
Alternative Strategy Formulation	X	X				X			
Continuous Improvement	X	X	X	X	X	X	X	X	X
Multi-Stakeholder Involvement	X	X	X	X	X	X	X	X	X
Interdisciplinary Approach				X	X	X	X	X	X
Storytelling Emphasis				X	X				
Long-Term Orientation							X	X	X
Design Approach	Linear	Iterative	Iterative	Linear	Iterative				
Technology Platform	Analog & Digital	Analog & Digital	Digital	Analog & Digital	Analog & Digital				

3

METHOD & EXPERIMENTAL DESIGN

This chapter aims to put forth the method and experimental design of this study. In doing so, *SQ4* is addressed. First, the motivation for the chosen mixed methods approach to the research is explained in section 3.1. Then, the quantitative components (section 3.2) and the qualitative components (section 3.3) for this study are outlined. In doing so, data collection methods, data descriptions, data pre-processing procedures, and analysis techniques are put forth. Finally, the experimental procedure for the study and the participatory game design workshop sessions are then described in section 3.4.

3.1 A Mixed Methods Approach

In order to investigate the main research question, a mixed methods approach to the research, involving both quantitative and qualitative approaches, has been chosen. In combining both quantitative and qualitative approaches to research, better depth of understanding and corroboration can be developed (Johnson, Onwuegbuzie, & Turner, 2007 as cited in Pardede, 2018). While mixed methods approaches to research can be integrated in a way that is either quantitatively or qualitatively dominant, this research calls for a balanced emphasis on both quantitative and qualitative methods. Additionally, during part of this research, as data will be collected during SGD workshop activities, analyzed separately, and then combined, a parallel approach to both quantitative and qualitative research activities must be enabled (Pardede, 2018). For this reason, and as classified by Creswell & Plano Clark (2007), a triangulation mixed methods approach with concurrent timing has been chosen for this research. However, while both quantitative and qualitative data will be collected concurrently in SGD workshop activities, qualitative data collection will be prioritized after SGD workshop activities via group semi-structured interviews with workshop participants. As such, qualitative group semi-structured interview data about participants' experience with the serious game design intervention for reduced police bias have the potential to allow further understanding of both the quantitative and qualitative data generated in SGD workshop activities. Therefore, and as also classified by Creswell & Plano Clark (2007), this research also calls for an explanatory mixed methods approach with sequential timing, where additional qualitative research activities follow initial quantitative and qualitative activities. As such, a combination of triangulation mixed methods approach with concurrent timing and explanatory mixed methods approach with sequential timing is used in this study. Table 6 below summarizes each basic mixed methods design as characterized by Creswell & Plano Clark (2007) and highlights the methods employed for this study. (as cited in Borrego, Douglas, & Amelink, 2009). Figure 7 outlines the triangulation mixed methods approach with concurrent timing, the explanatory mixed methods approach with sequential timing, and the combination of both of these mixed methods approaches which has been used in this research.

Table 6: Summary of mixed methods designs as outlined by Borrego et al. (2009) and adapted from Creswell & Plano Clark (2007)

Design Type	Timing of quan and qual phases	Relative weighting of quan and qual components	Mixing - when quan and qual phases are integrated	Notation
Triangulation	Concurrent	Equal	During interpretation or analysis	QUAN + QUAL
Embedded	Concurrent or Sequential	Unequal	One is embedded within the other	QUAN(qual) *or* QUAL(quan)
Explanatory	Sequential, quan then qual	Usually quan is given priority	Phase 1 informs phase 2	QUAN -> qual
Exploratory	Sequential, qual then quan	Usually qual is given priority	Phase 1 informs phase 2	QUAL -> quan

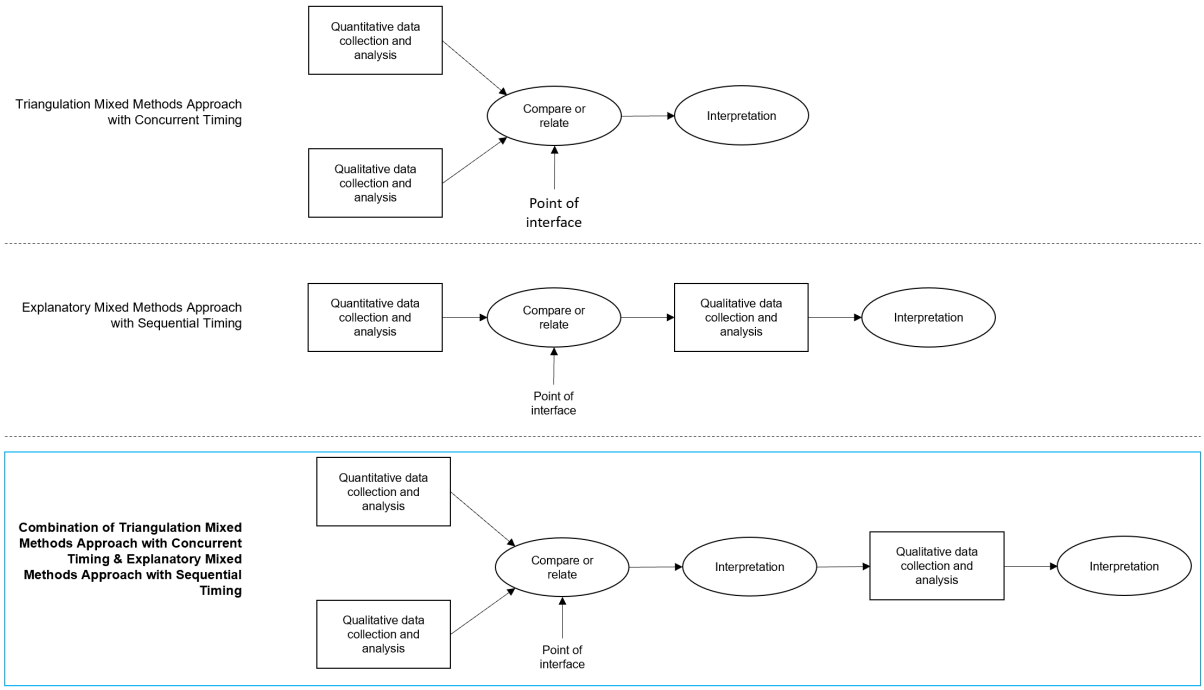


Figure 7: Triangulation mixed methods approach with concurrent timing, explanatory mixed methods approach with sequential timing, and combined triangulation mixed methods approach with concurrent timing and explanatory mixed methods approach with sequential timing diagrams adapted from Creswell & Plano Clark (2011)

3.2 Quantitative Components

3.2.1 Data Collection

3.2.1.1 Knowledge, Attitude, & Practice (KAP) Study

In order to address *SQ4*, quantitative methods, calling for quantitative data collection, can be implemented. As aforementioned in section 1.1.1, biased policing in the US is a public health issue (Helander & McNeill Brown, 2020; Cooper et al., 2004; Feldman, 2015; McGregor, 2016; Obasogie & Newman, 2017). The bias which drives unequal police practices is reflected by both beliefs and attitudes (Cantone, 2021). Beliefs, informed by knowledge acquired via culture, education, and experience, guide attitudes (Seitz & Angel, 2020; Fishbein, 1966). When operationalized, bias, which is informed by both beliefs and attitudes, leads to prejudiced behaviour or practice (Leitner et al., 2016). Thus, given the public health element and interaction of knowledge, attitudes, and practices in the case of biased US policing, a knowledge, attitude, and practice (KAP) study can be implemented to quantitatively approach this research.

As outlined by Muleme et al., (2017, p.318) a KAP study is fundamental to exposing "societal context specific dynamics in public health". Ideally, a KAP study precedes an intervention program, and results from the KAP study serve as: 1) input to the design of an effective intervention program and 2) baseline data for future evaluation of the intervention program (Andrade, 2020). According to ul Haq et al. (2012), a KAP study can be used to represent a population when information on a specific topic relevant to what is known, believed, and done is collected via survey. With a KAP study, knowledge is assessed in order to understand how much information a community holds about the particular topic (ul Haq et al., 2012). Community held "knowledge" which is deviant from fact indicates community beliefs (ul Haq et al., 2012). These community beliefs manifest into quantifiable attitudes via a "complex interaction of beliefs, feelings, and values" (ul Haq et al., 2012, p. 693). With the implementation of hypothetical questions, practice can then be quantified with a KAP study on the basis of highlighting the community's understanding of what types of behavior and action should emerge for improved public health surrounding the specific topic of interest (ul Haq, 2012). In Figure 8 below, such interactions between knowledge, attitude, and practice can be observed in the KAP study conceptual framework.

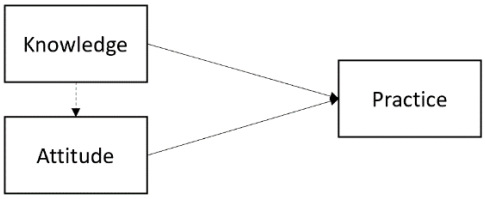


Figure 8: KAP study conceptual framework adapted from Arumugam (2019)

In addressing *SQ4* from a quantitative perspective, a KAP study can be implemented. With this quantitative method for data collection, the knowledge, attitudes, and practices of study participants towards their community, policing, and racism can be measured both before and after SGD workshops via survey. In surveying study participants before and after participation in the SGD intervention, insight into the effectiveness of participatory serious game design for reduced police bias can be measured on the basis of surveyed constructs. Additionally, insight into the effectiveness of both a more "traditional" approach (i.e., involving just community members) and a COP approach (i.e., involving both community members and law enforcement officers/associates) to SGD workshops can be generated with KAP survey data. In collecting KAP related data, participant gender, age, and race can also be gathered. In doing so, it is possible to gain insight into how different demographic variables relate to the knowledge, attitudes, and practices of participants after involvement in this intervention study.

With the use of Andrade et al.'s (2020) five step guideline, a KAP survey has been developed to generate insight into the knowledge, attitudes, and practices of study participants towards their community, policing, and racism. In order to develop the structured questionnaire, a need for the survey was identified, the target population was defined, questions were prepared, and questions were validated. All survey questions were drawn from the works of Iyengar et al. (2011), Kinder & Sanders (1996), and COPS (2018). Studies by Fontaine et al. (2017), Rohe et al. (1997), Buchanan et al. (2021), Edwards, Lee, & Esposito (2018), Pierson et al. (2020), Gaston (2018), and Goutille (2009) also informed question content and formatting. Details of the KAP survey development and validation process for this study has been outlined in Appendix A. The final validated KAP survey questions for this study can also be found in Appendix A. *Qualtrics* software was used to anonymously distribute KAP surveys. Techniques used for the analysis of KAP survey results are outlined in section 3.2.4.

3.2.2 Data Description

Participant Data Collected via Survey

Notably, while 43 participants attended the SGD workshop activities, some participants had issues filling out surveys. For example, five participants failed to appropriately link their online pre-intervention and post-intervention KAP surveys because they forgot to write down their unique participant IDs. One other participant failed to submit a pre-intervention KAP survey because they missed the introduction session, and instead, they only filled out a post-intervention KAP survey. As such, the data described for quantitative components relevant to KAP surveying of this research only represent 37 participants. Similarly, the data described for the quantitative components relevant to the SGD workshop effectiveness evaluation part of this research only represent 38 participants.

Participant Social Desirability Data

Prior to engaging in the initial KAP surveying and subsequent SGD workshop activities, all 43 participants were surveyed for their tendency to act in a socially desirable way. In filling out Sârbescu, Costea, & Rusu's (2012) shortened version of the MC-SDS survey, participants provided information on their self-reported concern for social approval.

Participant Demographic Data

Prior to initiating SGD workshop activities, all participants were surveyed for whether they participated in Workshop A protocol or Workshop B protocol. Additionally, they were also surveyed for their demographic attributes relevant to variables of "gender", "age", and "race". In total, 16 workshop participants identified as male, 20 identified as female, and 1 identified as "other" or a gender identity

that was not listed on the survey. With regard to age, 11 participants were 18-24 years old, 22 were 25-34 years old, 3 were 35-44 years old, and 1 was 55-64 years old. Additionally, 19 participants identified as "White or Caucasian", and 18 participants identified is either being Black, Indigenous, People of Color (BIPOC).

A breakdown of participant demographics for each workshop protocol (Workshop A and Workshop B) for each of the listed variables are presented below in Figure 9.

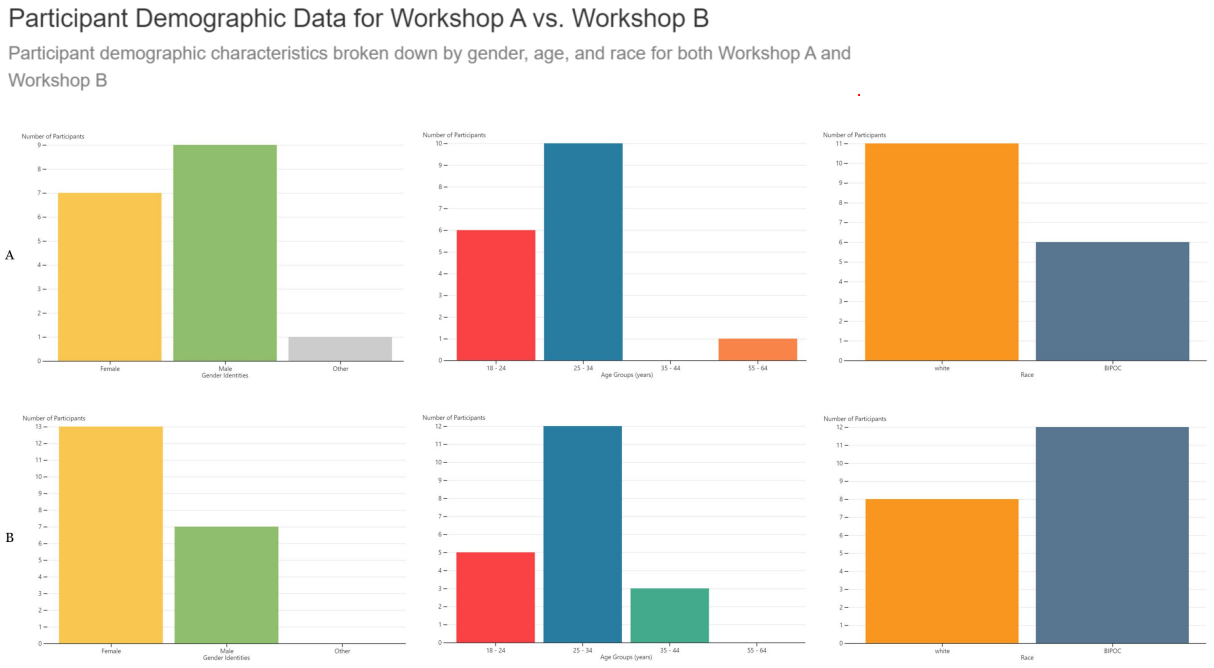


Figure 9: Breakdown of 17 (21 actual participants) Workshop A participant demographics for participant gender data, age data, and race data vs. all 20 (22 actual participants) Workshop B participant demographics for participant gender data, age data, and race data

Participant Knowledge, Attitude, and Practice (KAP) Data

All participants were surveyed before and after SGD workshop activities about their knowledge, attitudes, and practices towards community, policing, and racism. In doing so, data was collected on participants' responses to questions relevant to constructs of "Knowledge about Policing in Communities", "Attitude about Policing in Communities", "Attitude about Explicit Racism", and "Practice about Policing in Communities".

Participant Perceived Effectiveness of Participatory SGD Workshops Evaluation Data

Following the SGD workshop activities, all participants were questioned about their perceived effectiveness of the participatory SGD workshops via an evaluation survey. In filling out this survey, participants provided feedback via question and answer on constructs relevant to the "structure & flow", "usability", and "team communication" of the workshop. In total, this information provided insight into participants' perceived overall effectiveness of the participatory SGD workshop.

3.2.3 Data Pre-Processing

Upon collecting the data described in section 3.2.2, data pre-processing was conducted in order to render the data in a more usable format for subsequent analysis. As such, surveyed data was first downloaded from the *Qualtrics* online survey platform in .csv format for later use in the *Python* programing language. For data pre-processing in the *Python* programming language, the *Pandas* and *Numpy* libraries were primarily used.

Participant Social Desirability Data

With the *Python* programing language, the social desirability data was first cleaned for easy pre-processing. Then, the data was recoded to reflect a point system based on the answer key presented in Table F1 of Appendix E. For questions where participants provided answers that indicated a high social desirability, 1 was allotted. For questions where participants provided answers that indicated a low social desirability, 0 was allotted. Then, a "social desirability score" was created by adding up each participants' total score on the recoded MC-SDS survey, dividing that value by the number of questions present in the survey (13), and subtracting the resulting value from 1. The resulting "social desirability scores" were represented as values ranged between 0 and 1, where a value close to 0 represented high social desirability (i.e., more concern for social approval and less "honest" question answering) and a value close to 1 represented low social desirability (i.e., less concern for social approval and more "honest" question answering). As a result of this pre-processing step, each unique participant had an associated "social desirability score" which served as an indication as to how much they may have been seeking social approval in responding to other survey questions put forth in the workshop.

Participant Demographic Data

With the *Python* programing language, basic cleaning of the demographic data was conducted in preparation for subsequent pre-processing where both demographic data and other data relevant to social desirability, KAP, and workshop evaluation were involved.

Participant Knowledge, Attitude, and Practice (KAP) Data

The *Python* programming language was then used to clean data relevant to both KAP #1 and KAP #2 surveys. Then, the KAP data for each respective survey was recoded to reflect a point system. For questions pertaining to constructs of "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities", a 5-point Likert Scale was used where 1 was recoded for a response of "strongly disagree", 2 was recoded for a response of "disagree", 3 was recoded for a response of "neutral", 4 was recoded for a response of "agree", and 5 was recoded for a response of "strongly agree". For survey questions relevant to the construct of "Knowledge about Policing", the data was recoded to reflect a point system based on the answer key presented in Table A9 of Appendix A. For knowledge-based questions where participants provided the correct answer, 1 was allotted. For knowledge-based questions where participants provided an incorrect answer, 0 was allotted. For both KAP #1 (i.e., before the SGD workshop intervention) and KAP #2 (i.e., after the SGD workshop intervention), the mean response value of the items for each construct was calculated by adding up the values for responses to questions within a single construct and then dividing that value by the number of questions in each construct. The resulting construct values for "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" were represented as values between 1 and 5, where a value close to 1 represented very negative sentiment associated with the construct at hand and a value close to 5 represented very positive sentiment associated with the construct at hand. The resulting construct value for "Knowledge about Policing in Communities" was represented as a value between 0 and 1 where a value close to 0 represented lower knowledge about policing in communities and a value close to 1 represented higher knowledge about policing in communities. As a result of this pre-processing step, each unique participant had an associated score relevant to measured constructs of "Attitude about Explicit Racism", "Attitude about Policing in Communities", "Practice about Policing in Communities", and "Knowledge about Policing in Communities" for KAP #1 and KAP #2. The difference between each construct mean value for KAP #2 and KAP #1 was then calculated. These construct scores and their associated differences then served as an indication as to how effective the participatory SGD workshop intervention was relevant to these measurements.

Participant Effectiveness of Participatory SGD Workshops Evaluation Data

Using the *Python* programing language, the evaluation data of the effectiveness of the participatory SGD workshops was first cleaned for easy pre-processing. For questions pertaining to the constructs of "structure & flow", "usability", and "team communication" a 5-point Likert Scale was used where 1 was recoded for a response of "strongly disagree", 2 was recoded for a response of "disagree", 3 was recoded for a response of "neutral", 4 was recoded for a response of "agree", and 5 was recoded for a response of "strongly agree". The mean response value of the items for each construct was calculated by adding up the values for responses to questions within a single construct and then dividing that value by the number of questions in each construct. Similarly, an "overall effectiveness" construct pertaining to the overall effectiveness of the participatory SGD workshop based on overall structure & flow, usability, and team communication was created by adding up the values for responses to all questions (11) and then

dividing that value by the number of total questions (11) pertaining to the evaluation of the effectiveness of the participatory SGD workshops. The resulting construct values were represented as values between 1 and 5, where a value close to 1 represented very negative sentiment associated with the construct at hand and a value close to 5 represented very positive sentiment associated with the construct at hand. As a result of this pre-processing step, each unique participant had an associated score relevant to the measured constructs of "structure & flow", "usability", "team communication", and "overall effectiveness" in terms of how they evaluated the participatory SGD workshops. These construct scores then served as an indication as to how participants from different demographic backgrounds perceived the effectiveness of the participatory SGD workshop as an intervention. Notably, the calculated "social desirability score" for each participant was applied to data before subsequent analysis ensued.

Combining the Data

In the final step of data pre-processing, all cleaned and prepared data relevant to participant social desirability, participant demographics, participant KAP responses, and participant evaluation of the effectiveness of participatory SGD workshops were merged together into a single dataframe using the *Python* programing language. This data was easily paired on the basis of the unique "Random ID" in which was initially assigned to each participant anonymously and then later required for subsequent surveying. As a result of this final pre-processing step, the following data elements were present for each unique participant (indicated by their unique "Random ID"):

- An associated "social desirability score" which served as an indication as to how much they may have been seeking social approval in responding to other survey questions put forth in the workshop
- Demographic data outlining variables relevant to "gender", "age", and "race"
- Data outlining participation in Workshop A protocol or Workshop B protocol
- KAP #1 scores for measured constructs of "Attitude about Explicit Racism", "Attitude about Policing in Communities", "Practice about Policing in Communities", and "Knowledge about Policing in Communities"
- KAP #2 scores for measured constructs of "Attitude about Explicit Racism", "Attitude about Policing in Communities", "Practice about Policing in Communities", and "Knowledge about Policing in Communities"
- The difference between each construct mean value for KAP #2 and KAP #1
- An associated score relevant to the measured constructs of "structure & flow", "usability", "team communication", and "overall effectiveness" in terms of evaluating the participatory SGD workshops

3.2.4 Analysis

3.2.4.1 Within-Subjects: Paired Sampled T-Test

As described by Ross & Willson (2017, p. 17) "a paired samples t-test compares the mean of two matched groups of people or cases, or compares the mean of a single group, examined at two different points in time". With this study design, repeated measures can be taken from each participant as they participate in a treatment condition so that the changes within the same participants can be studied over time (Bhandari, 2021). Appendix F outlines the specifics of this analysis technique.

In this study, the quantitative data pertaining to constructs of "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" was analyzed from a within-subjects perspective with a paired samples t-test in order to gauge the effectiveness of the intervention for each individual participant. With the paired samples t-tests, the means of each construct measured in KAP #1 (pre SGD workshop intervention) and KAP #2 (post SGD workshop intervention) were compared for the same participant on the basis of their unique "Random ID". As such, multiple paired sample t-tests were conducted with the "Random ID" serving as the independent variable and each of the KAP related construct scores pertaining to "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" serving as the dependent variable for subsequent testing. The *Python* programming language was used in accordance with the *SciPy.stats* module for all analysis. In conducting this analysis, it was determined whether or not statistical evidence existed to support the alternative hypothesis that the mean difference between the paired construct observations from KAP #1 and KAP #2 were significantly different from zero. Notably, the calculated "social desirability score" for each participant was applied to both KAP #1 data and KAP #2 before subsequent t-testing. An alpha level (α) of 0.05 was used for all tests.

3.2.4.2 Between-Subjects: Analysis of Covariance (ANCOVA)

As described by Philippas (2014, p.157) the analysis of covariance (ANCOVA) is a technique that "analyzes grouped data having a response (dependent variable) and two or more predictor variables (covariates) where at least one of them in continuous (i.e., quantitative, scaled) and one of them is categorical (i.e., nominal, non-scaled)". With ANCOVA, it is possible to "model the response of a variable as a linear function of predictor(s), with the coefficients of the line varying among different groups" (Philippas, 2014, pp. 157-158). In ANCOVA, the covariates are included as statistical controls, so to explain the dependent variable variation, reduce error variation, and increase statistical power (Philippas, 2014). Ultimately, this analysis can be used with between-subjects classification designs, where participants are sampled from mutually-exclusive populations (e.g., male, female), in order to generate an understanding of how the sampled responses (i.e., dependent variable) could be affected by participants' different characteristics (i.e., covariates) (Schneider, Avivi-Reich, & Mozuraitis, 2015).

In this study, the quantitative data pertaining to constructs of "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" was analyzed from a between-subjects perspective with the analysis of covariance (ANCOVA) in order to gauge how participants' different workshop protocols (i.e., Workshop A and Workshop B) and different demographic characteristics relevant to gender, age, and race could have impacted participants' experiences with the SGD workshop intervention. As such, the effectiveness of the intervention for different groups of participants (on the basis of workshop protocol, gender, age, and race was investigated. With ANCOVA, the means of each construct measured in KAP #1 (pre SGD workshop intervention) were used for the covariate variable. The means of each construct measured in KAP #2 (post SGD workshop intervention) were used for the dependent response variable. Multiple ANCOVA tests were conducted where workshop protocol, gender, age, and race each served as the independent factor variable. The *Python* programming language was used in accordance with the *SciPy.stats* module for all analysis. In conducting this analysis, it was determined whether or not statistical evidence existed to support the alternative hypothesis that a significant relationship exists between participants' different demographic characteristics and how they experienced the SGD workshop intervention as a result of their KAP #2 mean construct scores (post SGD workshop intervention). Notably, the calculated "social desirability score" for each participant was applied to both KAP #1 data and KAP #2 before subsequent t-testing. An alpha level (α) of 0.05 was used for all tests.

3.3 Qualitative Components

3.3.1 Data Collection

3.3.1.1 Workshop & Focus Group Observation

Qualitative methods related to workshop and focus group observation can be used to address *SQ4*. As a research methodology, workshops can be used to prescribe group activities that challenge participants' ideas (Ørngreen, & Levinsen, 2017). With such workshops, a development dimension is present, as participants' collaborate to create tangible outcomes within a domain-specific context (Ørngreen, & Levinsen, 2017). Focus groups can be used in combination with workshops to collect process feedback and generate group discussion about workshop activities and outcomes (Onwuegbuzie et al., 2009). With focus groups, group context can be captured as participants' are free to express their views and interact with the views of others (Gibbs, 1997). In observing both of these experimental design elements, participant experience can be better understood. Notably, comparison in observations from the beginning to the end of the experimental procedure can lend insight into the effectiveness of the police-community serious game design process and how sentiments related to police-community relations have developed throughout the duration of the process. In observing workshop sessions and focus group discussion in this study, only field notes have been made. While audio and video recording are traditionally effective methods for observing such sessions, such techniques have been omitted from this study due to the sensitive nature of racism and policing. In omitting session recordings, participants have been encouraged to express their true sentiments, and socially desirable behavior has been discouraged. Workshop and focus group observations for this study were recorded using the Observation Protocol outlined in Appendix B. Techniques used for the analysis of workshop and focus group observation data are outlined in section 5.1.2.

3.3.1.2 Group Semi-Structured Interviews

Group semi-structured interviews can also be used to collect qualitative data about the effectiveness of a police-community serious game design process for reduced police bias. With a semi-structured interview form, structured questions informed by theory are paired with an opportunity for respondents to create their own narratives (Galletta, 2013). With such approach, focus and flexibility are balanced, as all dimensions of the research can be adequately addressed while still providing participants with an opportunity to expand upon their responses and create new meaning (Galletta, 2013). According to Treece & Treece (1986), semi-structured interviews are also advantageous for they provide opportunity to adapt to respondents' vocabularies, as the interviewer is able change words to questions without necessarily changing the meaning of such questions (as cited in Louise Barriball & White, 1994). For this study, where potential interview respondents are made up of both law enforcement officers and community members from a range of different backgrounds, such adaptability is particularly useful.

According to Carruthers (1990), interview data also has the potential to significantly enhance survey data. In collecting and analysing data from both surveys and interviews, triangulation can "be used not only to examine the same phenomenon from multiple perspectives but also to enrich our understanding by allowing for new or deeper dimensions to emerge" (Jick, 1979, pp. 603-604 as cited in Carruthers, 1990). As such, group semi-structured interviews about participants' experience with the serious game design intervention for reduced police bias have the potential to allow further understanding of KAP study results regarding community, policing, and racism. Triangulating interview data with focus group observations also has the potential to further explain workshop observed phenomena, providing an even deeper understanding of the success of the serious game design intervention.

In developing group semi-structured interview questions for this study, sections 2.1 and 2.2 of this report were used in combination with a literature review of concepts related to community, policing, and racism. In addition, interview questions were drawn and adapted from Gomez's (2016) study on policing, community fragmentation, and public health; COPS (2018) community survey on public safety and law enforcement; and Winn's (2009) game development process evaluation survey. Interview questions for this study were then validated via expert review from a single law enforcement and community expert from *CPE* for face validation. A list of questions used for group semi-structured interviews for this study can be found in Appendix C. Notably, given the sensitive nature of racism and policing, only written notes about the question response content were made in recording group semi-structured interviews. Techniques used for the analysis of group semi-structured interview data are outlined in section 3.3.3.3.

3.3.2 Data Description

Conceptboard Group Brainstorming Data

During both Brainstorming 1 and Brainstorming 2 SGD workshop activities, all workshop participants used ideation cards as a group to jumpstart their serious game concept, and in the process, sticky notes outlining group ideas were mapped on the *Art of Serious Game Design* Methodology Circle according to brainstorming categories of Learning, Storytelling, Gameplay, and User Experience. In doing so, each workshop session provided their categorized brainstorming data.

Observation Data: SGD Activities

During all SGD activities of all workshop sessions, observers from *CPE* sat in on workshops and took detailed notes according to the Observation Protocol outlined in Appendix B. In doing so, data on overall conversation, agreement/disagreement, and response patterns was collected.

Observation Data: Group Semi-Structured Interviews

During all evaluative group semi-structured interviews of all workshop sessions, observers from *CPE* sat in on workshops and took detailed notes according to the Observation Protocol outlined in Appendix B. In doing so, data on specific responses to each question, response patterns, agreement/disagreement, speech speed, speech tone, and body language was collected from participants.

3.3.3 Data Pre-Processing

Upon collecting the *Conceptboard* group brainstorming data and the observation data relevant to SGD activities described in section 3.3.2, some data pre-processing was conducted in order to render the data in a more usable format for subsequent analysis. Notably, observation data in the form of written notes

about overall conversation during SGD activities was not pre-processed, while data collected via brainstorming sticky notes was processed. First, the data collected via brainstorming sticky notes was transferred (via hand) from the *Conceptboard* workspace to a csv file for subsequent pre-processing with the *Python* programing language. For data pre-processing in the *Python* programming language, the *Pandas* and *Numpy* libraries were primarily used. With the *Python* programing language, basic cleaning of the brainstorming data was performed for subsequent processing. Then, the text for each sticky note was cleaned with the *clean-text Python* library in order to normalize the text representation by lowercasing text, removing additional whitespace, removing special characters, replacing currency symbols with written currency, and replacing numbers with written text (PyPI, 2021). Further text cleaning was performed via tokenization, and the text associated with each cleaned sticky note was broken down further into cleaned tokens of words according to Weng's (2019) protocol. As a result of this pre-processing step, each word within each unique sticky note of text data was cleaned and ready for subsequent analysis.

In addition, data pre-processing was done to the observation data relevant to the group semi-structured interview data. Observation data recorded during evaluative group semi-structured interviews relevant to specific responses to each question, response patterns, agreement/disagreement, speech speed, speech tone, and body language was pre-processed. As such, this data was transferred from the Observation Table in the Observation Protocol sheet to an xlsx file where it was later pre-processed in *Microsoft Excel*. With *Microsoft Excel*, the Observation Tables for all Workshop A and Workshop B sessions were joined and reformatted for easier processing. The number of total participants per workshop protocol was calculated. The number of participants for each demographic group was also calculated per workshop protocol. Then, the total responses made per each question were calculated, and the number of participants per each demographic group who responded to each question was also noted. As a result of this pre-processing step, all Observation Table notes on response content, response characteristics, response patterns, respondent demographics, and respondent non-verbal communication for both Workshop A and Workshop B protocols was organized and reformatted for subsequent analysis.

3.3.3 Analysis

3.3.3.1 Content Analysis: Word Frequency Distribution

As described by Krippendorff (2018, p. 24), content analysis is "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the context of their use". As such, content analysis can be used to understand the content of the text analyzed. With a summative approach to content analysis, words can be not only counted and compared, but also interpreted with underlying context (Hsieh & Shannon, 2005). In taking a summative approach to content analysis, the frequency distribution of words can be analyzed and interpretation can be made. (Dicle & Dicle, 2018).

In this study, the *Conceptboard* group brainstorming data collected via categorized sticky notes was analyzed with a summative approach to content analysis where the frequency distribution of words used during brainstorming sessions was calculated and then interpreted. Words used in all brainstorming categories for all total workshop sessions, Workshop A sessions, and Workshop B sessions were counted, and the most frequently used words were identified. Similarly, the most frequently used words for all total workshop sessions, Workshop A sessions, and Workshop B sessions during each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience were also identified. Then, using the observation data relevant to SGD activities, an interpretation of these results was put forth. The *Python* programming language was used with the *Pandas* library for all analysis. In conducting this analysis, the most frequently used words during brainstorming activities were identified and an interpretation of their use was outlined.

3.3.3.2 Sentiment Analysis

According to the Oxford Dictionary and as outlined by Mishev et al. (2020, p. 131662) "sentiment analysis is defined as the process of computationally identifying and categorizing opinions expressed in a text, primarily to determine whether the writer's attitude towards a particular topic or product is positive, negative, or neutral (Sentiment Analysis, 2021).". As such, this analysis technique can be used to understand the sentiment of the text analyzed.

In this study, *the Conceptboard* group brainstorming data collected via categorized sticky notes was analyzed with sentiment analysis using the *Valence Aware Dictionary for Sentiment Reasoning (VADER)* model from the *Natural Language Toolkit (NLTK*) library in the *Python* programing

language. The positive, negative, neutral, and normalized compound sentiment scores were calculated for all brainstorming categories for all total workshop sessions, Workshop A sessions, and Workshop B sessions. Then, the same sentiment metrics were calculated for all total workshop sessions, Workshop A sessions, and Workshop B sessions during each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience. While results were generated for positive, negative, neutral, and normalized compound sentiment scores, as outlined by Hutto & Gilbert (2014), the compound sentiment score was used as the primary indicator of how positive, neutral, or negative a sticky note was. In using the compound sentiment score and following the suggestions of Hutto & Gilbert (2014), a threshold value for positive sentiment was indicated as a compound score >= 0.05, neutral sentiment was indicated as a compound score > -0.05 and a compound score < 0.05, and negative sentiment was indicated as a compound score < -0.05. In conducting this analysis, the overall sentiment during brainstorming activities was calculated, and the sentiment during each different category of brainstorming relevant to Learning, Storytelling, Gameplay, and User Experience was also measured for all workshop sessions, Workshop A sessions, and Workshop B session

3.3.3.3 Micro-Interlocutor Analysis

As described by Onwuegbuzie et al. (2009, p. 3), micro-interlocutor analysis "incorporates and analyzes information from group interviews by delineating which participants respond to each question, the order of responses, and the nature of response as well as the nonverbal communication used by each of the focus group participants.". With this analysis technique, the group members can be used as the unit of analysis instead of the group itself (Onwuegbuzie et al., 2009). As such, insight can be generated into how different types of group members respond to the posed questions.

In this study, micro-interlocutor analysis was used on the Observable Table data collected during the group semi-structured interviews. For both Workshop A and Workshop B protocols, analysis was done at the aggregate level, accounting for all group semi-structured interview questions, and at the individual question level for interview questions Q1, Q2, Q3, and Q4. In analyzing this data, information on the demographics of respondents, the responses made, the order in which participants responded, the characteristics of response, and the non-verbal communication of participants was considered. At an aggregated level, the overall response rate for the group semi-structured interviews was calculated for each workshop protocol. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for each workshop protocol. Such response rate calculations were also made at a disaggregated level for interview questions Q1, Q2, Q3, and Q4 data. In conducting this analysis, insight into the response rates per different types of group members was generated, and a deeper and more coherent understanding of the response patterns, response characteristics, and response content that was generated for both Workshop A and Workshop B protocols was developed.

3.4 Experimental Procedure

The experimental procedure of this study involves the integration of serious game design workshops, focus group discussions, KAP surveying, and group semi-structured interviews. In addition to these elements, a survey identifying participants' concern for social approval has been included in this study to account for issues with social desirability. As identified in section 2.4, both SGD methods of Winn (2009) and Abeele et al. (2012) may be most appropriate to serve as the foundations for the police-community design workshops. However, due to time limitations of this study, the readily available resources associated with the Winn (2009) method, and the lack of readily available resources for the Abeele et al. (2012) method, Winn's (2009) SGD method has been chosen as the sole foundation for the participatory serious game design process.

While Workshop A and Workshop B protocols both have used the SGD method of Winn (2009) as the workshop foundations, these workshops have been composed with different types of people. Workshop A has been organized with just community members. Conversely, Workshop B has been organized with both community members and law enforcement officers/associates. Regardless, the overall session structures, involving focus group discussions, social desirability surveying, KAP surveying, and group semi-structured interviews, are the same for both workshop protocols. A session description of the serious game design workshop and associated focus group are outlined in Appendix D.

In conducting this research, participatory online workshop sessions were organized with volunteered community members and law enforcement officers/associates. Workshops were organized over two sessions, where the first session was 30 minutes in duration and the second workshop was 90 minutes

in duration. Without prior knowledge about workshop specific activities, volunteers indicated their dates of availability for participating in workshop sessions, and they were placed into either the Workshop A or Workshop B protocol. In order to ensure that all participants were comfortable engaging in the workshop settings and to further limit issues with social desirability, a list of workshop-specific participants was circulated, and participants had an opportunity to indicate whether or not they were uncomfortable engaging with any other members of their session (i.e., due to power dynamics, poor relations, etc.). Participants who had conflicts with the list of workshop participants were then promptly reallocated to a different workshop.

Before engaging in serious game design activities, all participants first took Sârbescu, et al.'s (2012) shortened version of the Marlowe-Crowne Social Desirability Scale (MC-SDS) to indicate their level of concern with social approval (Appendix E). They also took a KAP survey to generate insight into the knowledge, attitudes, and practices that they had towards their community, policing, and racism. Following KAP surveying, participants were briefly introduced to concepts of COP and serious game design. Participants then engaged in the predetermined SGD activities in a *Conceptboard* space similar to that illustrated in Figures 10, 11, and 12.

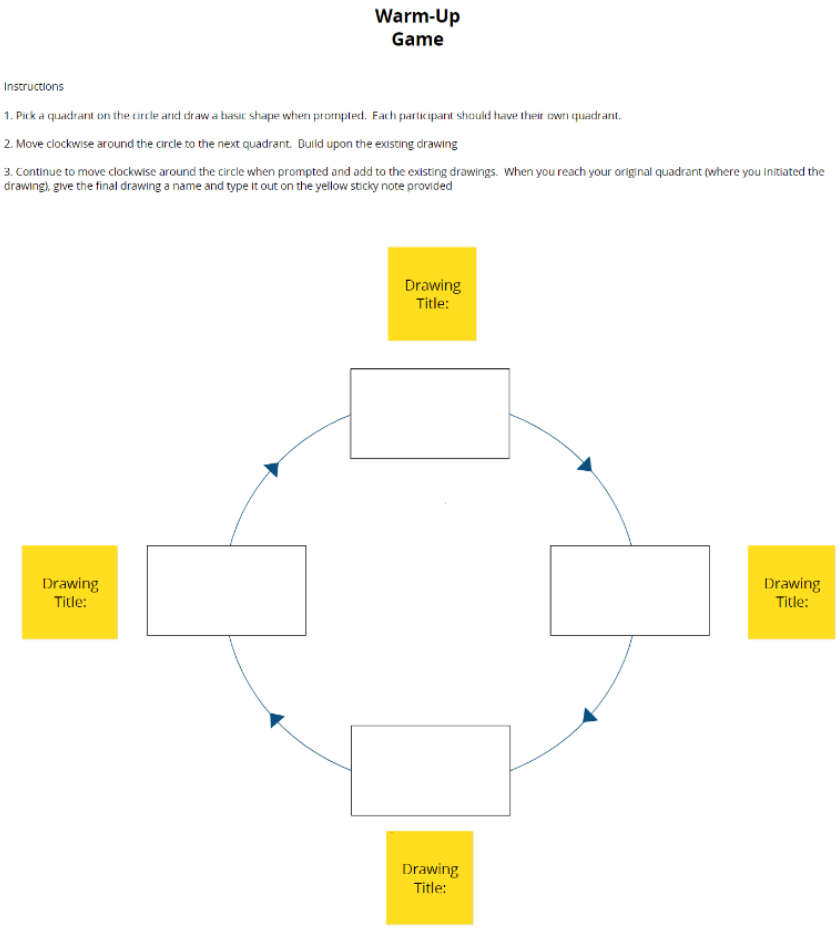


Figure 10: Warm-Up Game Section on *Conceptboard* space for workshops

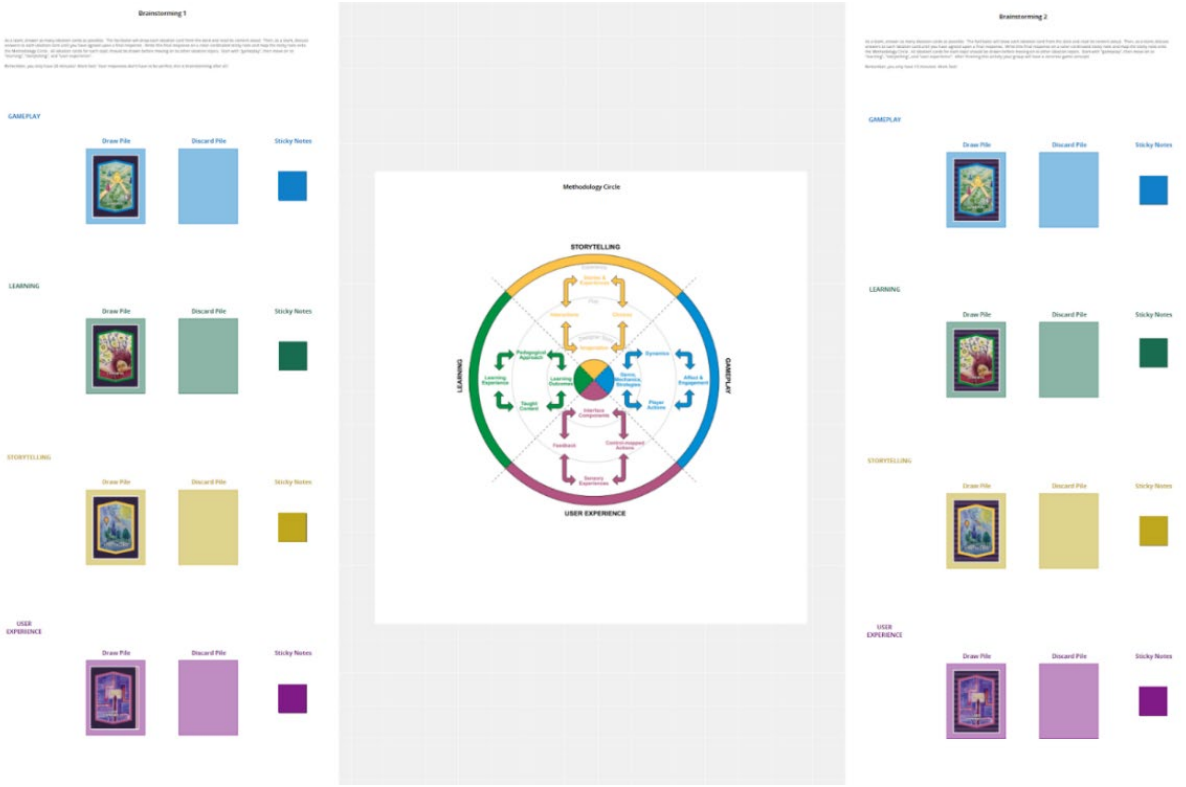


Figure 11: Brainstorming 1, Methodology Circle, and Brainstorming 2 Sections on *Conceptboard* space for workshops

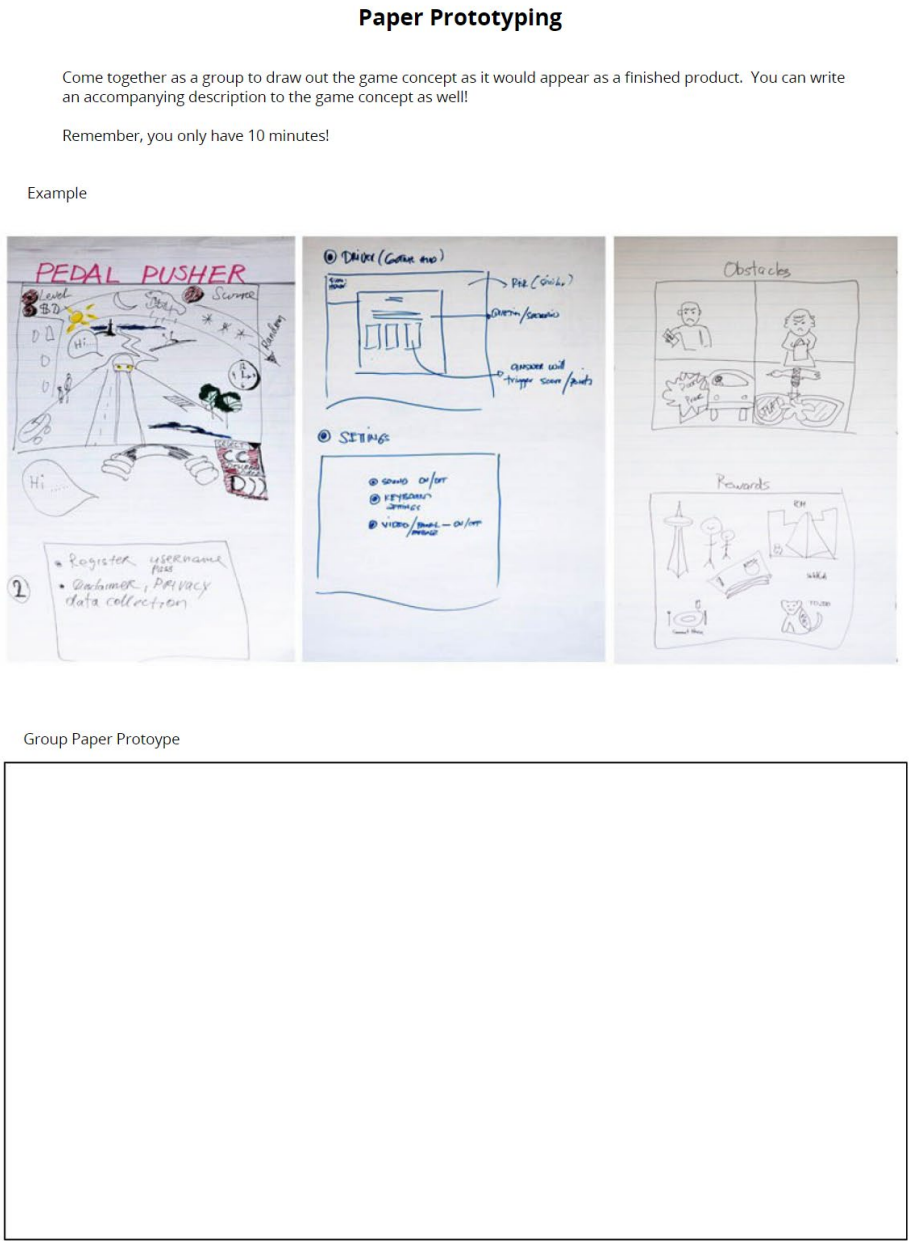


Figure 12: Paper Prototyping Section on *Conceptboard* space for workshops

During the SGD activities, focus group discussion about workshop activities and outcomes was generated. A second distribution of KAP surveying followed. Then, evaluative group semi-structured interviews were conducted with participants. During the SGD workshop activities, focus group discussion, and group semi-structured interviews observations were made in order to gauge how the serious game design process influenced participants' exploration of topics related to community, policing, and racial bias.

Following the design oriented research activities, quantitative and qualitative analysis was conducted in order to address *SQ5* and *SQ6*. Quantitative analysis provided insight into the changes in knowledge, attitudes, and practices of participants towards their community, policing, and racial bias as a result of involvement in the police-community serious game design process. Quantitative analysis also provided insight into the participants' perceived overall effectiveness of the participatory SGD workshop. Qualitative analysis methods based on workshop observation and group semi-structured interviews were used to investigate participants' experience with the serious game design intervention for reduced police bias. Finally, both qualitative and quantitative outcomes were combined in order to develop

further insight, address the Main Research Question, and draw conclusions about the effectiveness of serious game design as an intervention to policing and beyond.

In considering quantitative methods for measuring the effectiveness of the police-community serious game design process for reduced police bias, measured KAP survey constructs related "Knowledge about Policing in Communities", "Attitude about Policing in Communities", "Attitude about Explicit Racism", and "Practice about Policing in Communities" served as metrics for the effectiveness of the intervention. Furthermore, participants' self-evaluated perceived overall effectiveness ratings of the participatory SGD workshop also provided a quantitative metric as to how effective the intervention was. In considering qualitative methods for measuring the intervention's effectiveness, observed engagement and optimism amongst participants in workshop sessions served as a means for operationalizing "effectiveness". In addition, the meaningfulness in workshop session dialogue and the designed serious game concepts also served as a metric for intervention effectiveness.

Below, Figure 10 outlines an overview of the experimental procedure, and research activities are mapped in the sequential order in which they were carried out.

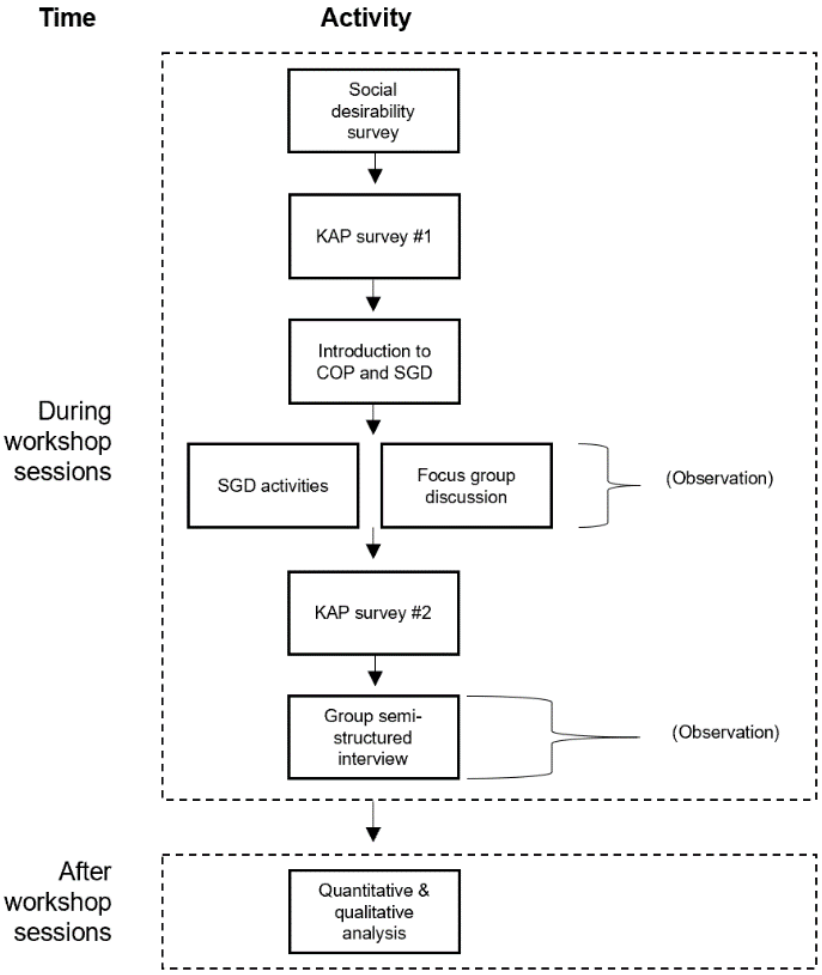


Figure 13: Visual overview of the experimental procedure and sequentially mapped research activities

4

QUANTITATIVE RESULTS

This chapter outlines the results obtained from the quantitative component of this study. Here, the results for within-subject analysis (section 4.1) and between-subject analysis (section 4.2) are detailed in order to quantitively outline the actual effectiveness of the SGD workshop intervention. Then, the perceived effectiveness of the SGD workshop intervention is explored quantitatively in section 4.3 with an analysis of the participant SGD workshop evaluation data. In analyzing the data and obtaining results, *SQ5* is partially addressed from a quantitative perspective, and insight into reducing police bias and improving police-community relations through participatory serious game design is generated.

4.1 Within-Subjects: Paired Samples T-Test Results

As described in section 3.2.4.1, quantitative data pertaining to constructs of "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" was analyzed from a within-subjects perspective with a paired samples t-test in order to gauge the effectiveness of the intervention for each individual participant. With the paired samples t-tests, the means of each construct measured in KAP #1 (pre SGD workshop intervention) and KAP #2 (post SGD workshop intervention) were compared for the same participant on the basis of their unique "Random ID". As such, multiple paired sample t-tests were conducted with the "Random ID" serving as the independent variable and each of the KAP related construct scores pertaining to "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" serving as the dependent variable for subsequent testing. Notably, the calculated "social desirability score" for each participant was applied to both KAP #1 data and KAP #2 before subsequent t-testing. An alpha level (α) of 0.05 was used for all tests. Before t-testing, all datasets were first tested for normality with a Shapiro-Wilk test, and all datasets passed for normality.

4.1.1 Knowledge about Policing in Communities

The box plots represented in Figure 14 illustrate the differences in pre-intervention mean "Knowledge about Policing in Communities" (i.e., "pre mean k_pc") and post-intervention mean "Knowledge about Policing in Communities" (i.e., "post mean k_pc"). As demonstrated below, both pre and post intervention mean "Knowledge about Policing in Communities" construct scores have similar spreads that appear to be symmetric and quite concentrated. In addition, the scores for this construct share the same median value for both pre and post intervention. Outliers are also present for both pre and post intervention construct scores.

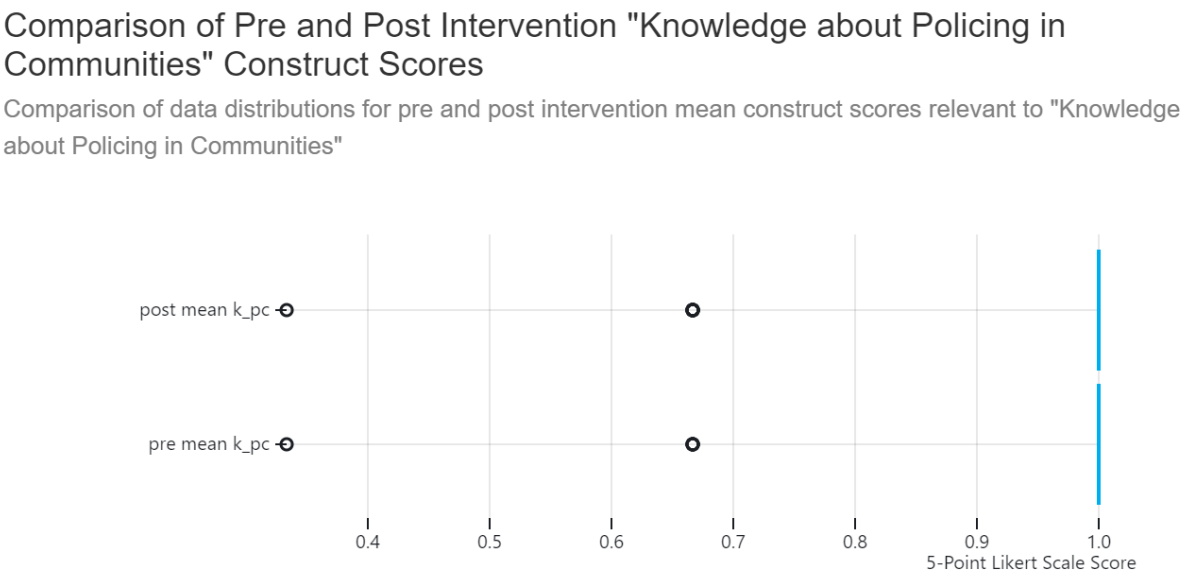


Figure 14: Box plot comparison of pre and post intervention "Knowledge about Policing in Communities" construct scores where "pre mean k_pc" represents pre-intervention mean "Knowledge about Policing in Communities" and "post mean k_pc" represents post-intervention mean "Knowledge about Policing in Communities" measured on a 5-point Likert Scale

Furthermore, the means of the "Knowledge about Policing in Communities" construct measured in KAP #1 and KAP #2 were compared for the same participant on the basis of their unique "Random ID". The results from the KAP #1 (M = 0.937, SD = 0.154) survey and KAP #2 (M = 0.910, SD = 0.169) survey indicate that there was not a significant difference in knowledge about policing in communities after participating in the SGD workshop intervention, $t(36)$ =1.14, p = .26.

4.1.2 Attitude about Explicit Racism

Figure 15 illustrate the differences in pre-intervention mean "Attitude about Explicit Racism" (i.e., "pre mean a_er") and post-intervention mean "Attitude about Explicit Racism" (i.e., "post mean a_er"). As outlined below, both pre and post intervention mean "Attitude about Explicit Racism" construct scores have similar spreads that appear to be symmetric. In addition, the scores for this construct share very similar median values for both pre and post intervention. Outliers are not present for neither pre nor post intervention measurements.

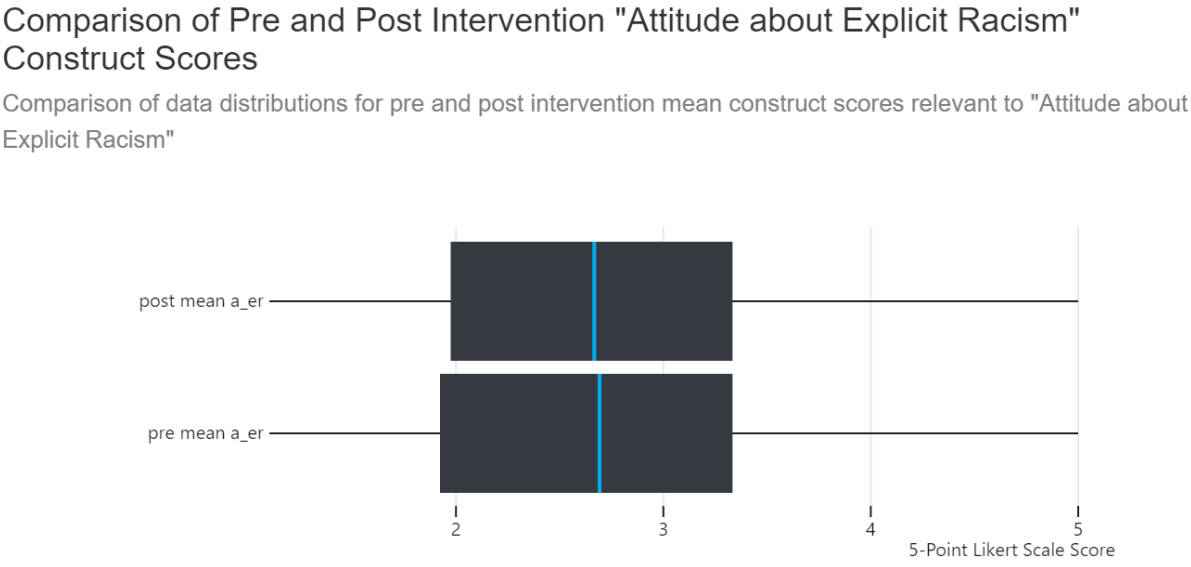


Figure 15: Box plot comparison of pre and post intervention "Attitude about Explicit Racism" construct scores where "pre mean a_er" represents pre-intervention mean "Attitude about Explicit Racism" and "post mean a_er" represents post-intervention mean "Attitude about Explicit Racism" measured on a 5-point Likert Scale

On the basis of each participant unique "Random ID", the means of the "Attitude about Explicit Racism" construct measured in KAP #1 and KAP #2 were compared for the same participant. The results from the KAP #1 (M = 2.72, SD = 0.915) survey and KAP #2 (M = 2.69, SD = 0.898) survey indicate that there was not a significant difference in attitude about explicit racism after participating in the SGD workshop intervention, $t(36) = 0.197$, $p = .84$.

4.1.3 Attitude about Policing in Communities

Figure 16 outlines the differences in pre-intervention mean "Attitude about Policing in Communities" (i.e., "pre mean a_pc") and post-intervention mean "Attitude about Policing in Communities" (i.e., "post mean a_pc") with box plots. As demonstrated below, both pre and post intervention mean "Attitude about Policing in Communities" construct scores have similar spreads. The post-intervention scores appear to be symmetrical in their spread. Contrastingly, the pre-intervention scores appear to be skewed right, indicating that the lower values for this construct score are closer together than the values that were scored as slightly higher. The scores for this construct have a similar median, however the post-intervention median value is greater than the pre-intervention median value. Finally, outliers appear to be present for both the pre and post intervention scores for this construct.

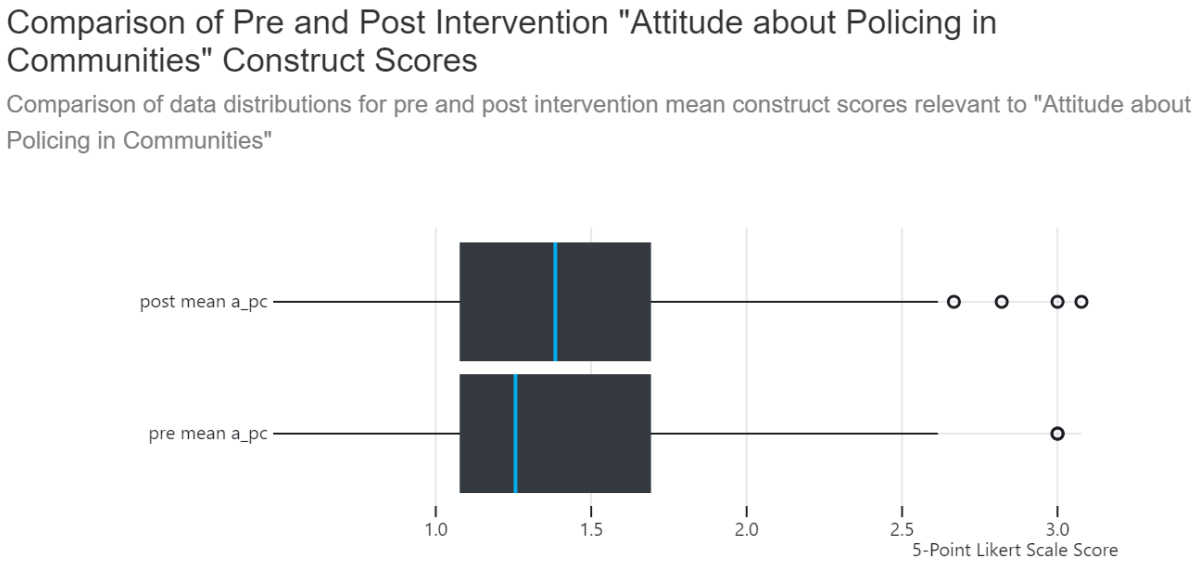


Figure 16: Box plot comparison of pre and post intervention "Attitude about Policing in Communities" construct scores where "pre mean a_pc" represents pre-intervention mean "Attitude about Policing in Communities" and "post mean a_pc" represents post-intervention mean "Attitude about Policing in Communities" measured on a 5-point Likert Scale

Furthermore, the means of the "Attitude about Policing in Communities" construct measured in KAP #1 and KAP #2 were compared for the same participant on the basis of their unique "Random ID". The results from the KAP #1 (M = 1.41, SD = 0.580) survey and KAP #2 (M = 1.53, SD = 0.513) survey indicate that there was not a significant difference in attitude about policing in communities after participating in the SGD workshop intervention, $t(36) = 0.700$, $p = .49$.

4.1.4 Practice about Policing in Communities

With box plots, Figure 17 outlines the differences in pre-intervention mean "Practice about Policing in Communities" (i.e., "pre mean p_pc") and post-intervention mean "Practice about Policing in Communities" (i.e., "post mean p_pc"). As illustrated below, the post-intervention "Practice about Policing in Communities" has a slightly larger spread than the pre-intervention "Practice about Policing in Communities". Furthermore, while the pre-intervention scores appear to be symmetric, the post-intervention scores appear to be slightly skewed right. Regardless, however, both pre and post intervention scores appear to share the same median value. Outliers are not present for neither pre nor post intervention measurements.

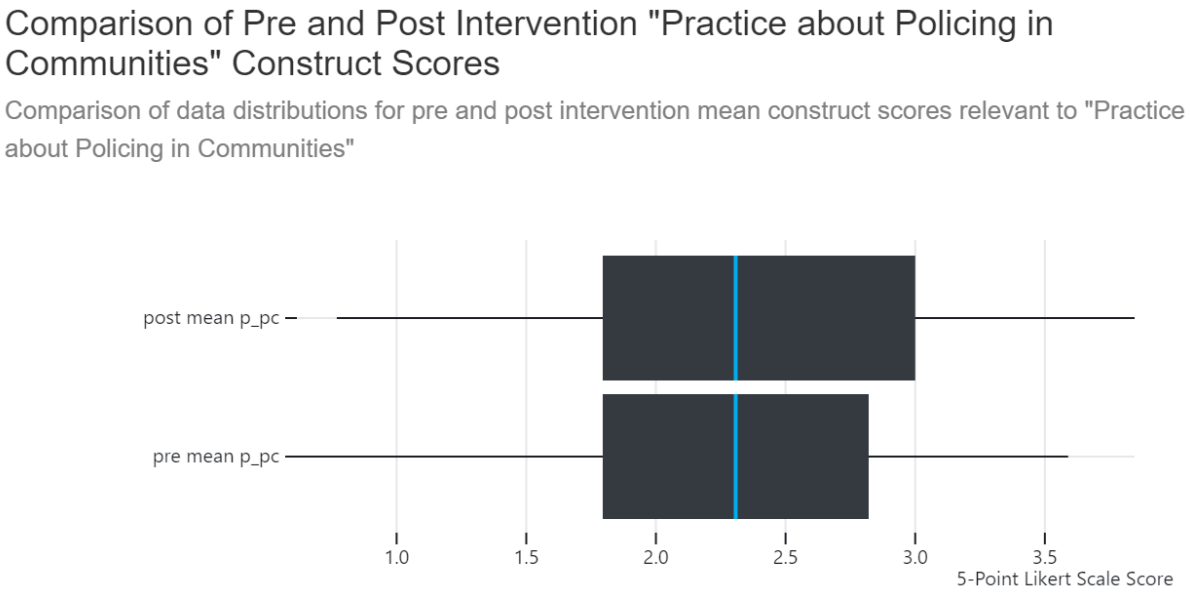


Figure 17: Box plot comparison of pre and post intervention "Practice about Policing in Communities" construct scores where "pre mean p_pc" represents pre-intervention mean "Practice about Policing in Communities" and "post mean p_pc" represents post-intervention mean "Practice about Policing in Communities" measured on a 5-point Likert Scale

Finally, the means of the "Practice about Policing in Communities" construct measured in KAP #1 and KAP #2 were compared for the same participant on the basis of their unique "Random ID". The results from the KAP #1 (M = 2.28, SD = 0.772) survey and KAP #2 (M = 2.33, SD = 0.824) survey indicate that there was not a significant difference in practice (i.e., behavior) about policing in communities after participating in the SGD workshop intervention, $t(36) = 0.981$, $p = .33$.

4.2 Between-Subjects: Analysis of Covariance (ANCOVA) Results

As described in section 3.2.4.2, the quantitative data pertaining to constructs of "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" was analyzed from a between-subjects perspective with the analysis of covariance (ANCOVA) in order to gauge how participants' different workshop protocols (i.e., Workshop A and Workshop B) and different demographic characteristics relevant to gender, age, and race could have impacted participants' experiences with the SGD workshop intervention. As such, the effectiveness of the intervention for different groups of participants (on the basis of workshop protocol, gender, age, and race) was investigated. With ANCOVA, the means of each construct measured in KAP #1 (pre SGD workshop intervention) were used for the covariate variable. The means of each construct measured in KAP #2 (post SGD workshop intervention) were used for the dependent response variable. Multiple ANCOVA tests were conducted where workshop protocol, gender, age, and race each served as the independent factor variable. Notably, the calculated "social desirability score" for each participant was applied to both KAP #1 data and KAP #2 before subsequent t-testing. An alpha level (α) of 0.05 was used for all tests. Before running ANCOVA, all datasets were first tested for normality with a Shapiro-Wilk test, and all datasets passed for normality. Appendix H outlines all between-subjects: analysis of covariance (ANCOVA) results in tabular form.

4.2.1 Knowledge about Policing in Communities

As can be observed in Figure 18, ANCOVA testing relevant to "Knowledge about Policing in Communities" revealed that only the factor variable of age resulted in a calculated p-value that indicated a significant effect of age group on post SGD workshop intervention "Knowledge about Policing in Communities" after controlling for pre SGD workshop intervention "Knowledge about Policing in Communities" (i.e., $p < .05$). All other factor variables relevant to gender, protocol, and race resulted in calculated p-values which indicated statistical insignificance (i.e., $p > .05$).

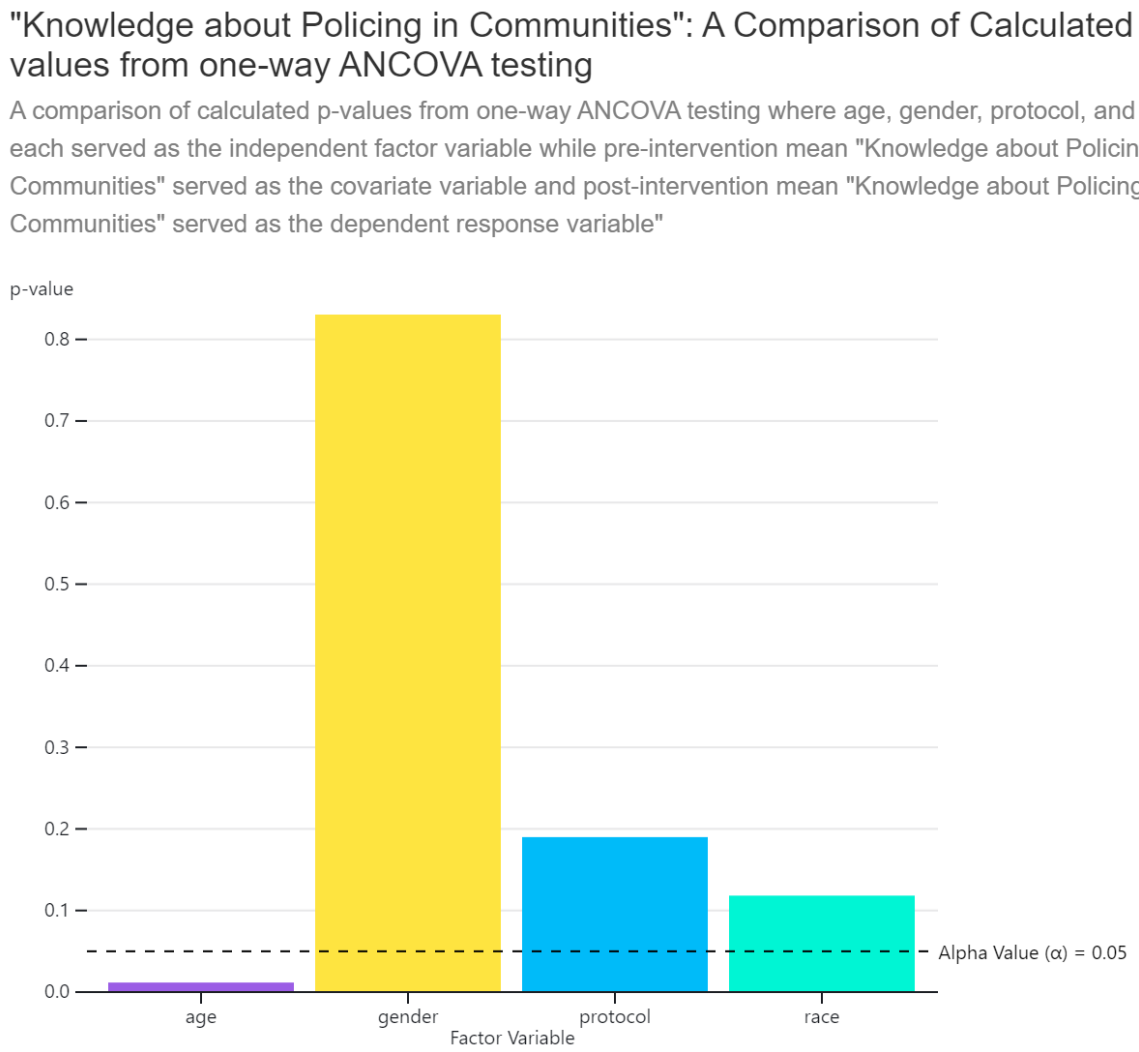


Figure 18: A comparison of calculated p-values from one-way ANCOVA testing where age, gender, protocol, and race each served as the independent factor variable while pre-intervention mean "Knowledge about Policing in Communities" served as the covariate variable and post-intervention mean "Knowledge about Policing in Communities" served as the dependent response variable

4.2.1.1 Gender

A one-way ANCOVA test was conducted to determine a statistically significant difference between gender identities on post SGD workshop intervention "Knowledge about Policing in Communities" controlling for pre SGD workshop intervention "Knowledge about Policing in Communities". This analysis indicates that there is no significant effect of gender identity on post SGD workshop intervention "Knowledge about Policing in Communities" after controlling for pre SGD workshop intervention "Knowledge about Policing in Communities", $F(2, 33) = 0.187$, $p = .830$.

4.2.1.2 Age

A one-way ANCOVA test was conducted to determine a statistically significant difference between age groups on post SGD workshop intervention "Knowledge about Policing in Communities" controlling for pre SGD workshop intervention "Knowledge about Policing in Communities". Results from this analysis indicate that there is a significant effect of age group on post SGD workshop intervention "Knowledge about Policing in Communities" after controlling for pre SGD workshop intervention "Knowledge about Policing in Communities", $F(3, 32) = 4.30$, $p = .012$.

In further investigating the significant effect of age group on post SGD workshop intervention "Knowledge about Policing in Communities", posteriori (i.e., post hoc) analysis was conducted first with a Tamhane T2 test. It can be observed in Table 7 that results from this analysis suggest that there could be different distributions of data between age groups of 25-34 years old and 18-24 years old when considering post SGD intervention "Knowledge about Policing in Communities". As also observed in Table 7, differences in data distributions could not be calculated for the age group 55-64 as there was only 1 participant who fell within this age demographic.

Table 7: Tamhane T2 post hoc test results for investigating which specific age group means for the "Knowledge about Policing in Communities" post intervention construct score differ from others

Age Groups	35-44 years old	25-34 years old	18-24 years old	55-64 years old
35-44 years old	1.000000	0.031257	0.811278	NaN
25-34 years old	0.031257	1.000000	0.301448	NaN
18-24 years old	0.811278	0.301448	1.000000	NaN
55-64 years old	NaN	NaN	NaN	1.0

In further investigating these outcomes, descriptive statistics indicated that there was a slight increase from 0.94 to 0.97 in the measured mean "Knowledge about Policing in Communities" for pre and post SGD workshop intervention levels for all 11 participants aged between 18-24 years old. Further descriptive statistics indicated that there was a slight decrease in the mean of this construct score from pre SGD workshop intervention levels to post SGD workshop intervention levels from 0.94 to 0.88 respectively for all 22 participants aged between 25 and 34 years old. Finally, or participants aged between 35 and 44 years old, the mean "Knowledge about Policing in Communities" construct score increased slightly from 0.89 to 1.0 for pre SGD workshop intervention scoring and post SGD workshop intervention scoring when considering all 3 participants in this age range.

4.2.1.3 Race

A one-way ANCOVA test was conducted to determine a statistically significant difference between racial identities on post SGD workshop intervention "Knowledge about Policing in Communities" controlling for pre SGD workshop intervention "Knowledge about Policing in Communities". This analysis indicates that there is no significant effect of racial identity on post SGD workshop intervention "Knowledge about Policing in Communities" after controlling for pre SGD workshop intervention "Knowledge about Policing in Communities" $F(1, 34) = 2.57$, $p = .118$.

4.2.1.4 Workshop Protocol

A one-way ANCOVA test was conducted to determine a statistically significant difference between workshop protocols on post SGD workshop intervention "Knowledge about Policing in Communities" controlling for pre SGD workshop intervention "Knowledge about Policing in Communities". Results from this analysis indicate that there is no significant effect of workshop protocol on post SGD workshop intervention "Knowledge about Policing in Communities" after controlling for pre SGD workshop intervention "Knowledge about Policing in Communities" $F(1, 34) = 1.79$, $p = .190$.

4.2.2 Attitude about Explicit Racism

Illustrated in Figure 19, ANCOVA testing relevant to "Attitude about Explicit Racism" revealed that all factor variables of age, gender, protocol, and race resulted in calculated p-values which indicated statistical insignificance (i.e., $p > .05$). As such, neither age, gender, protocol, nor race had a significant effect on this measured construct.

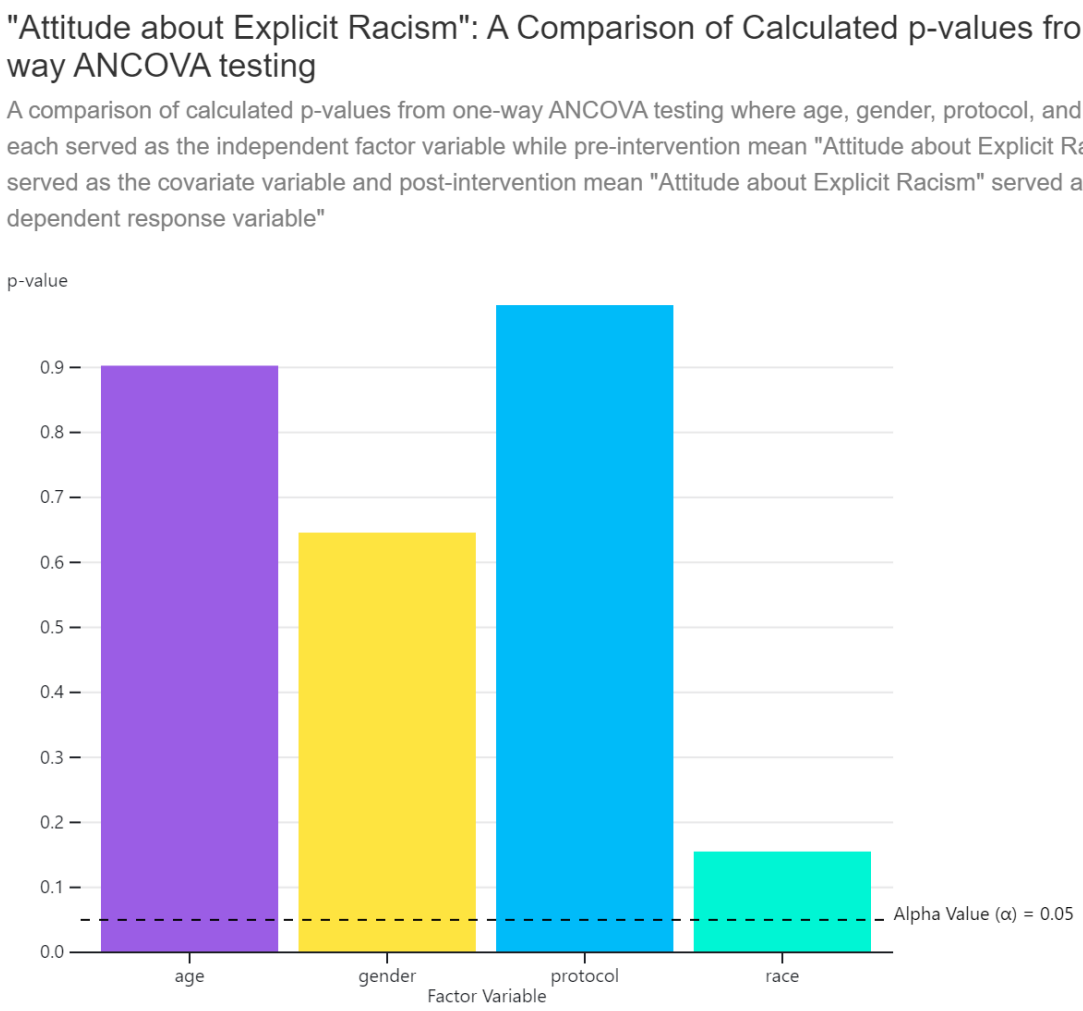


Figure 19: A comparison of calculated p-values from one-way ANCOVA testing where age, gender, protocol, and race each served as the independent factor variable while pre-intervention mean "Attitude about Explicit Racism" served as the covariate variable and post-intervention mean "Attitude about Explicit Racism" served as the dependent response variable

4.2.1.1 Gender

A one-way ANCOVA test was conducted to determine a statistically significant difference between gender identities on post SGD workshop intervention "Attitude about Explicit Racism" controlling for pre SGD workshop intervention "Attitude about Explicit Racism". This analysis indicates that there is no significant effect of gender identity on post SGD workshop intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Explicit Racism", F (2, 33) = 0.443, p = .646.

4.2.1.2 Age

A one-way ANCOVA test was also conducted to determine a statistically significant difference between age groups on post SGD workshop intervention "Attitude about Explicit Racism" controlling for pre SGD workshop intervention "Attitude about Explicit Racism". Results from this analysis indicate that there is no significant effect of age group on post SGD workshop intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Explicit Racism", F (3, 32) = 0.190, p = .903.

4.2.2.3 Race

Considering race, a one-way ANCOVA test was conducted to determine a statistically significant difference between racial identities on post SGD workshop intervention "Attitude about Explicit Racism" controlling for pre SGD workshop intervention "Attitude about Explicit Racism". This analysis indicates that there is no significant effect of racial identity on post SGD workshop intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Explicit Racism", $F(1, 34) = 2.11$, $p = .155$.

4.2.2.4 Workshop Protocol

In addition, a one-way ANCOVA test was conducted to determine a statistically significant difference between workshop protocols on post SGD workshop intervention "Attitude about Explicit Racism" controlling for pre SGD workshop intervention "Attitude about Explicit Racism". Results from this analysis indicate that there is no significant effect of workshop protocol on post SGD workshop intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Explicit Racism" $F(1, 34) = 0.000028$, $p = .996$.

4.2.3 Attitude about Policing in Communities

ANCOVA testing relevant to "Attitude about Policing in Communities" revealed that all factor variables relevant to age, gender, protocol, and race resulted in calculated p-values which indicated statistical insignificance (i.e., $p > .05$). As such, and as observed in Figure 20, neither age, gender, protocol, nor race had a significant effect on this measured construct.

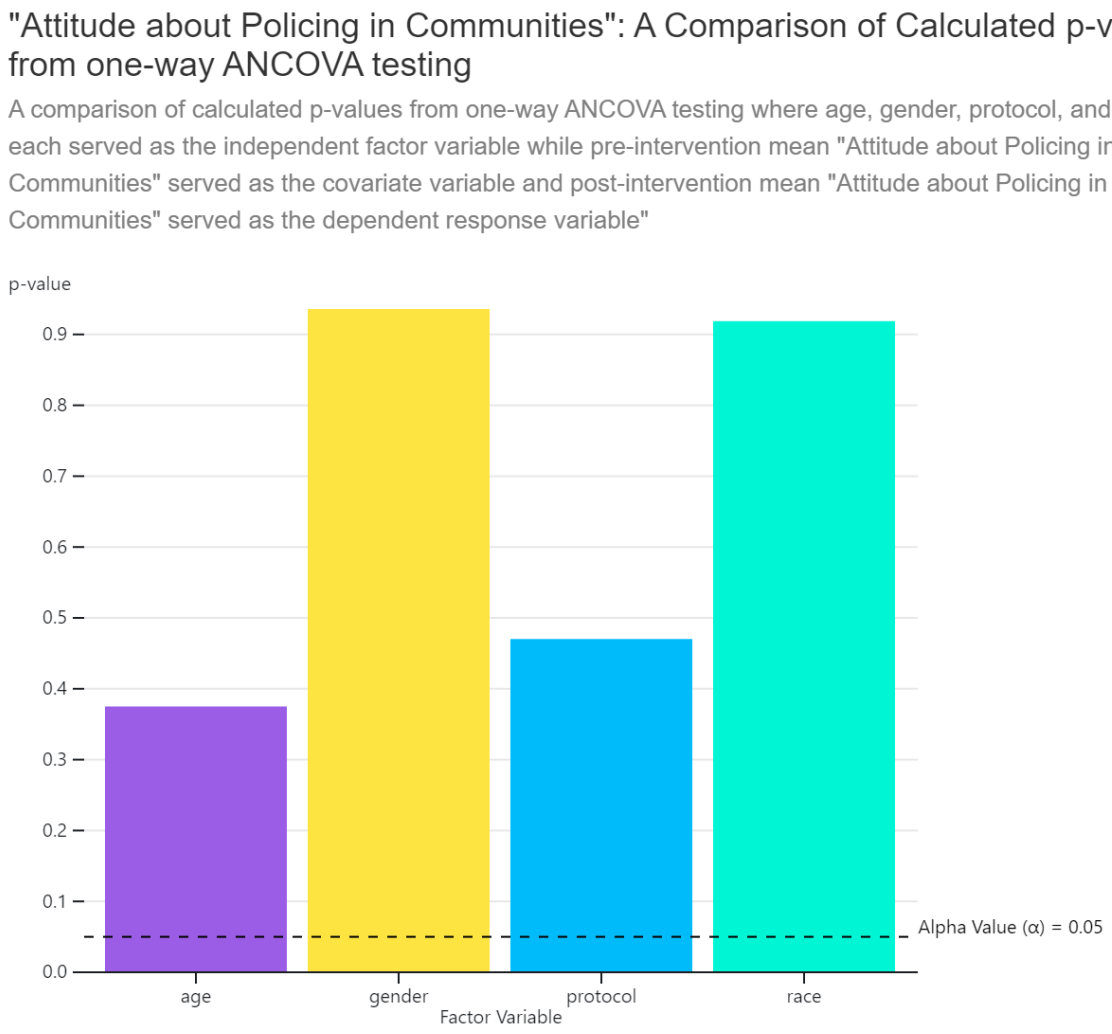


Figure 20: A comparison of calculated p-values from one-way ANCOVA testing where age, gender, protocol, and race each served as the independent factor variable while pre-intervention mean "Attitude about Policing in Communities" served as the covariate variable and post-intervention mean "Attitude about Policing in Communities" served as the dependent response variable

4.2.3.1 Gender

Considering gender, a one-way ANCOVA test was conducted to determine a statistically significant difference between gender identities on post SGD workshop intervention "Attitude about Policing in Communities" controlling for pre SGD workshop intervention "Attitude about Policing in Communities". This analysis indicates that there is no significant effect of gender identity on post SGD workshop intervention "Attitude about Policing in Communities" after controlling for pre SGD workshop intervention "Attitude about Policing in Communities", F (2, 33) = 0.00756, p = .925.

4.2.3.2 Age

A one-way ANCOVA test was conducted to determine a statistically significant difference between age groups on post SGD workshop intervention "Attitude about Policing in Communities" controlling for pre SGD workshop intervention "Attitude about Policing in Communities". Results from this analysis indicate that there is no significant effect of age group on post SGD workshop intervention "Attitude about Policing in Communities" after controlling for pre SGD workshop intervention "Attitude about Policing in Communities", F (3, 32) = 1.071, p = .375.

4.2.3.3 Race

In addition, a one-way ANCOVA test was conducted to determine a statistically significant difference between racial identities on post SGD workshop intervention "Attitude about Policing in Communities" controlling for pre SGD workshop intervention "Attitude about Policing in Communities". This analysis indicates that there is no significant effect of racial identity on post SGD workshop intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Policing in Communities", F (1, 34) = 0.0106, p = .919.

4.2.3.4 Workshop Protocol

A one-way ANCOVA test was conducted to determine a statistically significant difference between workshop protocols on post SGD workshop intervention "Attitude about Policing in Communities" controlling for pre SGD workshop intervention "Attitude about Policing in Communities". Results from this analysis indicate that there is no significant effect of workshop protocol on post SGD workshop intervention "Attitude about Policing in Communities" after controlling for pre SGD workshop intervention "Attitude about Policing in Communities" F (1, 34) = 0.533, p = .470.

4.2.4 Practice about Policing in Communities

Illustrated in Figure 21, ANCOVA testing relevant to "Practice about Policing in Communities" revealed that all factor variables relevant to age, gender, protocol, and race resulted in calculated p-values which indicated statistical insignificance (i.e., $p > .05$). As such, neither age, gender, protocol, nor race had a significant effect on this measured construct.

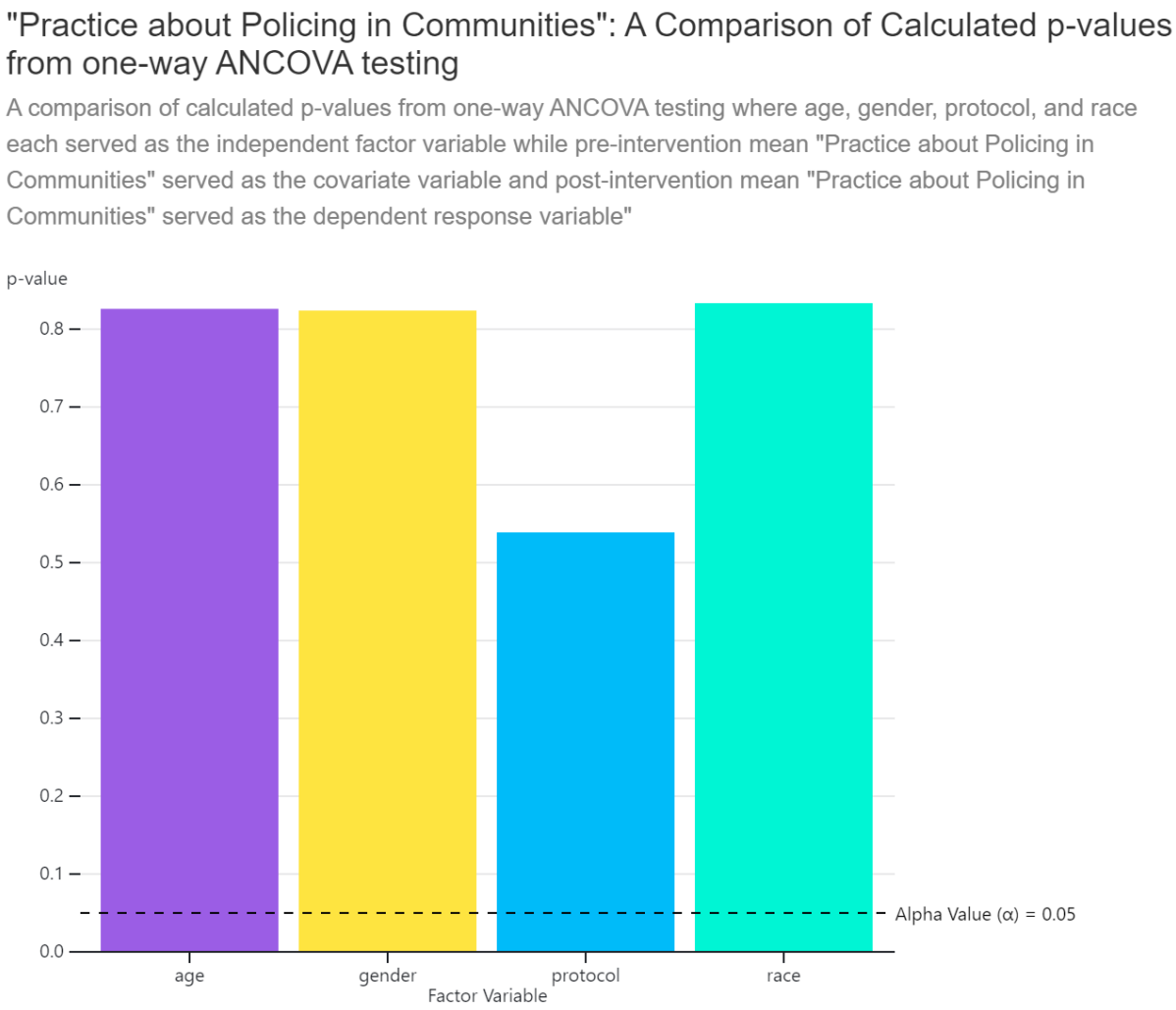


Figure 21: A comparison of calculated p-values from one-way ANCOVA testing where age, gender, protocol, and race each served as the independent factor variable while pre-intervention mean "Practice about Policing in Communities" served as the covariate variable and post-intervention mean "Practice about Policing in Communities" served as the dependent response variable

4.2.4.1 Gender

A one-way ANCOVA test was conducted to determine a statistically significant difference between gender identities on post SGD workshop intervention "Practice about Policing in Communities" controlling for pre SGD workshop intervention "Practice about Policing in Communities". This analysis indicates that there is no significant effect of gender identity on post SGD workshop intervention "Practice about Policing in Communities" after controlling for pre SGD workshop intervention "Practice about Policing in Communities $F(2, 33) = 0.195$, $p = .824$.

4.2.4.2 Age

Additionally, a one-way ANCOVA test was conducted to determine a statistically significant difference between age groups on post SGD workshop intervention "Practice about Policing in Communities" controlling for pre SGD workshop intervention "Practice about Policing in Communities". Results from this analysis indicate that there is no significant effect of age group on post SGD workshop intervention

"Practice about Policing in Communities" after controlling for pre SGD workshop intervention "Practice about Policing in Communities", F (3, 32) = 0.299, p = .826.

4.2.4.3 Race

Considering race, a one-way ANCOVA test was conducted to determine a statistically significant difference between racial identities on post SGD workshop intervention "Practice about Policing in Communities" controlling for pre SGD workshop intervention "Practice about Policing in Communities". This analysis indicates that there is no significant effect of racial identity on post SGD workshop intervention "Practice about Policing in Communities" after controlling for pre SGD workshop intervention "Practice about Policing in Communities"", F (1, 34) = 0.0449, p = .833.

4.2.4.4 Workshop Protocol

Finally, a one-way ANCOVA test was conducted to determine a statistically significant difference between workshop protocols on post SGD workshop intervention "Practice about Policing in Communities" controlling for pre SGD workshop intervention "Practice about Policing in Communities". Results from this analysis indicate that there is no significant effect of workshop protocol on post SGD workshop intervention "Practice about Policing in Communities" after controlling for pre SGD workshop intervention "Practice about Policing in Communities" F (1, 34) = 0.385, p = .539.

4.3 Perceived Effectiveness of Participatory SGD Workshops

With the participant effectiveness of participatory SGD workshops evaluation data described in section 3.2.3, participant associated scores relevant to the measured evaluated constructs of "structure & flow", "usability", "team communication", and "overall effectiveness" were investigated in order to understand how participants of different demographic backgrounds perceived the effectiveness of the participatory SGD workshop as an intervention.

In investigating this data, basic statistical analysis was first computed for the overall 38 participants' evaluations while considering each of their tendencies to respond in a socially desirable way. The results displayed in Table 8 indicated that on average and regardless of demographic characteristics or workshop protocol, participants scored the "overall effectiveness" of the participatory SGD workshops as approximately 2.69. In disaggregating the overall effectiveness, participants also gave an average approximate score of 2.73, 2.65, and 2.70 for participatory SGD constructs related to "structure & flow", "usability", and "team communication" respectively. These results can also be observed in Figure 22.

Table 8: Descriptive statistics for overall participants' evaluations of the perceived effectiveness of participatory SGD workshops (considering their tendencies to respond in socially desirable ways) for constructs relevant to "overall effectiveness", "structure & flow", "usability", and "team communication"

	Overall Effectiveness Score	Structure & Flow Score	Usability Score	Team Communication Score
count	38.000000	38.000000	38.000000	38.000000
mean	2.689087	2.731444	2.649966	2.696356
std	0.871057	0.916794	0.850329	0.907407
min	1.111888	1.153846	1.096154	0.961538
25%	2.045455	1.955128	2.033654	2.033654
50%	2.706294	2.692308	2.576923	2.692308
75%	3.305944	3.557692	3.173077	3.115385
max	4.909091	5.000000	4.750000	5.000000

Figure 22: Mean perceived participatory SGD workshop effectiveness score for constructs relevant to "structure & flow", "usability", and "team communication" for all workshop participants adjusted for social desirability tendencies

4.3.2 Gender

In disaggregating the evaluation results and considering gender, further analysis indicated that participants who gender identified as female on average scored the "overall effectiveness" of the participatory SGD workshops as approximately 2.66, and participants who gender identified as male on average scored the "overall effectiveness" of the participatory SGD workshops as approximately 2.67. The participant who gender identified as "other" or a gender identity that was not listed on the survey scored the "overall effectiveness" of the participatory SGD workshops as approximately 3.61.

In disaggregating the overall effectiveness, female participants also gave an average approximate score of 2.74, 2.60, and 2.65 for participatory SGD constructs related to "structure & flow", "usability", and "team communication" respectively. Similarly, male participants gave an average approximate score of 2.67, 2.64, and 2.71 for participatory SGD constructs related to "structure & flow", "usability", and "team communication" respectively. Finally, the single participant with the "other" gender identity scored the participatory SGD workshop constructs of "structure & flow", "usability", and "team communication" as approximately 3.70, 3.70, and 3.46 respectively. These results can also be observed in Figure 23.

Figure 23: Mean perceived participatory SGD workshop effectiveness score for constructs relevant to “structure & flow”, “usability”, and “team communication” for all workshop participants adjusted for social desirability tendencies visualized by gender

4.3.3 Age

In considering age and disaggregating the evaluation results, analysis also indicated that participants who were aged between 18-24 years old, 25-34 years old, 35-44 years old, and 55-64 years old on average scored the “overall effectiveness” of the participatory SGD workshops as approximately 3.05, 2.45, 3.72, and 1.11 respectively. 18-24 year old participants also gave an average approximate score of 3.06, 2.98, and 3.11 for participatory SGD constructs related to “structure & flow”, “usability”, and “team communication” respectively. 25-34 year old participants also gave an average approximate score of 2.49, 2.41, and 2.46 for participatory SGD constructs related to “structure & flow”, “usability”, and “team communication” respectively. For participants aged 35-44 years old, constructs of “structure & flow”, “usability”, and “team communication” were on average approximately scored respectively as 3.91, 3.76, and 3.55. Finally, for the participant aged 55-64 years old, approximate scores of 1.15, 1.10, and 1.10 were allotted respectively to constructs related to “structure & flow”, “usability”, and “team communication”. These results have been visualized in Figure 24.

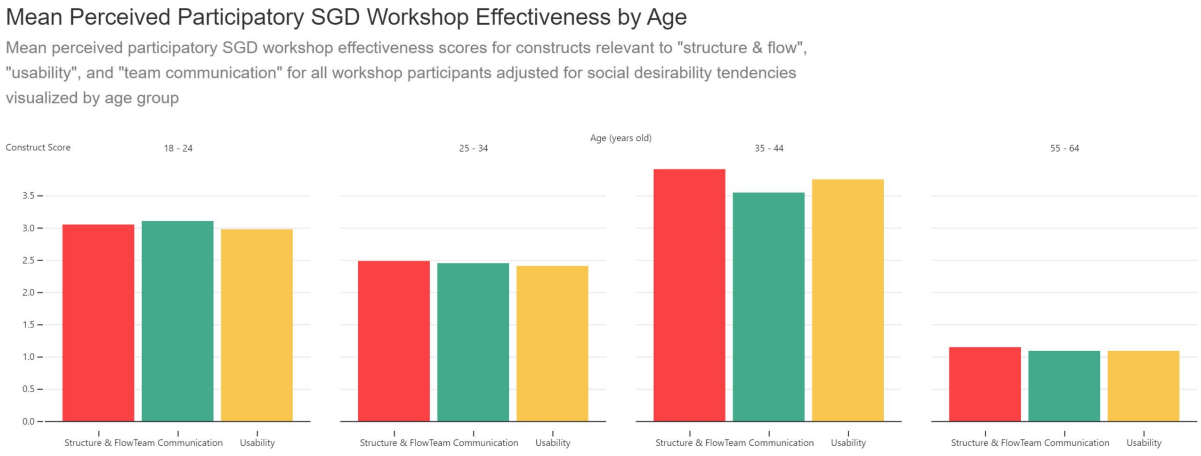


Figure 24: Mean perceived participatory SGD workshop effectiveness score for constructs relevant to “structure & flow”, “usability”, and “team communication” for all workshop participants adjusted for social desirability tendencies visualized by age groups

4.3.4 Race

Furth analysis also demonstrated that participants BIPOC participants on average scored the "overall effectiveness" of the participatory SGD workshops as approximately 2.66. Contrastingly, white participants on average scored the participatory SGD workshop "overall effectiveness" as 2.72. In disaggregating the overall effectiveness, BIPOC participants also gave an average approximate score of 2.66, 2.69, and 2.58 for participatory SGD constructs related to "structure & flow", "usability", and "team communication" respectively. For white participants, constructs of "structure & flow", "usability", and "team communication" were on average approximately scored respectively as 2.77, 2.72, and 2.66. These results have been visualized in Figure 25.

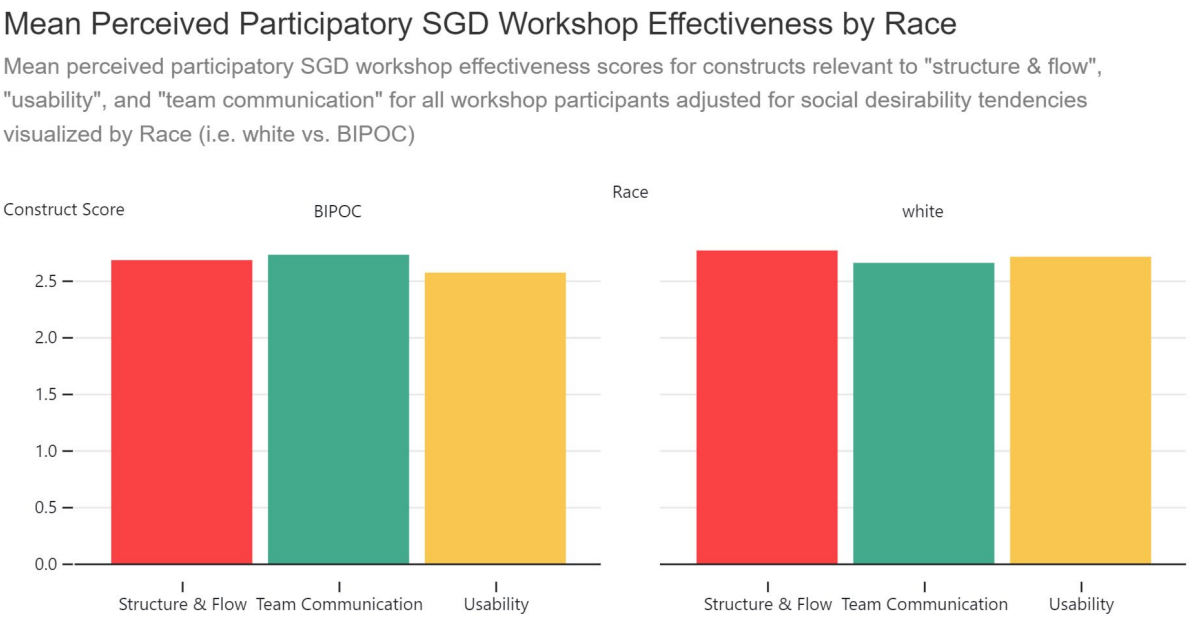


Figure 25: Mean perceived participatory SGD workshop effectiveness score for constructs relevant to "structure & flow", "usability", and "team communication" for all workshop participants adjusted for social desirability tendencies visualized by race (i.e., white vs. BIPOC)

4.3.5 Protocol

Finally, analysis also demonstrated that Workshop A protocol participants on average scored the "overall effectiveness" of the participatory SGD workshops as approximately 2.60, while Workshop B protocol participants on average scored this same construct as approximately 2.76. In addition, Workshop A protocol participants gave an average approximate score of 2.63, 2.52, and 2.66 for participatory SGD constructs related to "structure & flow", "usability", and "team communication" respectively. For Workshop B protocol participants, constructs of "structure & flow", "usability", and "team communication" were on average approximately scored respectively as 2.81, 2.75, and 2.72. These results have been visualized in Figure 26.

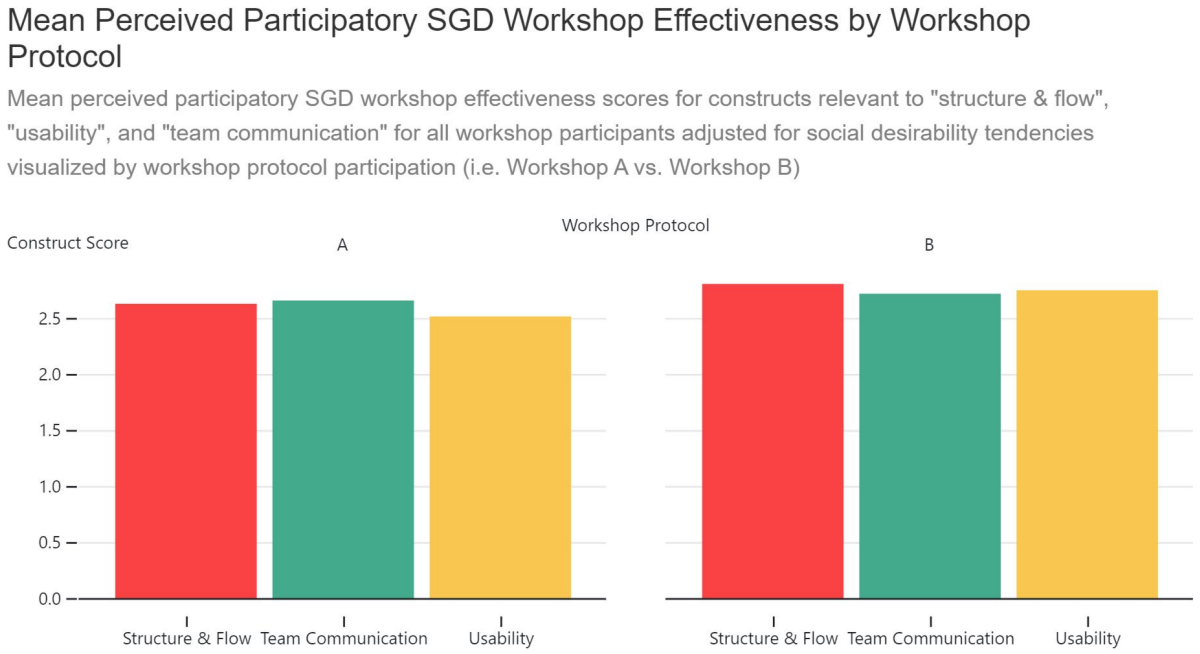


Figure 26: Mean perceived participatory SGD workshop effectiveness score for constructs relevant to "structure & flow", "usability", and "team communication" for all workshop participants adjusted for social desirability tendencies visualized by workshop protocol (i.e., Workshop A vs. Workshop B)

5

QUALITATIVE RESULTS

This chapter outlines the results obtained from the qualitative component of this study. Here, the results for the content analysis (section 5.1), sentiment analysis (5.2), and micro-interlocutor analysis (section 5.3) are detailed. In analyzing this data and obtaining results, *SQ5* is partially addressed from a qualitative perspective, and insight into reducing police bias and improving police-community relations through game design is generated.

5.1 Content Analysis Results

As described in section 3.3.3.1, the words from all brainstorming sticky notes were first counted and then interpreted using supporting observation note data from the SGD activities. Results were analyzed for all workshop sessions, Workshop A sessions only, and Workshop B sessions only at both the aggregated and disaggregated brainstorming levels (i.e., for the total brainstorming activity vs. for each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience).

5.1.1 All Workshop Sessions

Figure 16 below outlines a bar chart of the most frequently used words at the aggregated brainstorming level (i.e., for all brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience) for all of the workshop sessions (i.e., both Workshop A sessions and Workshop B sessions). As indicated by Figure 16, “police”, “community”, and “different” were the most commonly used words in brainstorming sessions across all workshop sessions.

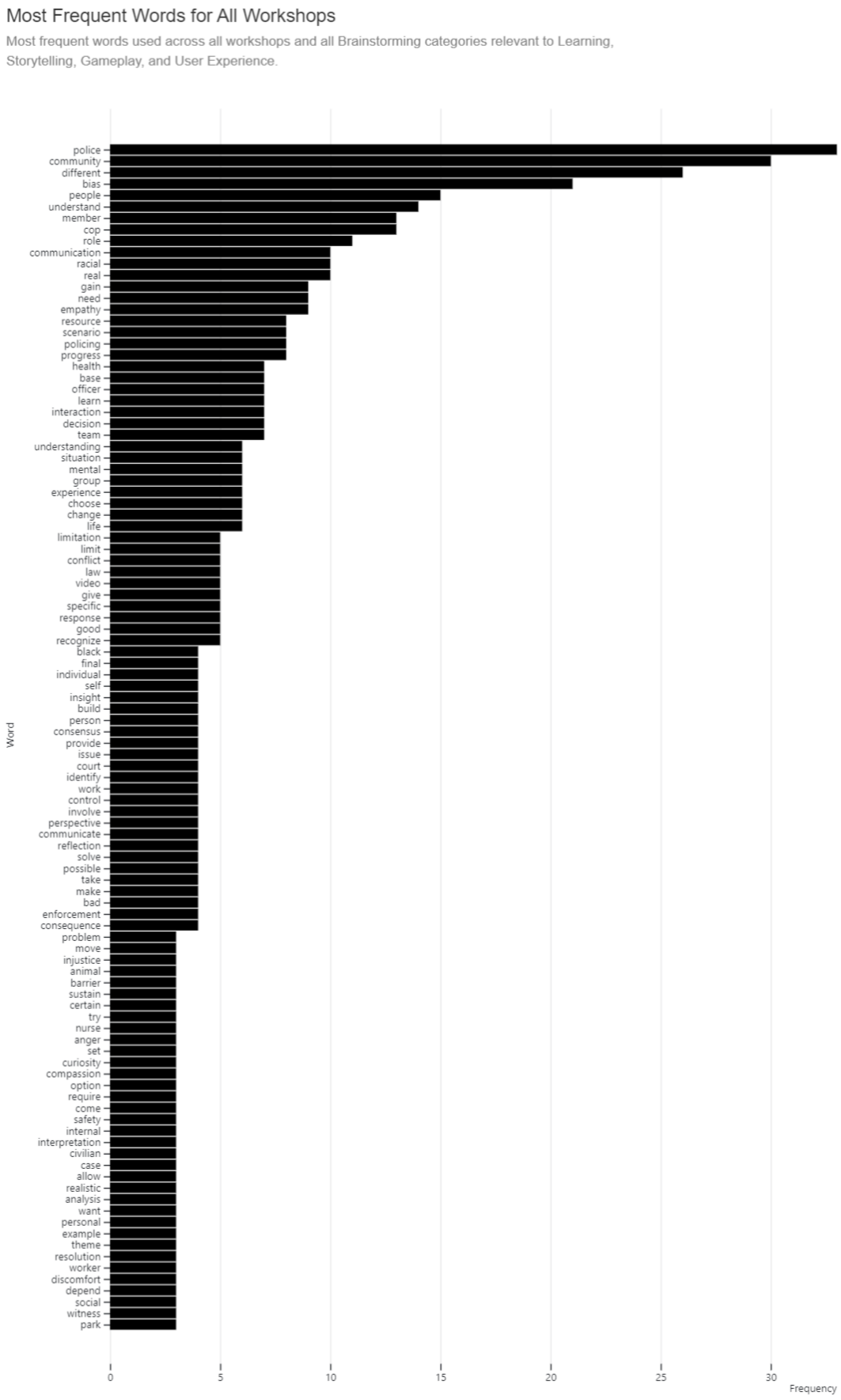


Figure 16: Most frequent words for all workshops and all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience

In disaggregating the brainstorming activities, Figure 17 indicates the most frequently used words for all of the workshop sessions per each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience. Notably, the plots presented in Figure 17 and subsequently in Figures 19 and 21 are plotted with different scales on the x-axis for each brainstorming activity. This was done in order to ensure that no brainstorming categories skewed the importance of different frequently used words within other brainstorming categories. For example, while the Learning top word of "bias" was used with even more frequency than the Gameplay top word of "different", this does not necessarily mean that the Gameplay word of "different" is less important than the Learning word of "bias". Rather, what is important within brainstorming categories is to understand which word was used more than others. It is less important to understand which word was used the most across brainstorming categories. As such, these plots were designed with different x-axis scales in order to preserve and mainly highlight the difference in word use frequency within each brainstorming category.

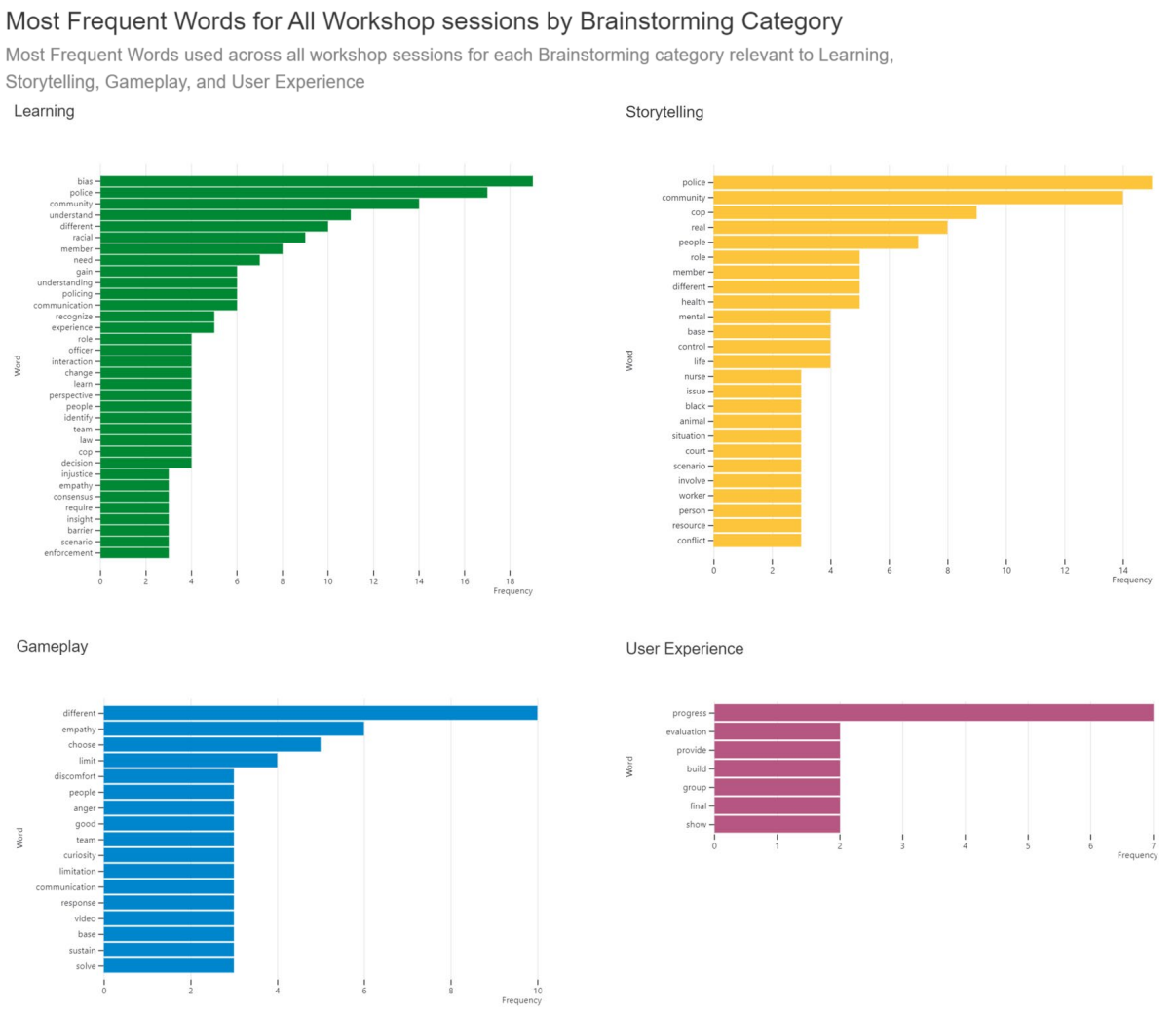


Figure 17: Most frequent words for all workshops per each Brainstorming category relevant to Learning (top left), Storytelling (top right), Gameplay (bottom left), and User Experience (bottom right)

For the Learning brainstorming activity, "bias", "police", and "community" were the most commonly used words. During the Storytelling brainstorming activity, "police", "community", and "cop" were the most commonly used words. "Different", "empathy", and "choose" were the most commonly used words during the Gameplay brainstorming activity. Finally, for the User Experience brainstorming activity, "progress" was the most commonly used word.

5.1.2 Workshop A Sessions

Figure 18 below outlines a bar chart of the most frequently used words at the aggregated brainstorming level (i.e., for all brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience) for Workshop A sessions. As indicated by Figure 18, "community", "police", and "bias" were the most commonly used words in brainstorming sessions across all workshop sessions.

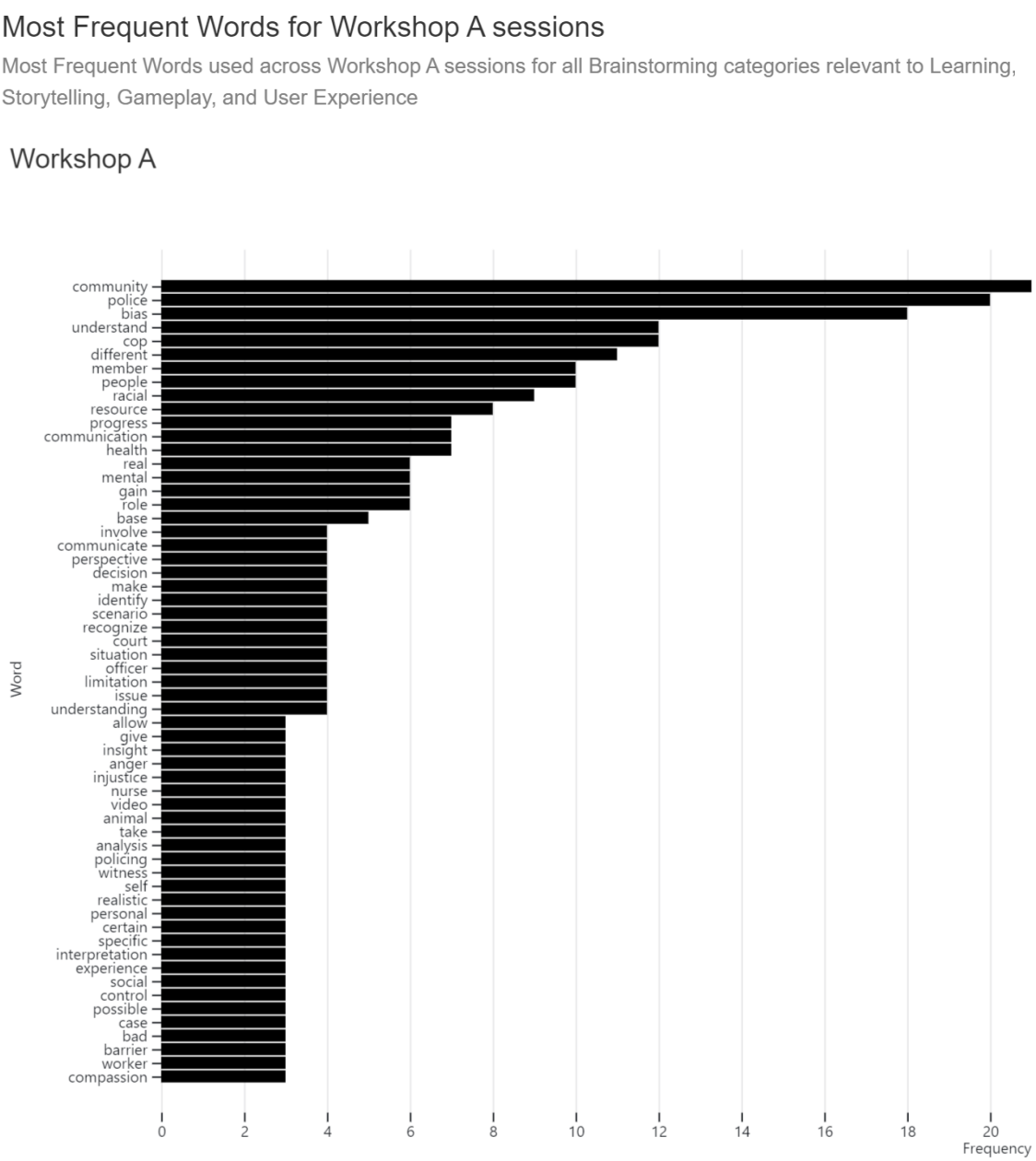


Figure 18: Most frequent words for Workshop A sessions and all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience

In disaggregating the brainstorming activities, Figure 19 indicates the most frequently used words for Workshop A sessions per each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience.

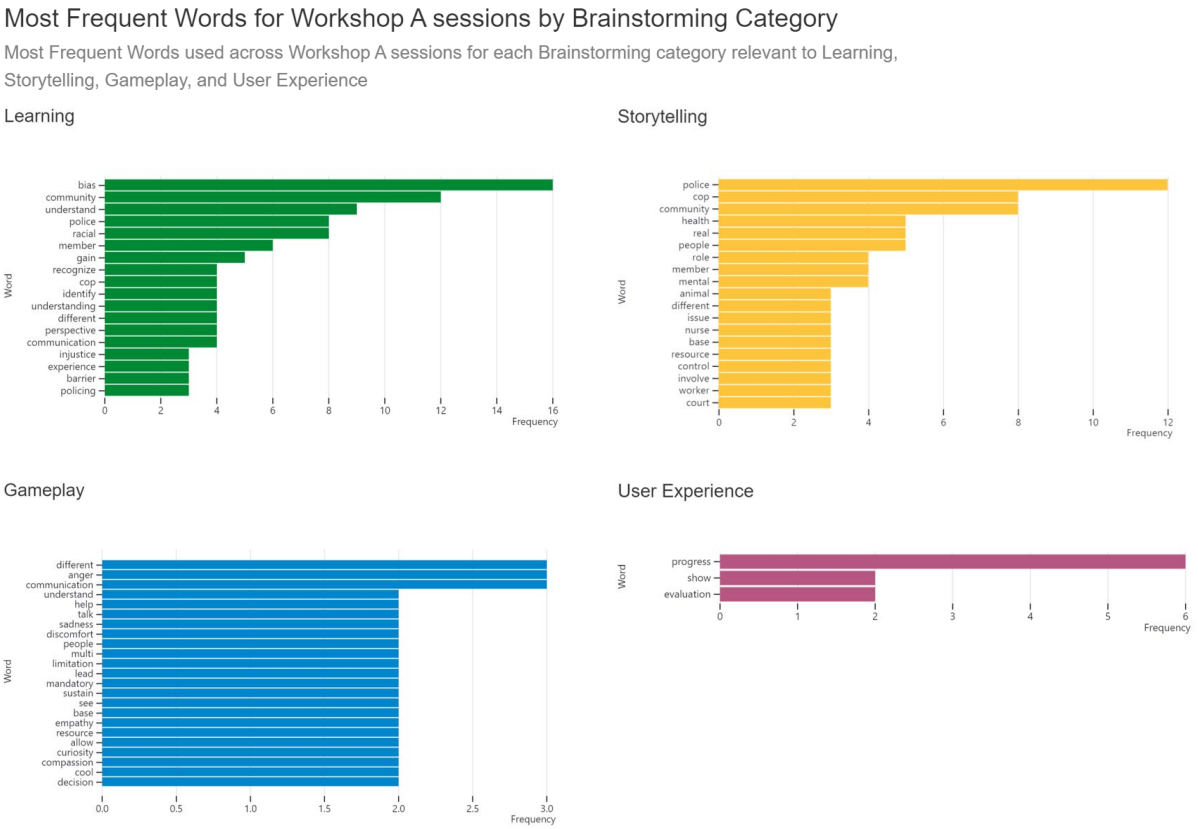


Figure 19: Most frequent words for Workshop A protocol per each Brainstorming category relevant to Learning (top left), Storytelling (top right), Gameplay (bottom left), and User Experience (bottom right)

Interestingly, it can be observed in Figure 19 that the words "community" and "police" were used with a higher frequency in brainstorming activities relevant to Learning and Storytelling. In the context of Learning, the word "community" was used 150% more frequently than the word "police". Contrastingly, in the context of Storytelling, the word "police" was used 150% more frequently than the word "community".

Learning

Specifically, for the Learning brainstorming activity, "bias", "community", and "understand" were the most commonly used words. Given the supporting observation notes from the SGD activities, it can be observed that the difficulties of understanding and teaching "bias" was discussed in the context of Workshop A Learning. Specifically, all Workshop A sessions spoke about how bias can be a very difficult concept to understand, and that it must be simplified in a way that is more approachable. Workshop A participants also recognized that while the learning goal of a serious game to address racial bias in policing should focus on players recognizing patterns of bias and exploring their own pre-existing biases, participants also spoke of how many instances of "bias" propagate because of a lack of understanding, vocabulary, and cultural competency to address such issue. As such, participants believed that addressing "bias" is not only a difficult task, but also a very important one. In considering the use of the word "community", supporting observation notes also indicated that participants believed that future players of their serious game designs should learn to better recognize patterns of racial injustice and bias that propagate throughout communities. Additionally, participants discussed the importance of probing police officers to learn and better understand their role in the community on a broader and more structural level. Finally, with regard to the use of the word "understand", participants believed that understanding racial bias and its many facets can be quite challenging, however it is important for individuals to learn and more deeply understand what their own biases are. In doing so,

police officers could better comprehend how their roles in communities fit into the grander scheme of society. Additionally, participants also discussed the importance of understanding different perspectives. As such, the goal of developing a serious game to address racial bias in policing should be about developing understanding.

Storytelling

For the Storytelling brainstorming activity, "police", "cop", and "community", were the most commonly used words. The word "police" was very frequently used when discussing the Storytelling component of serious game designs. Overwhelmingly, in the context of Storytelling, participants believed that the player of the serious games designed should take on the role of the police, and that a police officer should be the main character of the serious game. Participants also believed that within the game world, police officer characters should have to address a situation that happens within the community and interact with community members in the process. While all participants focused on police as key players and characters associated with the storyline, different participants used different language to describe this profession. Notably, many participants used the slang term "cop" when talking about the police. Unsurprisingly, this term was frequently used with a negative connotation to discuss instances where police officers have been involved in some sort of wrongdoing when addressing communities and community members. The word "community" was frequently brought up when discussing the importance of game world realities. Workshop A participants believed it very important to capture the history and current relationships that exist between communities and police in real world settings within the game world. However, in doing so, participants also felt that community members should take on a secondary role within the game, while an emphasis on police should be made.

Gameplay

For the Gameplay brainstorming activity, "different", "anger", "communication", and "understand" were the most commonly used words. With regard to the use of the word "different" in Workshop A Gameplay, the supporting observation notes outlined that participants believed that serious game designs should support different levels, different scenarios, different missions, and the ability to "level up" depending on a player's understanding of different taught content. Some Workshop A participants also spoke about a potential gameplay setting that would allow for different levels of subject matter comfort depending on a player's self-evaluated knowledge on anti-racism topics. Participants believed that such diversity in gameplay elements could help lead to shifting perspectives. Additionally, "anger" was a frequently used word when participants discussed the feelings that they wanted the gameplay of their serious game design to evoke. While participants felt it important for the player of the game to be frustrated and experience anger, they believed that it should be experienced in such a way that ultimately led to a changing viewpoint. Interestingly, when using the word "communication" in a gameplay context, Workshop A participants spoke about the importance of effective communication. As such, participants wanted to include communication as a key element within the gameplay context. Some participants suggested limiting communication between players or creating communication barriers with differences in language. Building upon this notion, participants then believed that creating space for more liberated communication throughout the progress of the game could be quite effective in highlighting the importance of communication and the dangers of communication breakdowns. Finally, the word "understanding" was used in the context of Workshop A gameplay to discuss emotions related to empathy and gameplay mechanics that would help players understand alternative perspectives. All participants believed that their serious game design should evoke some sort of emotion related to understanding. Participants also believed that it was important for players to achieve learning goals related to understanding diverse perspectives before they could progress in gameplay and level up.

User Experience

While all Workshop A sessions struggled to complete the User Experience ideation cards with the allotted available time, "progress" was the nonetheless the most commonly used word. Given the supporting observation notes from the SGD activities, it can be observed that Workshop A participants believed that their serious game designs should involve some sort of User Experience element that would allow players to see their progress after every level up or advancement in gameplay. In keeping track of such progress, participants believed that potential players should be presented with not only their game points and final scores, but also with a more detailed explicit evaluation of their progress relevant to taught content and their own biases. However, participants also acknowledged that such feedback could have the potential to be problematic, as players may be reluctant to have to face their own predispositions and biases in such a direct manner.

5.1.3 Workshop B Sessions

Figure 20 below outlines a bar chart of the most frequently used words at the aggregated brainstorming level (i.e., for all brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience) for Workshop B sessions. As indicated by Figure 20, "different", "police", and "community" were the most commonly used words in brainstorming sessions across all workshop sessions.

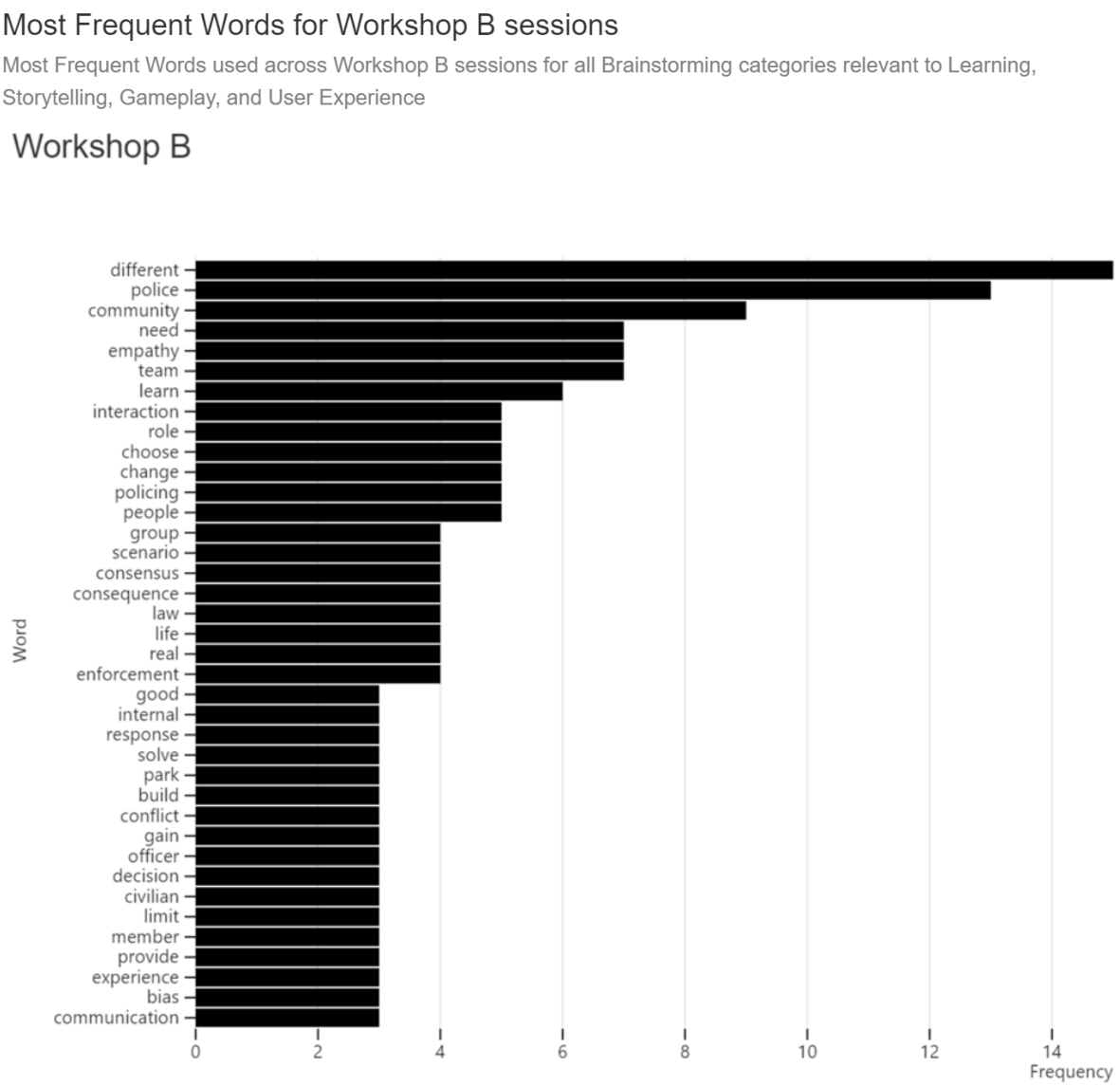


Figure 20: Most frequent words for Workshop B sessions and all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience

In disaggregating the brainstorming activities, Figure 21 indicates the most frequently used words for Workshop B sessions per each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience.

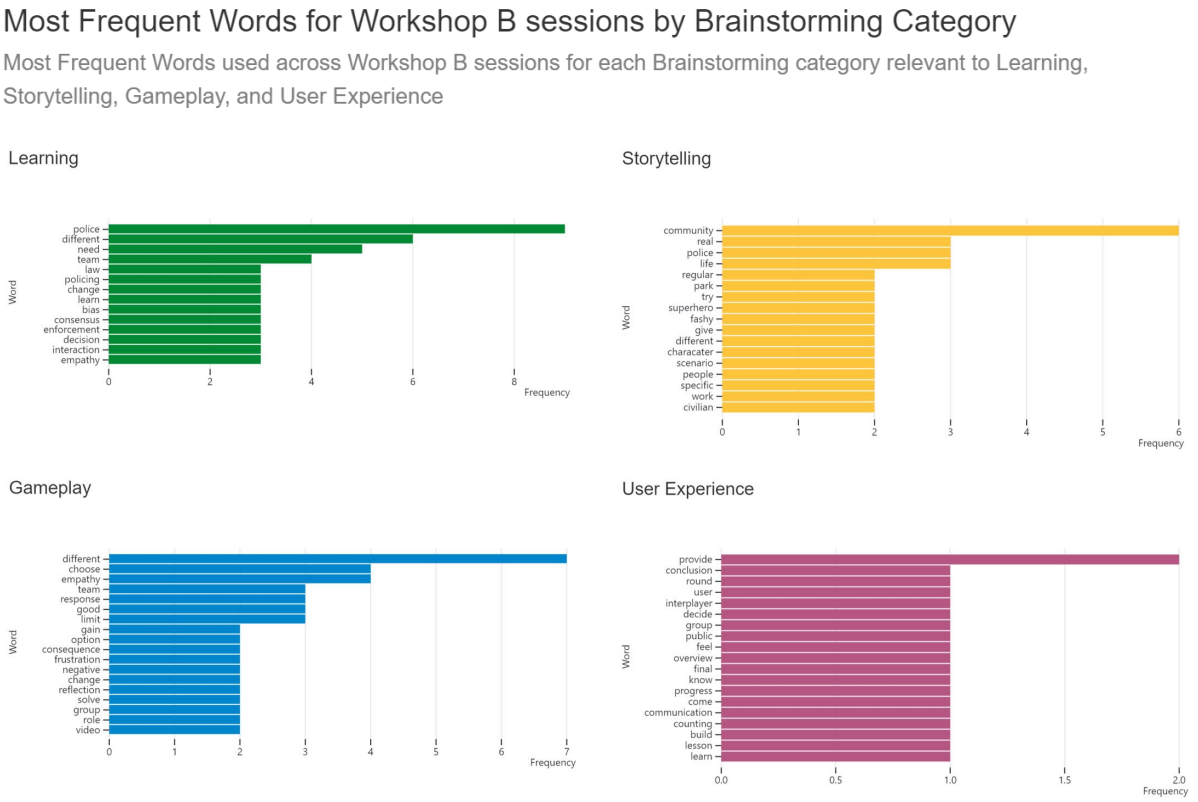


Figure 21: Most frequent words for Workshop B protocol per each Brainstorming category relevant to Learning (top left), Storytelling (top right), Gameplay (bottom left), and User Experience (bottom right)

In observing Figure 21, the use of the words "community" and "police" in the context of brainstorming activities relevant to Learning and Storytelling can be considered. Interestingly, in the context of Learning, the word "community" was not used at all. As such, the word "police" was used 100% more frequently than the word "community". Contrastingly, in the context of Storytelling, the word "community" was used 200% more frequently than the word "police".

Learning

For the Learning brainstorming activity, "police", "different", and "need" were the most commonly used words. Given the supporting observation notes from the SGD activities, it can be observed that participants in Workshop B believed that learning goals and challenges of a serious game design to address racial bias in policing should be catered towards police. Many participants echoed that police often have difficulty empathizing with the public, as their job has conditioned them to act out of self defense. As such, many participants believed that police need to unlearn a militant-like mentality to policing and adopt a more social approach to the profession. Additional conversation relevant to Learning was centered around police officers better understanding the historical context and role of policing in the US and the basics of police interactions. In addition, all participants strongly believed that police need to learn to take on different perspectives and challenge their own points of view by considering those of community members. As such, the word "different" was frequently used in the context of Workshop B Learning when discussing a need for police officers to explore the notion that different circumstances and personal presentation (i.e., appearance) result in different day-to-day experiences of people. Therefore, police officers "need" to not only learn to take on different perspectives, but they also need to reckon with the fact that not all police interactions are the same as a result of different peoples lived experience.

Storytelling

For the Storytelling brainstorming activity, "community", "real", "police" and "life" were the most commonly used words. In using the word "community" during Workshop B Storytelling, participants heavily emphasized the stances and perspectives of community members when addressing Storytelling elements in their game designs. They designed serious games with plots that heavily focused on members within the community as main characters. All participants agreed that the game should reflect the real experience of the black community, however different participants decided to approach such task with different storylines. In discussing how "real" the storylines should be to that of our world, some groups within Workshop B decided to take fantasy-like approaches where the black experience was woven into the game's story through parallels to the real world. Other groups within Workshop B took a more "real world" approach when designing their serious game, and their game stories were more representative of real life. Thus, in a similar context the word "life" was often used when participants were creating their game storylines and drawing comparisons to the present day realities of racial bias in policing. As such, participants decided to incorporate "police" elements within the storyline to varying degrees, depending on how abstract their game designs were. While some participants used the word "police" frequently during this SGD activity and actively included police as characters within their game design, other participants used abstract ideas of police, such as watchmen or overlords, when incorporating a police-like component into the games' stories.

Gameplay

For the Gameplay brainstorming activity, "different", "choose", and "empathy" were the most commonly used words. With regard to the use of the word "different" in Workshop B Gameplay, all participants spoke of role reversal in gameplay, where police officers would play their serious game designs as community members in order to develop different perspectives to police interactions. Additionally, some groups within Workshop B believed that different levels could be a useful means of ensuring that taught content was mastered before players could progress in gameplay. However, most groups preferred to design level-less games where the level of the game difficulty was not determined on demonstrated skill, but instead on random chance that would be based on different character demographics that the player was randomly assigned. Similar to real life, different people experience different levels of difficulty in living their day-to-day lives on the basis of their demographic characteristics, and not necessarily on the basis of the decisions that they choose. In this context, the word "choose" was frequently used when participants discussed player choices in gameplay. While most groups believed that the player should not be able to choose their game character, they did believe that the player should have the option to make their own choices and choose different actions in the game world. As a result of such choses actions, participants believed that different outcomes should result. However, participants also believed that these outcomes would be based on not only the action of the player, but also the demographics of the players' character, where some players would have innate advantages over other players on the very basis of their player demographics. Finally, when discussing the word "empathy" in a gameplay context, all Workshop B participants believed that feelings of empathy should be the end goal for their serious game designs. Similarly, however, participants also believed that before players should feel such emotion, they should first experience frustration, anger, and indignation beforehand in gameplay in order to elicit the emotions that many community members experience on a regular basis when confronted with racial bias in policing.

User Experience

Notably, almost all Workshop B sessions struggled to complete the User Experience ideation cards with the allotted available time. However, for the groups that did make it to this brainstorming activity, "provide" was the most commonly used word. In the context of User Experience, some participants believed that some sort of summary report of "lessons learned" should be provided at the end of the serious game sessions. Other participants believed that providing individual players' statistics could be polarizing, as some players may be hesitant to face their own racial biases in such a direct manner.

5.2 Sentiment Analysis Results

As described in section 3.3.3.2, text collected via categorized sticky notes was analyzed with sentiment analysis to determine the positive, negative, neutral, and normalized compound sentiment scores. Results were generated for all workshop sessions, Workshop A sessions only, and Workshop B sessions only at both the aggregated and disaggregated brainstorming levels (i.e., for the total brainstorming activity vs. for each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience).

5.2.1 All Workshop Sessions

Figure 22 highlights a histogram of the total sentiment measured across all workshop sessions (i.e., both Workshop A sessions and Workshop B sessions) at the aggregated brainstorming level (i.e., for all brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience). In creating Figure 22, a sample size of N = 240 sticky notes was used. As indicated by this figure, the peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.2. Figure 22 also indicates that the spread of the sticky notes is from a compound sentiment score of -1.0 to 1.0. In terms of symmetry, the data appear to be distributed approximately normally, and visible outliers are not present.

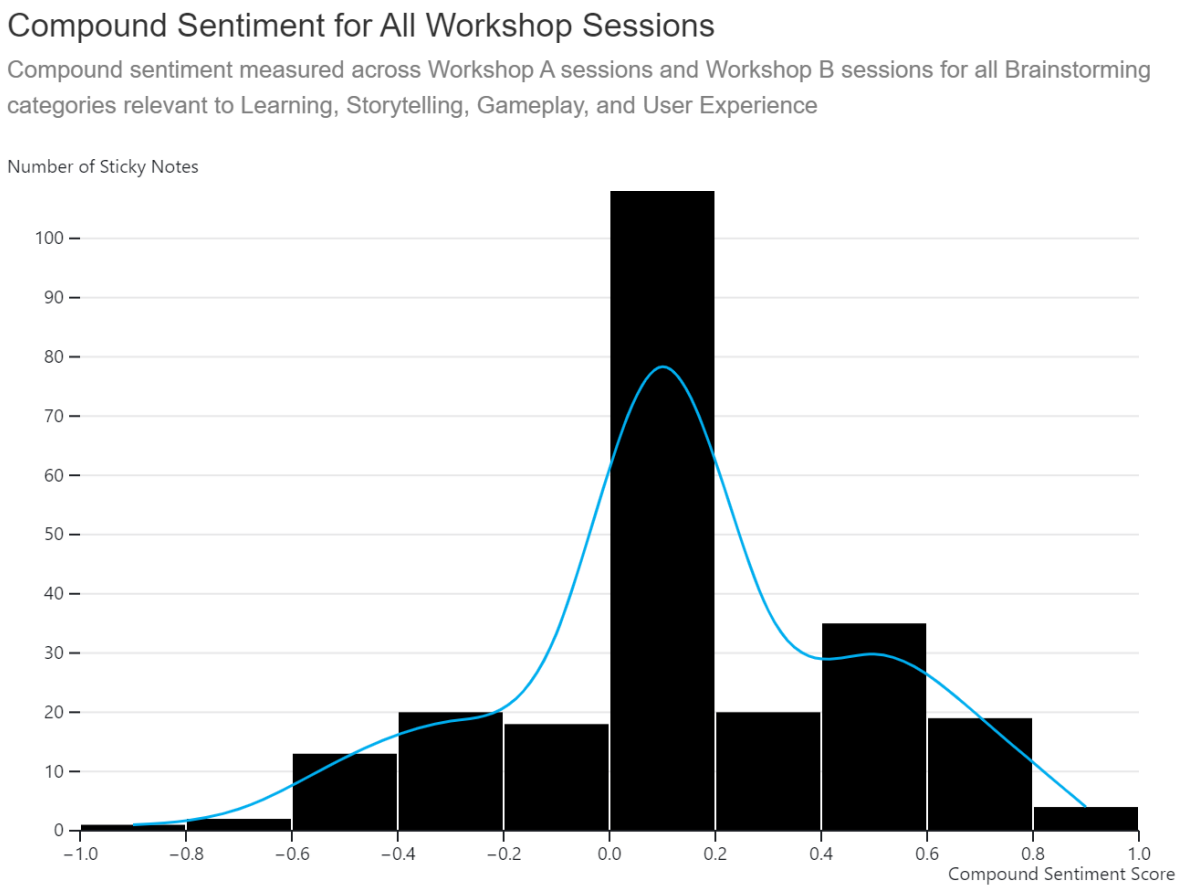


Figure 22: Compound Sentiment of sticky notes for all workshops and all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience

Similarly, a bar chart in Figure 23 outlines the total sentiment measured across all workshop sessions (i.e., both Workshop A sessions and Workshop B sessions) at the aggregated brainstorming level (i.e., for all brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience) as either neutral, positive, or negative. As indicated by this figure, it can be observed that in total and at the aggregated level, 101 sticky notes had a neutral sentiment, 85 sticky notes had a positive sentiment, and 54 sticky notes had a negative sentiment with them.

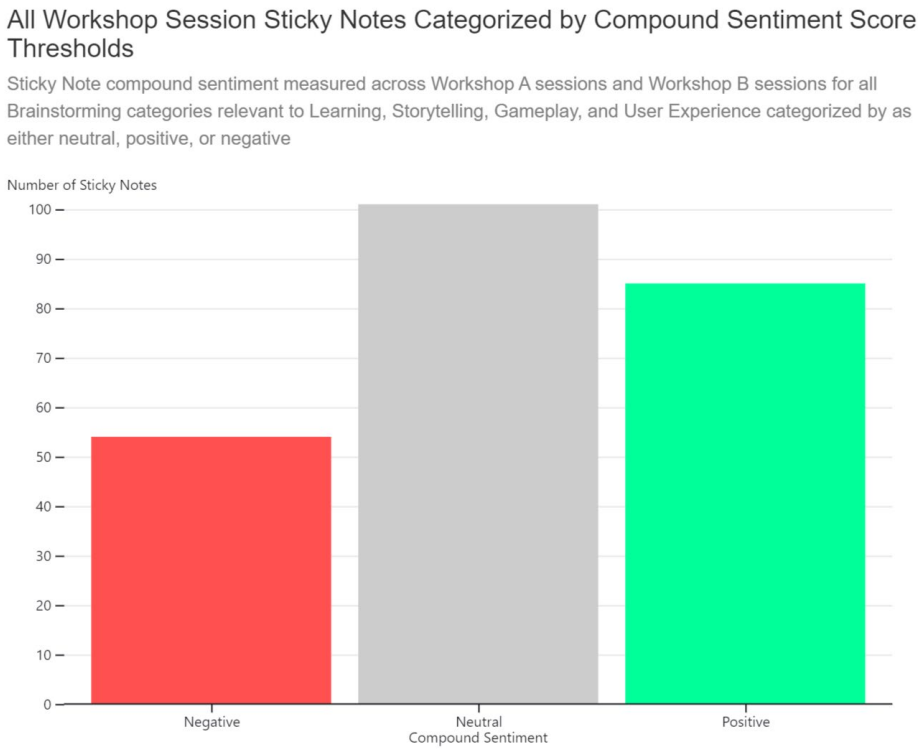


Figure 23: All workshop session sticky notes for all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience categorized by compound sentiment score thresholds where positive sentiment is indicated by a compound score >= 0.05, neutral sentiment is indicated by a compound score > -0.05 and a compound score < 0.05, and negative sentiment is indicated by a compound score < -0.05

In disaggregating the brainstorming activities, Table 9 indicates the average positive, negative, neutral, and normalized compound sentiment scores for all workshop sessions per each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience. Notably, the average compound sentiment score for brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience measure at approximately 0.129, 0.00368, 0.0901, and 0.305 respectively.

Table 9: Compound (highlighted), positive, negative, and neutral sentiment of sticky notes for all workshops and all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience

Brainstorming Category	Compound Sentiment	Positive Sentiment	Negative Sentiment	Neutral Sentiment
Learning	0.128985	0.109456	0.065101	0.825456
Storytelling	0.003679	0.038702	0.048596	0.912667
Gameplay	0.090063	0.198274	0.156952	0.644750
User Experience	0.304635	0.337550	0.018150	0.644300

More specifically, Figure 24 also outlines the histograms of the compound sentiment measured across all workshop sessions (i.e., both Workshop A sessions and Workshop B sessions) at the disaggregated brainstorming level (i.e., for each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience). In creating this figure, a sample size of N = 79 sticky notes was utilized for the Learning brainstorming activity. As indicated by Figure 24, the peak number of Learning sticky notes had the most common compound sentiment score between values of 0.0 and 0.2. The spread of the

Learning sticky notes was from a compound sentiment score of -0.6 to 1.0. Additionally, in terms of symmetry, the Learning data were distributed with a slight positive skew, however no visible outliers were detected with visual review. For the Storytelling data, a sample size of N = 57 sticky notes was used. The peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.1 for the Storytelling brainstorming category. Additionally, the spread of the Storytelling sticky notes was from a compound sentiment score of -0.8 to 0.6. In terms of symmetry, the Storytelling data were distributed with a slight negative skew, and upon visual review, some outliers appear to possibly be present. With regard to Gameplay, a sample size of N = 84 sticky notes was used. As also indicated by Figure 24 and relevant to Gameplay, the peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.2. The spread of the Gameplay sticky notes was from a compound sentiment score of -1.0 to 1.0. The Gameplay data were distributed with a slight negative skew, and outliers however no visible outliers were detected with visual review. Finally, in considering the User Experience data, a sample size of N = 20 sticky notes was utilized. For User Experience, the data appeared to be bimodal where the largest peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.1, and the second largest peak number of sticky notes had the most common compound sentiment score between values of 0.4 and 0.5. The spread of the User Experience sticky notes was from a compound sentiment score of -0.3 to 0.9. In terms of symmetry, the user Experience data were distributed with a slight positive skew. Upon visual inspection, it also appears that outliers are not present.

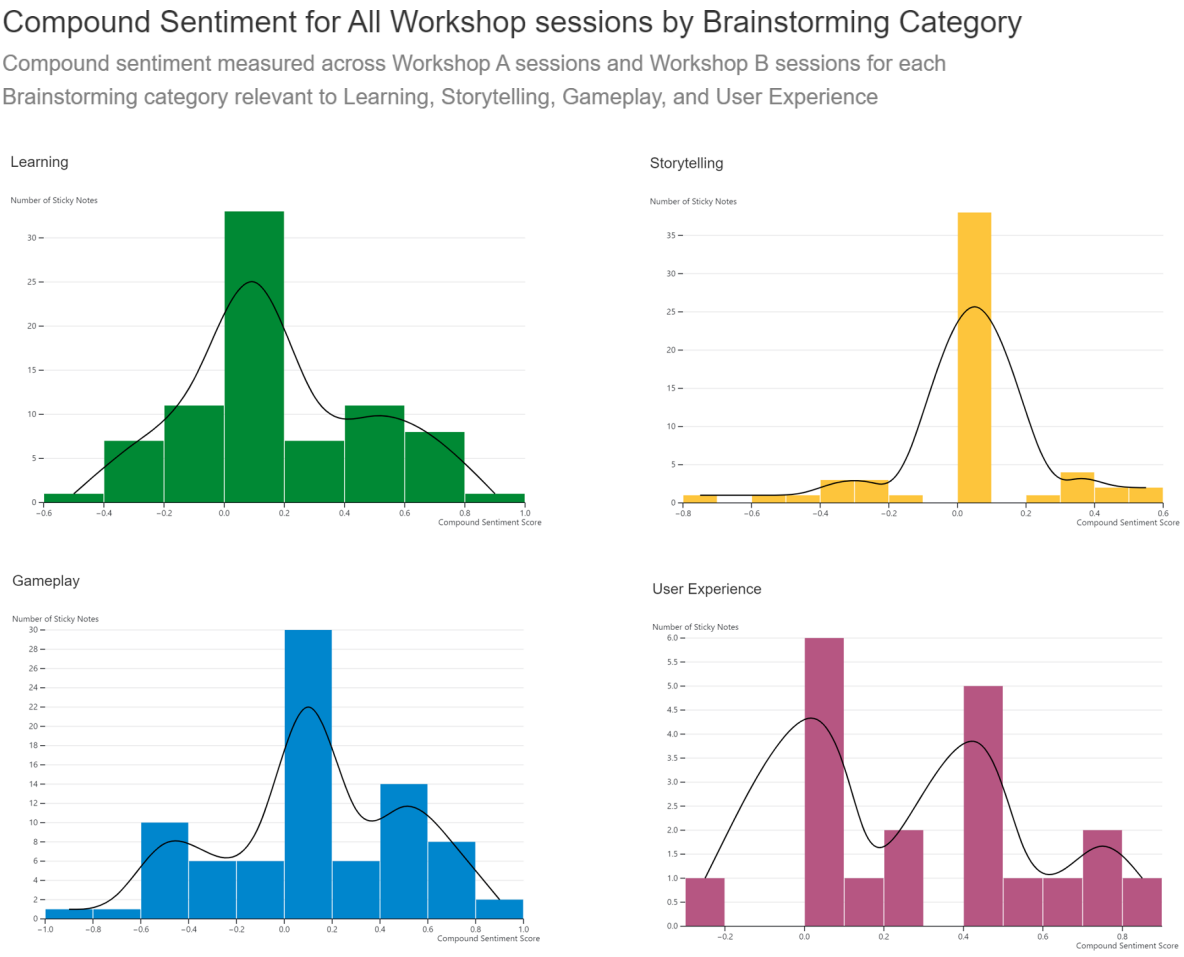


Figure 24: Compound Sentiment of sticky notes for all workshops per each Brainstorming category relevant to Learning, Storytelling, Gameplay, and User Experience

Figure 25 outlines the total sentiment measured across all workshop sessions (i.e., both Workshop A sessions and Workshop B sessions) at the disaggregated brainstorming level (i.e., for each brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience) as either neutral, positive, or negative. In this figure it can be observed that only the Gameplay and User Experience brainstorming activities resulted in more positive sentiment sticky notes than neutral sentiment sticky notes when comparing across all brainstorming activities. However, it can also be observed that across all brainstorming activities, more positive sentiment sticky notes were produced than negative sentiment sticky notes for all Workshop A and Workshop B sessions. Overall, while the majority of sticky notes related to Gameplay and User Experience had positive sentiments associated with them, the majority of sticky notes related to both Learning and Storytelling had a neutral sentiment associated with them.

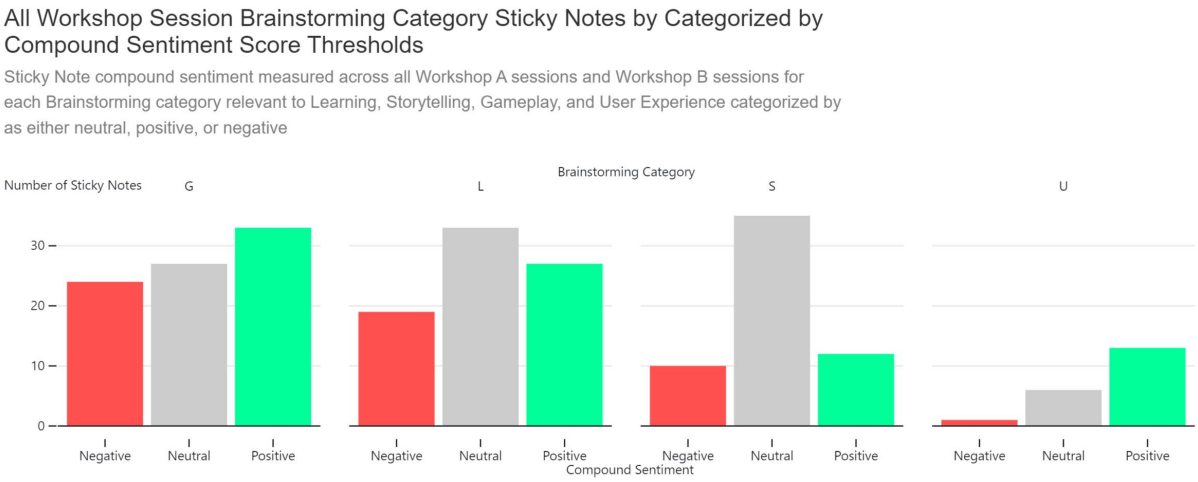


Figure 25: All workshop session sticky notes for each Brainstorming category relevant to Learning (L), Storytelling (S), Gameplay (G), and User Experience (U) categorized by compound sentiment score thresholds where positive sentiment is indicated by a compound score >= 0.05, neutral sentiment is indicated by a compound score > -0.05 and a compound score < 0.05, and negative sentiment is indicated by a compound score < -0.05

5.2.2 Workshop A Sessions

Figure 26 highlights a histogram of the total sentiment measured across all Workshop A sessions at the aggregated brainstorming level (i.e., for all brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience). In creating Figure 26, a sample size of N = 137 sticky notes was used. As indicated by this figure, the peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.2. This figure also indicates that the spread of the sticky notes is from a compound sentiment score of -1.0 to 1.0. In terms of symmetry, the data are distributed with a very slight negative skew, and outliers are not present.

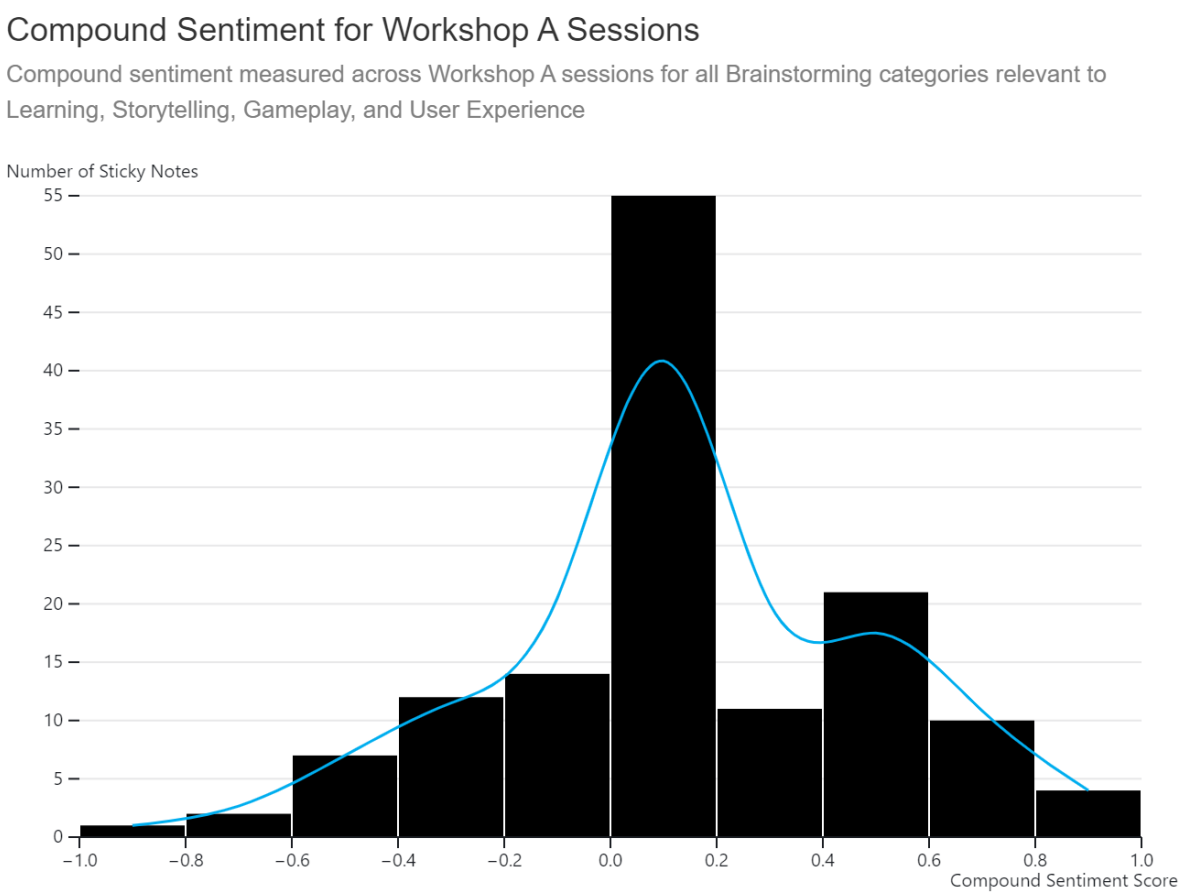


Figure 26: Compound Sentiment of sticky notes for Workshop A and all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience

Similarly, a bar chart in Figure 27 outlines the total sentiment measured across all Workshop A sessions at the aggregated brainstorming level (i.e., for all brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience) as either neutral, positive, or negative. As indicated by this figure, it can be observed that in total and at the aggregated level, 52 sticky notes had a neutral sentiment, 49 sticky notes had a positive sentiment, and 36 sticky notes had a negative sentiment with them. As such, approximately 26.28% of all Workshop B sticky notes had negative sentiment, 37.95% of all Workshop B sticky notes had a neutral sentiment, and 35.77% of all Workshop B sticky notes had a positive sentiment.

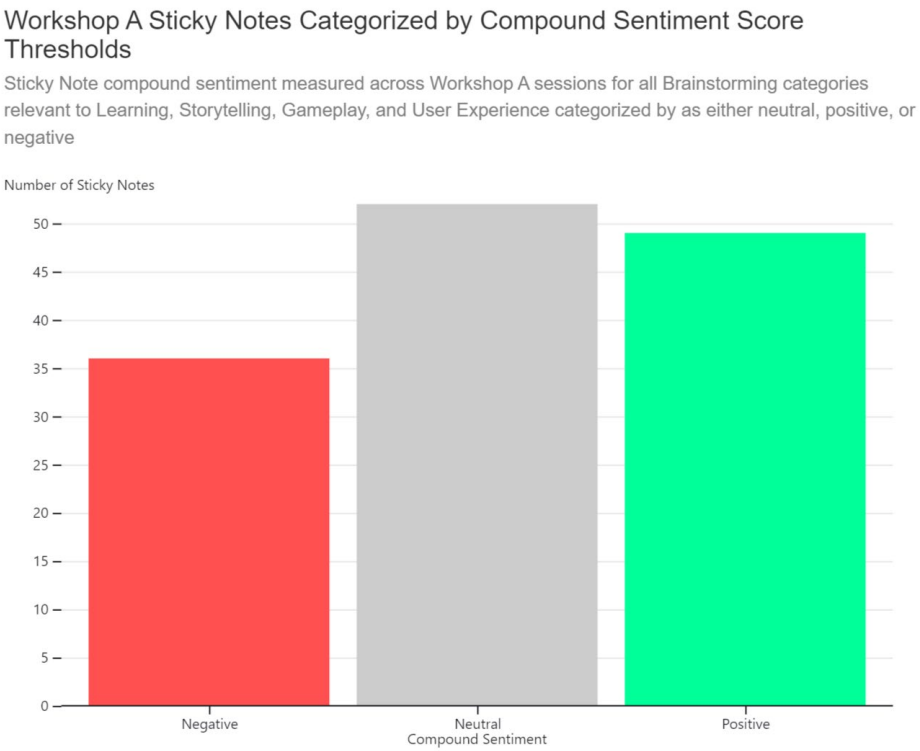


Figure 27: All Workshop A sticky notes for all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience categorized by compound sentiment score thresholds where positive sentiment is indicated by a compound score >= 0.05, neutral sentiment is indicated by a compound score > -0.05 and a compound score < 0.05, and negative sentiment is indicated by a compound score < -0.05

Table 10 indicates the average positive, negative, neutral, and normalized compound sentiment scores for all Workshop A sessions per each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience. Importantly, the average compound sentiment score for brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience measure at approximately 0.127, -0.0460, 0.0812, and 0.406 respectively.

Table 10: Compound (highlighted), positive, negative, and neutral sentiment of sticky notes for Workshop A and all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience

Brainstorming Category	Compound Sentiment	Positive Sentiment	Negative Sentiment	Neutral Sentiment
Learning	0.126589	0.107326	0.089870	0.802826
Storytelling	-0.045965	0.018853	0.067794	0.913324
Gameplay	0.081181	0.208977	0.153884	0.637140
User Experience	0.406486	0.441214	0.025929	0.532857

Figure 28 also outlines the compound sentiment measured across all Workshop A at the disaggregated brainstorming level (i.e., for each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience). Here, a sample size of N = 46 sticky notes was utilized for the Learning brainstorming activity. The peak number of Learning sticky notes had the most common compound sentiment score between values of 0.0 and 0.2. The spread of the Learning sticky notes was from a compound sentiment score of -0.6 to 1.0. Additionally, in terms of symmetry, the Learning data were distributed approximately normally, and outliers were not present. For the Storytelling data, a sample

size of N = 34 sticky notes was used. The peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.1 for the Storytelling brainstorming category. Additionally, the spread of the Storytelling sticky notes was from a compound sentiment score of -0.8 to 0.6. In terms of symmetry, the Storytelling data were distributed with a slight negative skew, and upon visual inspection, some outliers may be present. With regard to Gameplay, a sample size of N = 43 sticky notes was used. As indicated by Figure 28 and relevant to Gameplay, the peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.2. The spread of the Gameplay sticky notes was from a compound sentiment score of -1.0 to 1.0. The Gameplay data were distributed with a slight negative skew, but outliers were not present. Finally, in considering the User Experience data, a sample size of only N = 14 sticky notes was utilized. For User Experience, the peak number of sticky notes had the most common compound sentiment score between values of 0.4 and 0.5. The spread of the User Experience sticky notes was from a compound sentiment score of -0.3 to 0.9. In terms of symmetry, the user Experience data appeared to be distributed with a slight negative skew, and outliers appeared to possibly be present upon visual inspection.

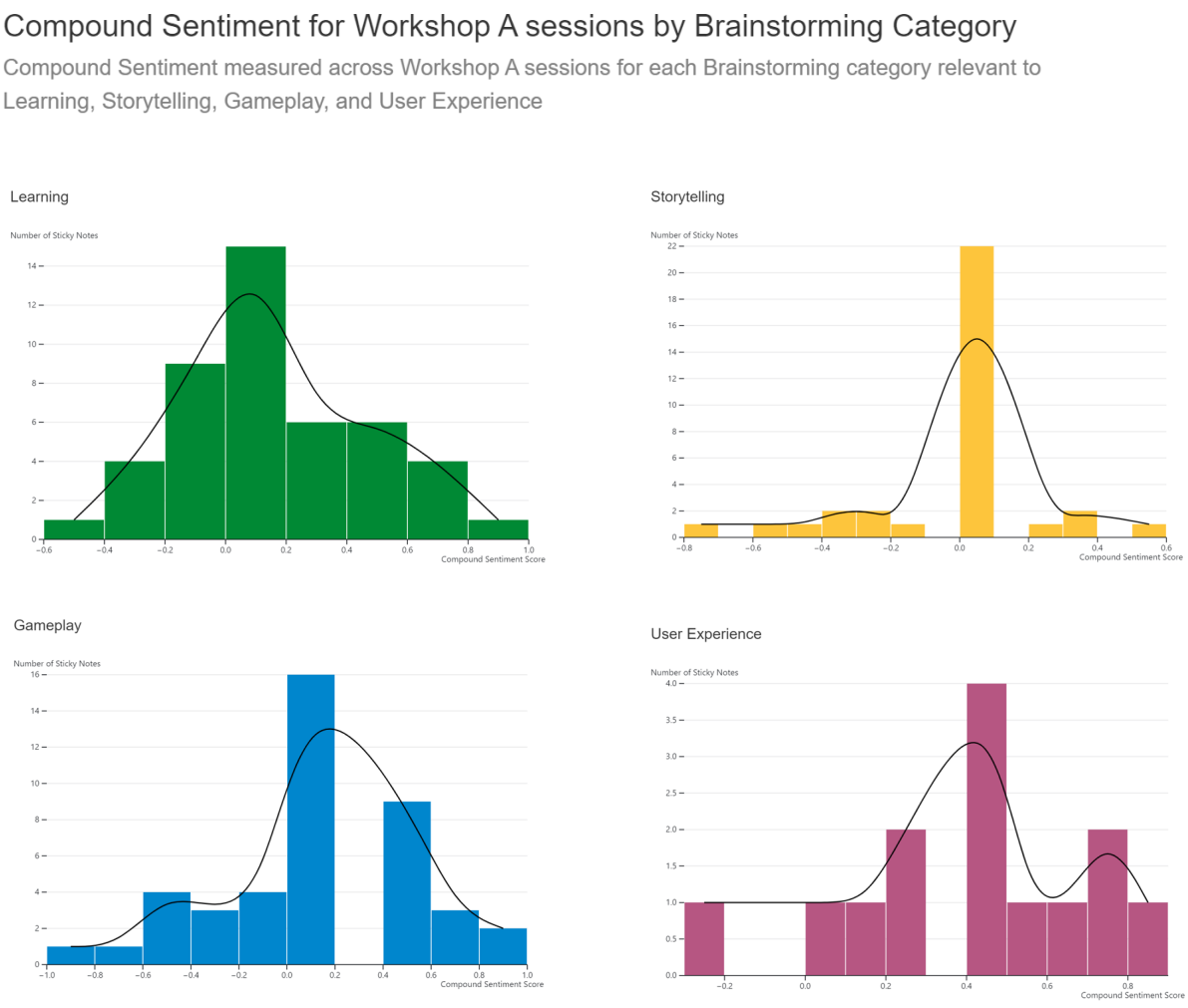


Figure 28: Compound Sentiment of sticky notes for Workshop A per each Brainstorming category relevant to Learning, Storytelling, Gameplay, and User Experience

Figure 29 outlines the total sentiment measured across all Workshop A sessions at the disaggregated brainstorming level (i.e., for each brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience) as either neutral, positive, or negative. In this figure it can be observed that brainstorming activities pertaining to Gameplay, Learning, and User Experience resulted in more positive sentiment sticky notes than neutral sentiment sticky notes and negative sentiment sticky notes when comparing across all brainstorming activities. Similarly, and apart from the User Experience brainstorming activity where Workshop A participants generated an equal number of negative sentiment sticky notes and neutral sentiment sticky notes, all other brainstorming activities resulted in more neutral sentiment sticky notes than negative sentiment sticky notes. Overall, the majority of sticky notes related to Workshop A brainstorming activities relevant to Gameplay, Learning, and User Experience had positive sentiments associated with them, and the majority of sticky notes related to Workshop A Storytelling had neutral sentiment associated with them.

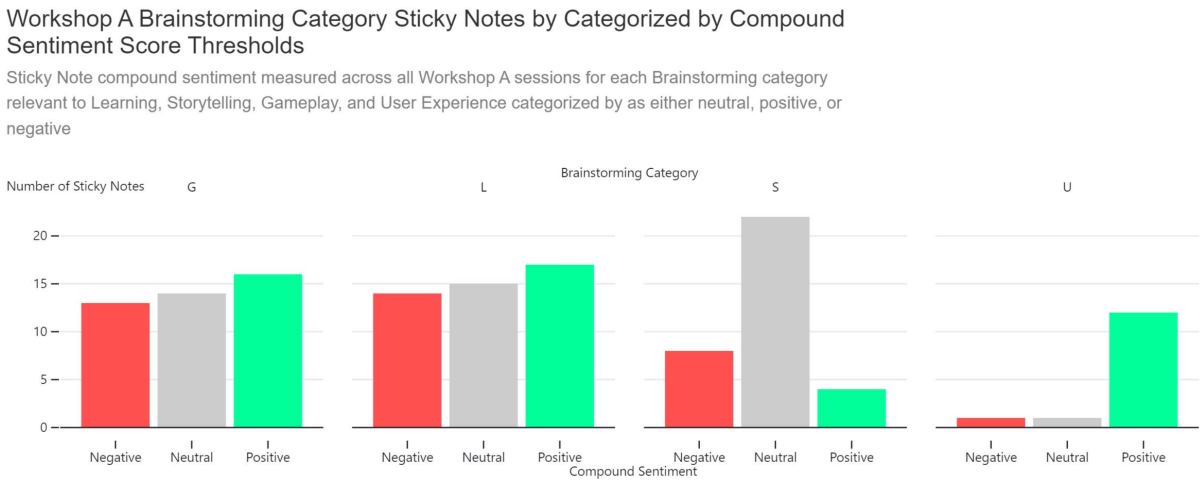


Figure 29: All workshop A session sticky notes for each Brainstorming category relevant to Learning (L), Storytelling (S), Gameplay (G), and User Experience (U) categorized by compound sentiment score thresholds where positive sentiment is indicated by a compound score >= 0.05, neutral sentiment is indicated by a compound score > -0.05 and a compound score < 0.05, and negative sentiment is indicated by a compound score < -0.05

5.2.3 Workshop B Sessions

Figure 30 highlights a histogram of the total sentiment measured across all Workshop B sessions at the aggregated brainstorming level (i.e., for all brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience). In creating Figure 30, a sample size of N = 103 sticky notes was used. As indicated by this figure, the peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.1. The spread of the sticky notes is from a compound sentiment score of -0.6 to 0.8. In terms of symmetry, the data appear to be distributed approximately normally, and outliers are not present.

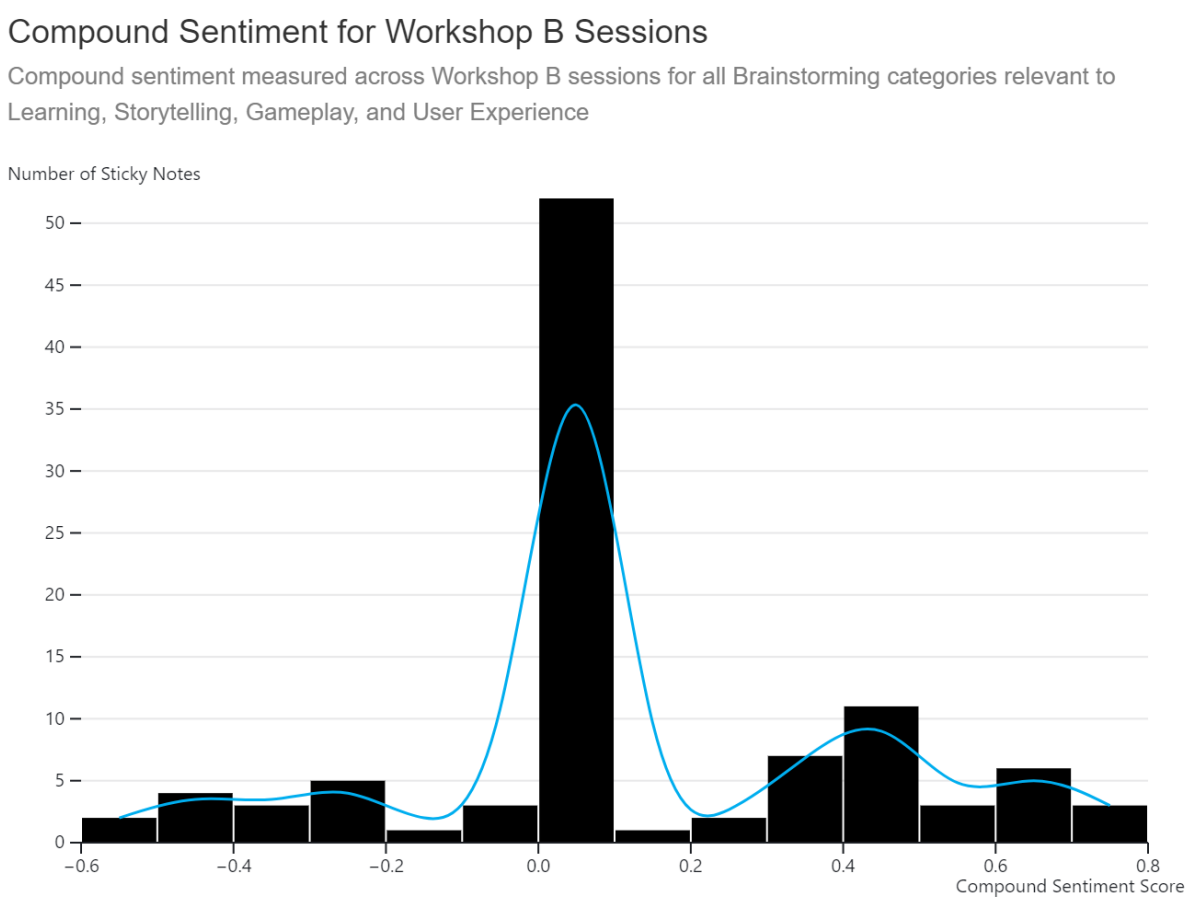


Figure 30: Compound Sentiment of sticky notes for Workshop B and all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience

Similarly, a bar chart in Figure 31 outlines the total sentiment measured across all workshop sessions (i.e., both Workshop A sessions and Workshop B sessions) at the aggregated brainstorming level (i.e., for all brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience) as either neutral, positive, or negative. As indicated by this figure, it can be observed that in total and at the aggregated level, 49 sticky notes had a neutral sentiment, 36 sticky notes had a positive sentiment, and 18 sticky notes had a negative sentiment with them. As such, approximately 17.48% of all Workshop B sticky notes had negative sentiment, 47.57% of all Workshop B sticky notes had a neutral sentiment, and 34.95% of all Workshop B sticky notes had a positive sentiment.

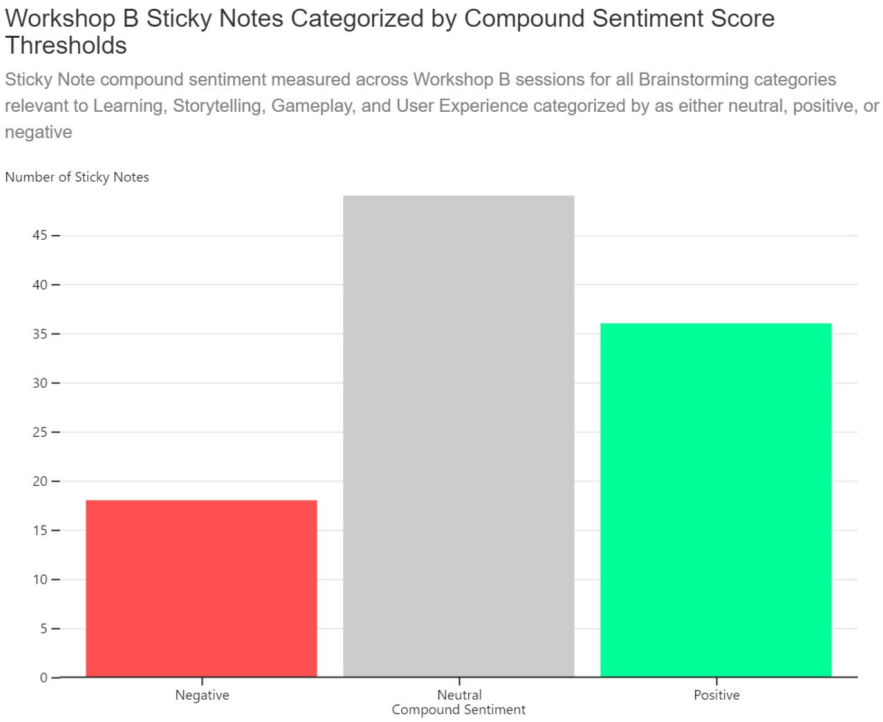


Figure 31: All Workshop A sticky notes for all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience categorized by compound sentiment score thresholds where positive sentiment is indicated by a compound score >= 0.05, neutral sentiment is indicated by a compound score > -0.05 and a compound score < 0.05, and negative sentiment is indicated by a compound score < -0.05

The average positive, negative, neutral, and normalized compound sentiment scores for Workshop B sessions per each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience are highlighted in Table 11. As indicated here, the average compound sentiment score for brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience measure at approximately 0.132, 0.0771, 0.0994, and 0.0661 respectively.

Table 11: Compound (highlighted), positive, negative, and neutral sentiment of sticky notes for Workshop B and all Brainstorming categories relevant to Learning, Storytelling, Gameplay, and User Experience

Brainstorming Category	Compound Sentiment	Positive Sentiment	Negative Sentiment	Neutral Sentiment
Learning	0.132324	0.112424	0.030576	0.857000
Storytelling	0.077065	0.068043	0.020217	0.911696
Gameplay	0.099378	0.187049	0.160171	0.652732
User Experience	0.066983	0.095667	0.000000	0.904333

Figure 32 highlights the compound sentiment measured across all Workshop B at the disaggregated brainstorming level (i.e., for each brainstorming activity relevant to Learning, Storytelling, Gameplay, and User Experience). Here, a sample size of N = 33 sticky notes was utilized for the Learning brainstorming activity. The peak number of Learning sticky notes had the most common compound sentiment score between values of 0.0 and 0.1. The spread of the Learning sticky notes was from a compound sentiment score of -0.4 to 0.8. Additionally, in terms of symmetry, the Learning data

appeared to be distributed with a slight positive, and outliers were not present. For the Storytelling data, a sample size of N = 23 sticky notes was used. The peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.1 for the Storytelling brainstorming category. Additionally, the spread of the Storytelling sticky notes was from a compound sentiment score of -0.4 to 0.6. In terms of symmetry, the Storytelling data were distributed approximately normally, and outliers were not present. With regard to Gameplay, a sample size of N = 41 sticky notes was used. Relevant to Gameplay, the peak number of sticky notes had the most common compound sentiment score between values of 0.0 and 0.1. The spread of the Gameplay sticky notes was from a compound sentiment score of -0.6 to 0.8. The Gameplay data appeared to be distributed approximately normally, and outliers were not present. Finally, in considering the User Experience data, a sample size of only N = 6 sticky notes was utilized. For User Experience, the peak number of sticky notes had the most common compound sentiment score between values of 0.4 and 0.5. The spread of the User Experience sticky notes was from a compound sentiment score of -0.3 to 0.9. In terms of symmetry, the user Experience data were distributed with a slight negative skew, and upon visual inspection, it appeared as if outliers could possibly be present.

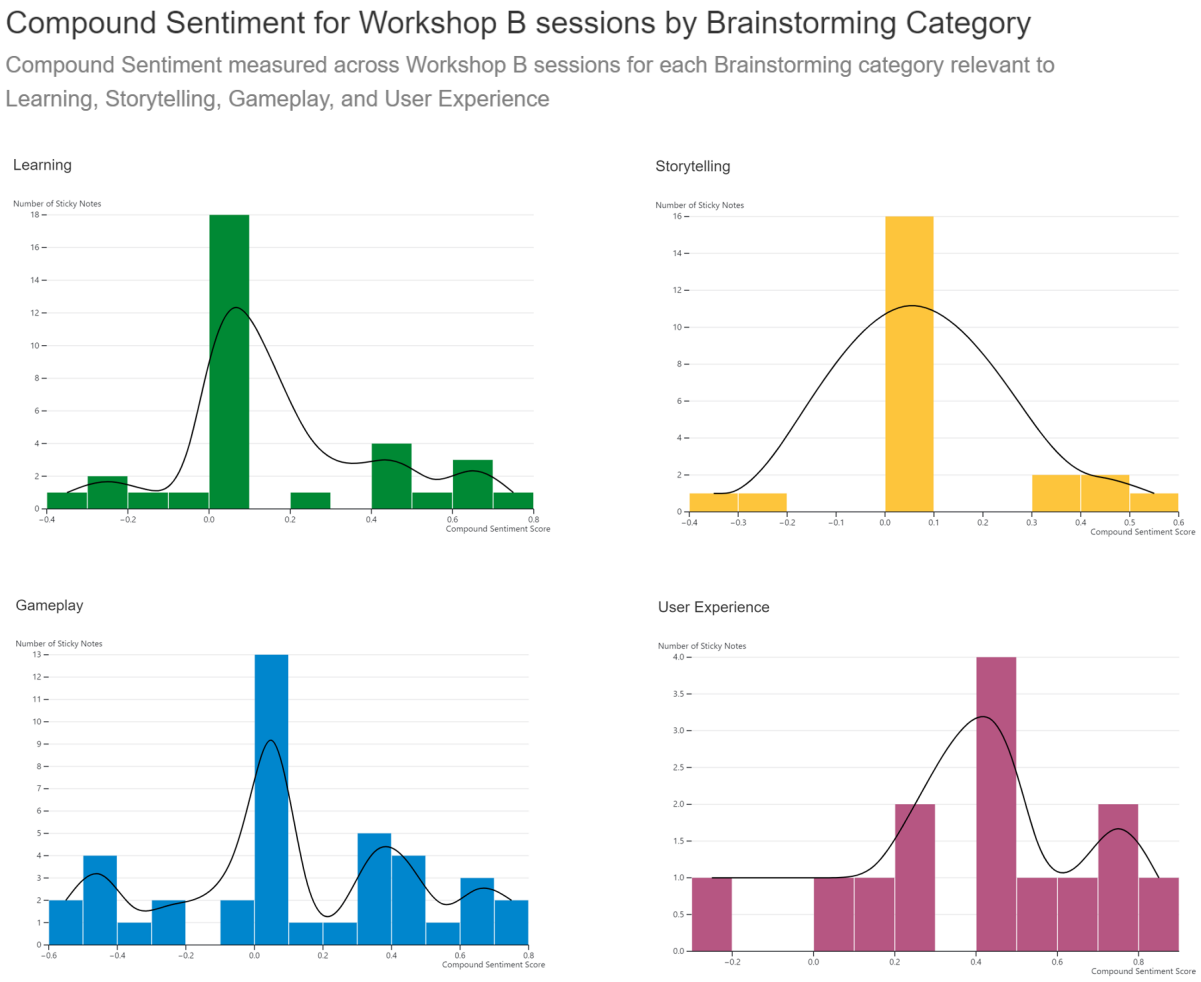


Figure 32: Compound Sentiment of sticky notes for Workshop B per each Brainstorming category relevant to Learning, Storytelling, Gameplay, and User Experience

Figure 33 outlines the total sentiment measured across all Workshop B sessions at the disaggregated brainstorming level (i.e., for each brainstorming activities relevant to Learning, Storytelling, Gameplay, and User Experience) as either neutral, positive, or negative. In this figure it can be observed that brainstorming activities pertaining to only Gameplay resulted in more positive sentiment sticky notes than neutral sentiment sticky notes and negative sentiment sticky notes when comparing across all brainstorming activities. All other brainstorming activities related to Learning, Storytelling, and User Experience resulted in participants generating more neutral sentiment sticky notes than positive sentiment sticky notes and negative sentiment sticky notes. Regardless, for all brainstorming activities, negative sentiment sticky notes were generated the least when compared to the number of neutral sentiment sticky notes and positive sentiment sticky notes. Overall, the majority of sticky notes related to Workshop B brainstorming activities relevant to Gameplay had positive sentiments associated with them, and the majority of sticky notes related to Workshop A Learning, Storytelling, and User Experience had neutral sentiment associated with them.

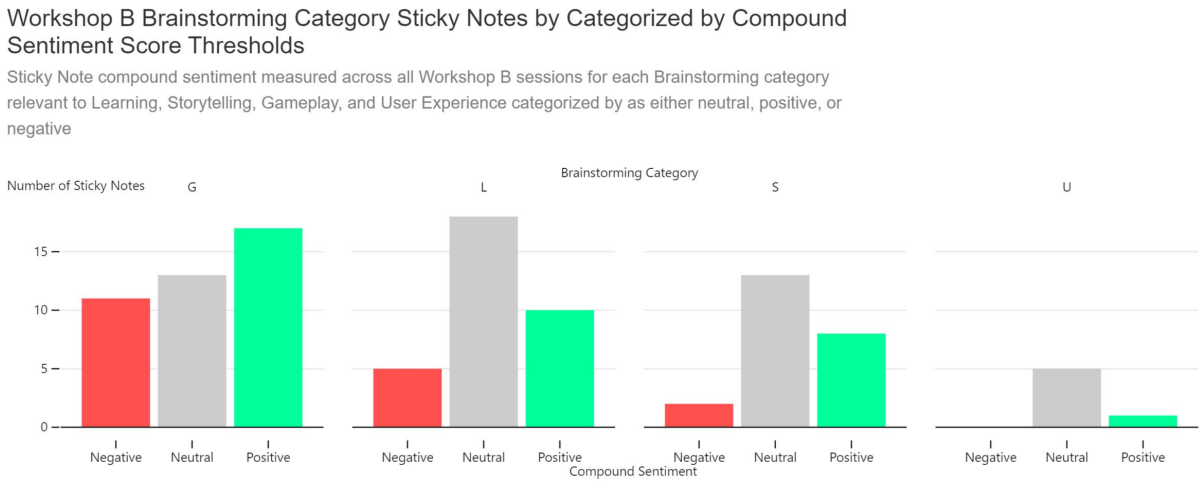


Figure 33: All workshop B session sticky notes for each Brainstorming category relevant to Learning (L), Storytelling (S), Gameplay (G), and User Experience (U) categorized by compound sentiment score thresholds where positive sentiment is indicated by a compound score >= 0.05, neutral sentiment is indicated by a compound score > -0.05 and a compound score < 0.05, and negative sentiment is indicated by a compound score < -0.05

5.3 Micro-Interlocutor Analysis Results

As described by Onwuegbuzie et al. (2009, p. 3), micro-interlocutor analysis "incorporates and analyzes information from group interviews by delineating which participants respond to each question, the order of responses, and the nature of response as well as the nonverbal communication used by each of the focus group participants.". With this analysis technique, the group members can be used as the unit of analysis instead of the group itself (Onwuegbuzie et al., 2009). As such, insight can be generated into how different types of group members respond to the posed questions.

In this study, micro-interlocutor analysis was used on the Observable Table data collected during the evaluative group semi-structured interviews. For both Workshop A and Workshop B protocols, analysis was done at the aggregate level, accounting for all evaluative group semi-structured interview questions, and at the individual question level for interview questions Q1, Q2, Q3, and Q4. In analyzing this data, information on the demographics of respondents, the responses made, the order in which participants responded, the characteristics of response, and the non-verbal communication of participants was considered. At an aggregated level, the overall response rate for the group semi-structured interviews was calculated for each workshop protocol. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for each workshop protocol. Such response rate calculations were also made at a disaggregated level for interview questions Q1, Q2, Q3, and Q4 data. In conducting this analysis, insight into the response rates per different types of group members was generated, and a deeper and more coherent understanding of the response patterns, response characteristics, and response content that was generated for both Workshop A and Workshop B protocols was developed.

5.3.1 Aggregated Semi-Structured Interview Question Comparison: Workshop A vs. Workshop B

5.3.1.1 Workshop A

As indicated by Table 12, Workshop A sessions consisted of a total of 21 participants where 3-4 participants made up an individual session. In total for Workshop A sessions, 1 participant was older, 20 participants were younger, 8 participants were BIPOC, 13 participants were white, 8 participants were visibly female, and 13 participants were visibly male. A breakdown of the responses made by different respondent demographics for all Workshop A sessions can be found in Table 13.

Table 12: Workshop A participant demographic breakdown

Workshop A	
Total Participants	21
Older Participants	1
Younger Participants	20
BIPOC Participants	8
White Participants	13
Visibly Female Participants	8
Visibly Male Participants	13

Table 13: Workshop A participant responses breakdown

Workshop A	All Responses	Responses Question 1	Responses Question 2	Responses Question 3	Responses Question 4
Total Participants	56	14	15	12	15
Older Participants	0	0	0	0	0
Younger Participants	56	14	15	12	15
BIPOC Participants	22	5	7	4	6
White Participants	34	9	8	8	9
Visibly Female Participants	25	6	6	6	6
Visibly Male Participants	31	8	9	6	9

Table 13 illustrates that of the 56 total responses to all questions across all Workshop A sessions, 0 responses were made from older people, while all 56 responses were made from younger people. Of these responders, 22 were BIPOC, and 34 were white. In addition, 25 of these responders were visibly female, and 31 were visibly male. Given that all 21 total participants had an opportunity to respond to all of the four questions posed and only 56 responses were made, the overall group semi-structured interview session had an overall calculated response rate of 67%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for the overall group semi-structured interview session. As outlined in Table 14, 0% of older participants, 70% of younger participants, 69% of BIPOC participants, 65% of white participants, 78% of visibly female participants, and 60% of visibly male participants responded to the overall group semi-structured interview session.

Table 14: Workshop A participant response rates for all evaluative group semi-structure interview questions

Workshop A	Responses	Potential Responses	Response Rate
Total Questions	56	84	0.67
Older Participants	0	4	0.00
Younger Participants	56	80	0.70
BIPOC Participants	22	32	0.69
White Participants	34	52	0.65
Visibly Female Participants	25	32	0.78
Visibly Male Participants	31	52	0.60

5.3.1.2 Workshop B

As indicated by Table 15, Workshop B sessions consisted of a total of 22 participants where 3-4 participants also made up an individual session. In total for Workshop B sessions, 1 participant was older, 21 participants were younger, 12 participants were BIPOC, 10 participants were white, 13 participants were visibly female, and 9 participants were visibly male. A breakdown of the responses made by different respondent demographics for all Workshop B sessions can be found in Table 16.

Table 15: Workshop B participant demographic breakdown

Workshop B	
Total Participants	22
Older Participants	1
Younger Participants	21
BIPOC Participants	12
White Participants	10
Visibly Female Participants	13
Visibly Male Participants	9

Table 16: Workshop B participant responses breakdown

Workshop B	All Responses	Responses Question 1	Responses Question 2	Responses Question 3	Responses Question 4
Total Participants	73	17	18	20	18
Older Participants	2	1	0	1	0
Younger Participants	71	16	18	19	18
BIPOC Participants	38	9	10	9	10
White Participants	35	8	8	11	8
Visibly Female Participants	43	10	10	12	11
Visibly Male Participants	30	7	8	8	7

Table 16 demonstrates that of the 73 total responses to all questions across all Workshop B sessions, 2 responses were made from older people, while 71 responses were made from younger people. Of these responders, 38 were BIPOC, and 35 were white. In addition, 43 of these responders were visibly female, and 30 were visibly male. Given that all 22 total participants had an opportunity to respond to all of the four questions posed and only 73 responses were made, the overall group semi-structured interview session had an overall calculated response rate of 83%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for the overall group semi-structured interview session. As outlined in Table 17, 50% of older participants, 85% of younger participants, 79% of BIPOC participants, 88% of white participants, 83% of visibly female participants, and 83% of visibly male participants responded to the overall group semi-structured interview session.

Table 17: Workshop B participant response rates for all evaluative group semi-structure interview questions

Workshop B	Responses	Potential Responses	Response Rate
Total Questions	73	88	0.83
Older Participants	2	4	0.50
Younger Participants	71	84	0.85
BIPOC Participants	38	48	0.79
White Participants	35	40	0.88
Visibly Female Participants	43	52	0.83
Visibly Male Participants	30	36	0.83

5.3.2 Q1: How do you think the involvement of both community members and law enforcement officers in these workshops could impact the outcome of a serious game concept?

5.3.2.1 Workshop A

In considering Q1, Table 18 illustrates that of the 14 total responses to Q1 across all Workshop A sessions, no responses were made from older people, and responses were made from younger people. Of these responders, 5 were BIPOC, and 9 were white. In addition, 6 of these responders were visibly female, and 8 were visibly male. Given that 21 total participants had an opportunity to respond to Q1 and only 14 responses were made, Q1 had an overall calculated response rate of 67%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for Q1. 0% of older participants, 70% of younger participants, 63% of BIPOC participants, 69% of white participants, 75% of visibly female participants, and 62% of visibly male participants responded to Q1.

Table 18: Workshop A participant response rates for evaluative semi-structured interview Q1

Workshop A	Responses	Potential Responses	Response Rate
Q1	14	21	0.67
Older Participants	0	1	0.00
Younger Participants	14	20	0.70
BIPOC Participants	5	8	0.63
White Participants	9	13	0.69
Visibly Female Participants	6	8	0.75
Visibly Male Participants	8	13	0.62

Interestingly, for all Workshop A working groups, young people responded first to Q1. Similarly, young people also made the last responses to Q1. In responding to Q1, 9 respondents believed that by involving both community members and law enforcement officers/associates in these serious game design workshops, more interesting and creative game concepts could emerge as a result of an expanding scope, various perspectives, different values, and different priorities. All of these 9 respondents were young white people. Contrastingly, 2 respondents believed that in bringing both law enforcement officers/associates and community together to design a serious game concept, a deadlock could emerge resulting in serious game designs that are not representative enough of community ideas and experiences. In responding to this question, one participant was visibly agitated, as they sighed and frowned when delivering their response. Both of these 2 respondents with negative outlooks on game concept outcomes as a result of community and law enforcement SGD collaboration were young BIPOC people. Of the remaining 3 respondents, it was believed that the most interesting serious game design concepts could emerge if law enforcement officers/associates designed their own game separately from community members, who also designed their own game concept separately. These three respondents were made up of young people with both white and BIPOC identities.

5.3.2.2 Workshop B

Table 19 illustrates that of the 17 total responses to Q1 across all Workshop B sessions, one response was made from an older person, and all responses were made from younger people. Of these responders, 9 were BIPOC, and 8 were white. In addition, 10 of these responders were visibly female, and 7 were visibly male. Given that 22 total participants had an opportunity to respond to Q1 and only 17 responses were made, Q1 had an overall calculated response rate of 77%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for Q1. 100% of older participants, 76% of younger participants, 75% of BIPOC participants, 80% of white participants, 77% of visibly female participants, and 78% of visibly male participants responded to Q1.

Table 19: Workshop B participant response rates for evaluative semi-structured interview Q1

Workshop B	Responses	Potential Responses	Response Rate
Q1	17	22	0.77
Older Participants	1	1	1.00
Younger Participants	16	21	0.76
BIPOC Participants	9	12	0.75
White Participants	8	10	0.80
Visibly Female Participants	10	13	0.77
Visibly Male Participants	7	9	0.78

For all Workshop B working groups, young people made both the first and last responses to Q1. In responding to Q1, 9 respondents believed that serious game design concepts to address racial bias in policing could benefit from involving both community members and law enforcement officers/associates in these serious game design workshops, as multiple perspectives could lead to more dynamic serious game concepts. Of these 9 respondents with such optimistic views, 3 people were BIPOC, and the remaining 6 people were white. Contrastingly, 5 participants were sceptical of the added value that bringing both law enforcement officers/associates and community together would have on created serious game design concepts. Many participants expressed the opinion that as a result of the inherent power dynamic between community members and police officers, a serious game design concept designed in this manner will always be skewed towards the perspective of police officers. Of respondents with this pessimistic opinion, 3 people were BIPOC, and 2 people were white. Of the remaining 3 respondents, clear stances of the effectiveness of such workshop design for creating serious game design concepts were not stated. Instead, discussions about diversifying community representation, maintaining small working group sizes, and breaking down power structures were all brought up as potential means to ensure that such workshop setting could produce the most interesting and useful serious game design concept. Respondents with such views were all BIPOC people.

5.3.3 Q2: Do you think that the serious game conceptual design workshop encouraged participants to openly explore aspects of racial bias in policing? Please explain.

5.3.3.1 Workshop A

As outlined in Table 20, of the 15 total responses to Q2 across all Workshop A sessions, no responses were made from older people, and all responses were made from younger people. Of these responders, 7 were BIPOC, and 8 were white. In addition, 6 of these responders were visibly female, and 8 were visibly male. Given that 21 total participants had an opportunity to respond to Q2 and only 15 responses were made, Q2 had an overall calculated response rate of 71%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for Q2. 0% of older participants, 75% of younger participants, 88% of BIPOC participants, 62% of white participants, 75% of visibly female participants, and 69% of visibly male participants responded to Q2.

Table 20: Workshop A participant response rates for evaluative semi-structured interview Q2

Workshop A	Responses	Potential Responses	Response Rate
Q2	15	21	0.71
Older Participants	0	1	0.00
Younger Participants	15	20	0.75
BIPOC Participants	7	8	0.88
White Participants	8	13	0.62
Visibly Female Participants	6	8	0.75
Visibly Male Participants	9	13	0.69

For all Workshop A working groups, young people made both the first and last responses to Q2. Interestingly, all of the young people who provided the final comments on Q1 were also all visibly male. In responding to Q2, 6 respondents believed that the workshop encouraged them to actively explore aspects of racial bias in policing while designing their serious game design concepts. Of these respondents who took such view, 5 respondents had white identities and 1 respondent had a BIPOC identity. Contrastingly, 9 participants believed that in participating in this workshop, their attention was mainly drawn to serious game design concepts, and they did not actively think about racial bias in policing as a result. As put by one participant, "it was easy to get lost in the sauce" of game concepts. Of the participants who did not feel as if the workshop encouraged them to actively explore aspects of racial bias in policing, 6 participants had BIPOC identities, and 1 participant had a white identity.

5.3.3.2 Workshop B

Similarly, Table 21 illustrates that of the 18 total responses to Q2 across all Workshop B sessions, no responses were made from older people, and all responses were made from younger people. Of these responders, 10 were BIPOC, and 8 were white. In addition, 10 of these responders were visibly female, and 8 were visibly male. Given that 22 total participants had an opportunity to respond to Q2 and only 18 responses were made, Q2 had an overall calculated response rate of 82%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for Q2. As outlined by Table 37 0% of older participants, 86% of younger participants, 83% of BIPOC participants, 80% of white participants, 77% of visibly female participants, and 89% of visibly male participants responded to Q2.

Table 21: Workshop B participant response rates for evaluative semi-structured interview Q2

Workshop B	Responses	Potential Responses	Response Rate
Q2	18	22	0.82
Older Participants	0	1	0.00
Younger Participants	18	21	0.86
BIPOC Participants	10	12	0.83
White Participants	8	10	0.80
Visibly Female Participants	10	13	0.77
Visibly Male Participants	8	9	0.89

For all Workshop B working groups, young people made both the first and last responses to Q2. Interestingly, in responding to Q2, 16 respondents believed that the workshop encouraged them to actively explore aspects of racial bias in policing while designing their serious game design concepts. Of these respondents who took such view, 10 respondents were BIPOC and 6 were white. The remaining 2 participants believed that in participating in this workshop, it was difficult to think about racial bias in policing in relation to game concepts. Specifically, these participants found themselves too focused on serious game design, and as a result, they did not actively explore racial bias as it relates to policing. The respondents who struggled to keep racially biased policing at the forefront of their minds while designing their serious game were both white.

5.3.4 Q3 Do you think that serious game design based workshops could be effective in improving police-community relationships and improving overall community satisfaction with police? Why?

5.3.4.1 Workshop A

As outlined in Table 22, of the 12 total responses to Q3 across all Workshop A sessions, no responses were made from older people, and all responses were made from younger people. Of these responders, 4 were BIPOC, and 8 were white. In addition, 6 of these responders were visibly female, and 6 were visibly male. Given that 21 total participants had an opportunity to respond to Q3 and only 15 responses were made, Q3 had an overall calculated response rate of 57%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for Q3. As outlined in Table 38, 0% of older participants, 60% of younger participants, 50% of BIPOC participants, 62% of white participants, 75% of visibly female participants, and 46% of visibly male participants responded to Q2.

Table 22: Workshop A participant response rates for evaluative semi-structured interview Q3

Workshop A	Responses	Potential Responses	Response Rate
Q3	12	21	0.57
Older Participants	0	1	0.00
Younger Participants	12	20	0.60
BIPOC Participants	4	8	0.50
White Participants	8	13	0.62
Visibly Female Participants	6	8	0.75
Visibly Male Participants	6	13	0.46

For all Workshop A working groups, young people made both the first and last responses to Q3. Interestingly, of the young participants who were first to respond to Q3, all were white. In responding to Q3, 8 respondents believed that the serious game based workshop could be an effective tool in improving police-community relationship and overall community satisfaction with the police. 6 respondents with this opinion were white while 2 respondents with this opinion were BIPOC. Contrastingly, 3 respondents believed that SGD based interventions may not be the most effective means of creating dialogue, and they found it hard to understand the utility of such intervention type in the context of racially biased policing. Of respondents who held a more sceptical view of SGD based intervention in the context of police-community relations, 2 respondents had BIPOC identities, and 1 respondent was white. Finally, one white participant held the view that such SGD based interventions could either be incredibly positive or create even more conflict between communities and police.

5.3.4.2 Workshop B

Similarly, Table 23 illustrates that of the 20 total responses to Q3 across all Workshop B sessions, 1 response was made from an older person, while all other responses were made from younger people. Of these responders, 9 were BIPOC, and 11 were white. In addition, 12 of these responders were visibly female, and 8 were visibly male. Given that 22 total participants had an opportunity to respond to Q3 and only 20 responses were made, Q3 had an overall calculated response rate of 91%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for Q3. 100% of older participants, 90% of younger participants, 75% of BIPOC participants, 110% of white participants (i.e., the same white participant responded to Q3 more than once), 92% of visibly female participants, and 89% of visibly male participants responded to Q3.

Table 23: Workshop B participant response rates for evaluative semi-structured interview Q3

Workshop B	Responses	Potential Responses	Response Rate
Q3	20	22	0.91
Older Participants	1	1	1.00
Younger Participants	19	21	0.90
BIPOC Participants	9	12	0.75
White Participants	11	10	1.10
Visibly Female Participants	12	13	0.92
Visibly Male Participants	8	9	0.89

For all Workshop B working groups, young people made both the first and last responses to Q3. In responding to Q3, most respondents believed that these serious game based workshops for improving police-community relationship and overall community satisfaction with the police had varying degrees of neutral to positive utility. 7 participants strongly believed that the serious game based workshop could be an effective tool when applied to police-community relations. Of the respondents with this view, all were white. Holding a slightly less optimistic perspective, 4 participants believed that SGD based interventions could be useful in improving police-community relationships only if activities are carried over longer periods of time, so that relationship repair can actuate and so that communities do not interpret the intervention as a "throw away event". In expressing this opinion, these participants also acknowledged that the underlying systemic issues which influence police-community relationships will always be present, thus always influencing the effectiveness of any intervention type, whether it be serious game related or not. Of the 4 participants with this slightly less optimist perspective, 3 were BIPOC and 1 was white. With a more neutral, yet still positive outlook, 1 BIPOC participant believed that such intervention could be effective in improving police-community relations and overall community satisfaction only in instances where recent significant police harm had not yet occurred. This participant believed that SGD intervention could be effective only if the police department involved in the intervention already had a more positive reputation with the community. 6 participants had neutral perspectives with regard to Q3, and they believed that while SGD intervention could be useful, police and communities are at such conflict with one another that getting them to engage fully and effectively in any setting is incredibly difficult. As such, they believed that at this point in time, society is not ready for such interventions. Of the 6 participants with this view, 3 were BIPOC and 3 were white. Finally, when posed with Q3 2 participants had negative outlooks on the utility of SGD based intervention for improving police-community relations and overall community satisfaction with police. These participants believed that basic trust, which currently does not exist, would be needed before this type of intervention could be useful in a police-community context. These participants also found the idea of serious game based solutions as alienating, as they believed that they made light of very serious issues. Of the participants with such view, both BIPOC and white identities were represented.

5.3.5 Q4: How could these serious game conceptual design workshops be improved for better police-community relationship outcomes?

5.3.5.1 Workshop A

Related to Q4, Table 24 illustrates that of the 15 total responses to this question across all Workshop A sessions, no responses were made from older people, and all responses were made from younger people. Of these responders, 6 were BIPOC, and 9 were white. In addition, 6 of these responders were visibly female, and 9 were visibly male. Given that 21 total participants had an opportunity to respond to Q4 and only 15 responses were made, Q4 had an overall calculated response rate of 71%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for Q4. As outlined in Table 40, 0% of older participants, 75% of younger participants, 75% of BIPOC participants, 69% of white participants, 75% of visibly female participants, and 69% of visibly male participants responded to Q4.

Table 24: Workshop A participant response rates for evaluative semi-structured interview Q4

Workshop A	Responses	Potential Responses	Response Rate
Q4	15	21	0.71
Older Participants	0	1	0.00
Younger Participants	15	20	0.75
BIPOC Participants	6	8	0.75
White Participants	9	13	0.69
Visibly Female Participants	6	8	0.75
Visibly Male Participants	9	13	0.69

For all Workshop A working groups, young people made both the first and last responses to Q4. In responding to Q4, respondents believed the serious game conceptual design workshop could be improved for better police-community relationship outcomes if the intervention was held over a longer duration of time with multiple sessions. In expanding the workshop, participants also believed that a greater emphasis and time could be allotted to providing more learning material on racial bias and on serious games. Finally, participants also believed that conducting such workshops in-person as opposed to online could be more inclusive, as more participants who may not feel competent with online platforms would be able to be involved.

5.3.5.2 Workshop B

Finally, Table 25 illustrates that of the 18 total responses to Q4 across all Workshop B sessions, no responses were made from older people, and all responses were made from younger people. Of these responders, 10 were BIPOC, and 8 were white. In addition, 11 of these responders were visibly female, and 7 were visibly male. Given that 22 total participants had an opportunity to respond to Q4 and only 18 responses were made, Q4 had an overall calculated response rate of 82%. With the total responses made per each demographic group and the number of participants present per each demographic group, the response rate for each demographic group was also calculated for Q4. 0% of older participants, 86% of younger participants, 83% of BIPOC participants, 80% of white participants, 85% of visibly female participants, and 78% of visibly male participants responded to Q4.

Table 25: Workshop B participant response rates for evaluative semi-structured interview Q4

Workshop B	Responses	Potential Responses	Response Rate
Q4	18	22	0.82
Older Participants	0	1	0.00
Younger Participants	18	21	0.86
BIPOC Participants	10	12	0.83
White Participants	8	10	0.80
Visibly Female Participants	11	13	0.85
Visibly Male Participants	7	9	0.78

For all Workshop B working groups, young people made both the first and last responses to Q4. In responding to Q4, respondents believed the serious game conceptual design workshop could be improved for better police-community relationship outcomes if the intervention was held over a longer duration of time with multiple sessions. In expanding the workshop, participants also believed that more learning material about serious games could be provided. They also believed that it could be interesting to learn about the serious games that participating police officers/associates have been exposed to in their line of work. Participants thought that conducting such workshops in-person as opposed to online could be more inclusive, as more participants who may not feel competent with online platforms would be able to be involved. Finally, participants suspected that in addressing the inherent power dynamic between police and community members early on and ensuring that police went into workshops in a more vulnerable state, more productive workshop outcomes could emerge.

6

DISCUSSION

In an effort to address complex societal problems in a variety of different domains, many actors have turned to participatory methods as a means to incorporate community level knowledge and produce community-oriented solutions while addressing societal concerns (Smith, Felderhof, & Bosch, 2008; Suprun et al., 2018; McCabe & Halog, 2018). Simultaneously, serious games are being used in a variety of different fields to provide risk-free spaces where societal problems can be explored (Abt, 1987). More specifically, research in serious games and serious game design has indicated that games for educational purposes are most successful when participants partake in the design and development of the game instead of gameplay (Gamson, 1971). As such, a participatory approach to serious game design has the potential to serve as a means to exploration for complex societal problems like that of this research's application study: biased policing in the US.

But ***how can a participatory serious game design process for reduced police bias, involving both communities and police, be developed in the United States as a means for societal intervention?***

With biased US policing serving as an application study, a curiosity to explore participatory serious game design as a means for societal intervention motivated this study and led to the research question posed above. More specifically, the objective of this research was to explore the effectiveness of participative serious game design as a societal intervention method for biased US policing. Through a mixed methods approach involving inferential statistical analysis, descriptive statistical analysis, content analysis, sentiment analysis, and micro-interlocutor analysis, workshop data was explored. In doing so, an effort was made to develop a deeper understanding of if and how participatory serious game design can be leveraged effectively as a means to societal intervention in the context of US policing and beyond.

As such, this sixth chapter first discusses the key results obtained in chapter 4 and chapter 5 (section 6.1) as they relate to *SQ5* and *SQ6* (section 6.2). Then, scientific strengths and limitations to this research are described (section 6.3). In considering both the findings of this work and the scientific strengths and limitations presented, direction for future research is also outlined (section 6.3). Finally, implications of this research are then outlined in both academic and socio-political contexts (section 6.4). The research contribution of this work is highlighted, and the relevancy of this work in the context of society and public policy is also brought to light.

6.1 Discussion of Key Results

As a result of both quantitative and qualitative components to this work, several key findings resulted from this research. As indicated in section 4.1, no significant differences in measured KAP construct scores were measured before and after the SGD workshop intervention for individual participants. However, as also indicated by section 4.2 section 4.3, section 5.1, section 5.2, and section 5.3, analysis at the "demographic" level resulted in interesting outcomes pertaining to age group, race, and protocol (i.e., participation in either Workshop A or Workshop B).

6.1.1 Age as a Factor to Participatory SGD Workshop Effectiveness

Firstly, with regard to age, it was observed that, overall, younger people displayed higher levels of engagement in the participatory SGD workshops than older people. Despite this, however, participants aged between 25-34 years old had a difficult time benefiting from the intervention. This was demonstrated in section 4.2.1.2 by the fact that this age group had a statistically different mean score for the "Knowledge about Policing in Communities" construct when compared to the 18-24 year old participants and the 35-44 year old participants. Surprisingly, 25-34 year old participants on average, scored worst on their "Knowledge about Policing in Communities", while all other age groups scored better on their "Knowledge about Policing in Communities" post intervention. Similarly, in considering

the perceived effectiveness of the participatory SGD workshop intervention in section 4.3.3, 25-34 year old participants evaluated the intervention with a significantly lower average score than participants aged between 18-24 years old and 35-44 years old. While the incredibly small sample sizes for the 18-24 year old age group (N=11) and the 35-44 years old age group (N=3) likely influenced the results obtained, something may still be able to be said with regard to the ability to engage participants in the 25-34 year old age group. While the participatory SGD workshops may not have necessarily been effective at engaging participants aged between 18-24 years old and 35-44 years old, they were most definitely ineffective, in comparison, at evoking optimistic engagement amongst 25-34 year old participants. While generational differences in optimism, pessimism, and scepticism could have been at play in obtaining these results, there is not a substantial enough amount of scientific literature to support the interpretation of these findings. As such, interpreting these results accurately is beyond the scope of this project.

6.1.2 Race as a Factor to Participatory SGD Workshop Effectiveness

Secondly, in considering racial identities, it was observed that white people were far more optimistic about the participatory SGD workshop intervention than BIPOC people. As demonstrated in section 5.3.4, when asked if the participatory SGD workshop intervention could be useful for improving police-community relations, white respondents were on average more optimistic about the intervention's utility than BIPOC respondents for both Workshop A and Workshop B protocols. These results were also corroborated by evaluations of the perceived effectiveness of the participatory SGD workshop in section 4.3.4, where white people, on average, scored the "overall effectiveness" of the intervention higher than BIPOC people on average. These results could be interpreted to be in accordance with existing research on white fragility (i.e., where white people are intolerant of engaging in racial discomfort) and existing research on serious gaming. Literature within the realm of white fragility and racism demonstrates that, often times, white people find it incredibly difficult to engage in dialogue about racism (DiAngelo, 2018; Unzueta & Lowery, 2008; Augoustinos & Every, 2007; Van Dijk, 1992). Simultaneously, literature related to serious games has indicated that serious games provide a risk-free space, abstracted from reality, where complex societal problems can be explored in a non-threatening nor confrontational setting (Abt, 1987). As such, results from this study could corroborate existing literature on the value of serious games. While white people have a difficult time engaging in race-related dialogue, when asked to do so through a serious game related means, white participants, on average, perceived the experience as "very useful" when compared to BIPOC participants. As such, it could be interpreted in this study that the serious game space provided a level of abstraction from reality that assisted white people in exploring an otherwise difficult topic.

6.1.3 Protocol as a Factor to Participatory SGD Workshop Effectiveness

As demonstrated by section 5.3, for both Workshop A protocol and Workshop B protocol, aside from the fact that younger people in general engaged more in overall discussion and semi-structured group interview responses, results from this study also indicated that no specific demographic group dominated workshop dialogue. As such, both workshop protocols may have been effective at fostering a safe and inclusive space where all participants felt able to engage and contribute to discussion, regardless of their backgrounds. Regardless, however, results did indicate that while both workshop protocols encouraged participants to engage in meaningful dialogue, Workshop B protocol was more effective at doing so. Particularly, Workshop B protocol resulted in higher response rates and more interesting discourse to the semi-structured group interview question at both the aggregated and individual question levels for all demographic groups presented. In considering these results in relation to existing literature around group dynamics, it could be interpreted that the existence of diverse participants with conflicting perspectives in the Workshop B protocol could have fueled more dynamic, creative, and novel ideas. According to Troyer & Youngreen (2009), when properly managed, conflict can play an important role in generating innovative and creative solutions to complex problems. Similarly, studies by Nemeth (1986) and Nemeth & Wachtler (1983) suggest that conflict is also critical in avoiding groupthink (i.e., "the tendency for groups to emphasize consensus over conflict" (Troyer & Youngreen, 2009, p. 410)), as novel solutions to problem-solving can emerge when opinions and ideas that are contrary to those held by the group majority are introduced (as cited in Troyer & Youngreen, 2009). Warr & O'Neill (2006), de Bono (1971), and Lawson (1980) have also cited that the more divergent a group's thinking, the more creative the group is. As such, it could be possible that Workshop A protocol was less effective at engaging participants in more frequent and interesting conversation as a result of the participants' conforming identities (i.e., all community members). Similarly, it is possible that the Workshop B protocol was more successful in eliciting creativity and novel ideas as a result of

the minor and inevitable conflict associated with engaging two different groups with somewhat opposite stances on a controversial issue (i.e., community members and law enforcement officers/associates).

Specific to section 5.3.4 and when asked if the participatory SGD workshop intervention could be useful for improving police-community relations, Workshop protocol B respondents, with both BIPOC and white identities, were on average more optimistic about the intervention's utility than Workshop A protocol respondents of both BIPOC and white identities. While Workshop B protocol respondents that were white found the intervention more useful on average than Workshop B protocol respondents that were BIPOC, BIPOC respondents who participated in the Workshop B protocol, on average, still found some utility in the participatory SGD workshop intervention. Contrastingly, while many white people who participated in the Workshop A protocol were optimistic about the participatory SGD workshop intervention, many BIPOC people who participated in the Workshop A protocol and responded to semi-structured group interview question Q3 struggled to find utility in the experience. As stated by one participant "real people want real solutions to real problems". In understanding this statement in the context of another participant's comments about "throw-away" events and the black community's disinterest in participating in exercises that appear to be performative and useless in nature, it could be interpreted that some Workshop A protocol participants felt as if the participatory SGD workshop was an unreal and meaningless exercise. Comparatively, Workshop B protocol participants may have interpreted the participatory SGD workshop as more "real" and useful as a result of law enforcement officer/associate presence. As demonstrated by Sinclair & Irani (2013), accountability can lead to trustworthiness, which in turn can result in positive attitudes. In interpreting these results in such context, it is possible that the presence of law enforcement officers/associates signaled police level accountability in the Workshop B protocol. As a result, participants may have been more trustworthy of the workshop, and in turn, they may have developed more positive attitudes about the exercise in comparison to Workshop A protocol participants who did not experience police level accountability in their sessions.

Also interesting and as demonstrated by section 5.3.3, on average, Workshop B protocol participants indicated that the participatory SGD workshop intervention encouraged them to actively explore racial bias in policing. Of the Workshop B protocol participants who responded to this question, 80% of respondents held this belief. Comparatively, Workshop A protocol participants indicated that they found it difficult to openly explore aspects of racial bias in policing while designing their serious game design concepts. For Workshop A protocol participants who responded to this question, only 40% of respondents believed that they either consciously or sub-consciously explored aspects of racial bias in policing while engaging in SGD activities. The remaining 60% of Workshop A protocol respondents believed that it was easy to become lost in the game elements during the intervention, and as a result, they did not focus significant attention towards the topic of racial bias in the context of policing. Interestingly, while studies suggest that pressure can lead to "*choking*", Worthy et al. (2009), Markman et al. (2006), Beilock & Carr (2005), Beilock et al. (2004), and Wine (1971), all suggest that in many instances, when placed under pressure, individuals can also excel and perform better than they would have without pressure. Pressure can alter motivation, and individuals can become fixated on minimizing negative social outcomes (Worthy et al., 2009). As such, it could be possible that Workshop B protocol participants performed better in focusing their attention on "racial bias in policing" as a result of the additional pressure in which the protocol elicited when having participants work in a mixed setting, where not all participants shared the same views, opinions, and beliefs about the incredibly controversial issue at hand.

The contrastingly different experiences with the participatory SGD workshop intervention between Workshop B and Workshop A protocols are also reflected in the SGD brainstorming activities and overall game design concepts that resulted. As indicated by section 5.1 and the content analysis results, Workshop B protocol participants focused on serious game "Learning" content that aimed to have police learn more about internal and systemic issues with US policing as an institution. To achieve such learning goals, the "Storytelling" elements for all Workshop B protocol groups heavily focused on community and community perspective. As such, all Workshop B protocol groups used community members as the main game character while designing their serious game concepts. Police-like components were then secondary to the Workshop B protocol group serious game designs, and these components were often worked into game plots to varying degrees of reality with different levels of abstract nuance. On the other hand, Workshop A protocol participants focused on serious game "Learning" content that aimed to have police learn more about communities. To achieve such learning goals, the "Storytelling" elements for all Workshop A protocol groups heavily focused on the police and police perspective. As such, all Workshop A protocol groups used police officers as the main game

characters while designing their serious game concepts, and these components were often worked into the game plots in a very literal sense.

As a result of these strikingly different serious game designs that emerged between Workshop B and Workshop A protocols, sentiment analysis in section 5.2 revealed that at the aggregated brainstorming level, Workshop A protocol participants produced more positive sentiment sticky notes as a percentage of total sticky notes and more negative sentiment sticky notes as a percentage of total sticky notes when compared to Workshop B protocol participants. Similarly, Workshop B protocol participants produced more neutral sentiment sticky notes as a percentage of total sticky notes when compared with Workshop A protocol. These results could be interpreted to mean that the Workshop A protocol participants developed brainstorming ideas and subsequent serious game design concepts that were more polarized than those of the Workshop B protocol participants. While usually discussed in social media related research, "echo chambers" as described by the Oxford dictionary, are "environments in which a person encounters only beliefs or opinions that coincide with their own, so that their existing views are reinforced, and alternative ideas are not considered" ("echo chamber", 2021). According to Del Vicario et al. (2016), Sunstein (2006), and Zollo (2015), echo chambers, and more generally conversation between like-minded individuals, can negatively influence people's emotions and further enforce group polarization. Contrastingly, social media related research also demonstrates that the more heterogenous a person's network, the more neutral and less polarized their views (Lee & Choi, 2019). In abstracting these findings and relating them back to this research, it could be interpreted that Workshop B protocol participants, who participated in a more heterogenous workshop where both community members and law enforcement officers/associates were present, produced more neutral and less polarized sentiment in their brainstorming as a result of the heterogeneity present in their workshop "network". Similarly, it is possible that the homogenous nature of the Workshop A protocol created an echo chamber-like environment were like-minded people only interacted with other like-minded people, thus contributing to more polarized sentiment in the Workshop A protocol brainstorming session.

6.2 Summary of Key Results as Related to *SQ5* and *SQ6*

In summarizing the key results, it can be observed that age, race, and protocol could have all impacted the effectiveness of the participatory SGD workshop intervention.

Specifically, however, ***what can the results from this study conclude about reducing police bias and improving police-community relations through participatory serious game design-based intervention?***

Also more generally, ***what can the outcomes from this study deduce about participatory serious game design as a means for societal intervention in general?***

In directly addressing both *SQ5* and *SQ6*, the following points can be made:, first in relation to reducing police bias and improving police-community relations through participatory serious game design, and then in relation to participatory serious game design as a means for intervention in general:

Reducing Police Bias & Improving Police-Community Relations Through Participatory Serious Game Design-Based Intervention

1. A mixed participant SGD workshop protocol, where both community members and law enforcement officers/associates are both present (i.e., similar to that of Workshop B), may have the potential to be more effective at reducing police bias in the long-term than a workshop protocol that fails to bring both community and police together. With such workshop set-up, it is possible to engage two "opposing" groups (who often fail to interact meaningfully) in dialogue about a controversial issue in a less discomforting and more abstract way. As put by one participant, "anytime you get law enforcement and community members in the same room, it's a positive thing". In further exposing police officers to actual community members and encouraging relationship building, it is possible to mitigate against potential racial resentment and prejudiced attitudes as a result of exposure. Furthermore, as observed by Workshop B protocol sessions, mixed participant SGD workshops have the potential to result in more nuanced, creative, and less accusational serious game designs which could be more effective in teaching police officers about communities, institutional racism, and their own biases during both game design and potential game play post game development.

2. Regardless of the previous point, current relations between police and BIPOC communities, (specifically black communities) are in a dire state of mistrust. As such, any attempt at intervention between these two groups will likely be received with caution and scepticism. However, in demonstrating that intervention efforts are long-term oriented and not just "throw away" events, more willing engagement could be facilitated and relationships between police and community could be improved.
3. Demonstrated police accountability in intervention efforts can exhibit to communities that such efforts to improve police-community relations are serious. In maintaining police accountability throughout interventions, communities may be more willing to welcome police efforts in engagement and relation building.

Participatory Serious Game Design as a Means for Societal Intervention

1. Participatory serious game design has the potential to be an effective intervention method if it is implemented in a way that appropriately caters to the desired audience. In using participatory serious game design as a means to societal intervention, participant "buy-in" must be facilitate early on to ensure effective engagement. For example, in the instance of this research, the workshop facilitator was white and perceived by many participants to be "too European". As such, many participants perceived the facilitator as too disconnected from the issue of racial-based police bias in the US, thus subsequently effecting their perception of the intervention itself. Therefore, a facilitator who may be more embedded in the culture of the participants that are to engage in the intervention may result in a more effective intervention as a result of better participant "buy-in".

2. Participatory serious game design could be particularly effective as an intervention method when the topic of intervention is taboo, discomforting, or difficult to talk about, as the game-like nature of such intervention can provide an adequate amount of abstraction from reality that helps facilitate less anxiety-inducing dialogue. In a similar vein, the reality abstracted game-like nature of participatory serious game design interventions could also be effective at fostering a safe and inclusive spaces where all participants can feel able to engage and contribute to discussion, regardless of their backgrounds.

3. In ensuring a diversity of participant backgrounds and perspectives, echo chambers, polarity, and groupthink can be avoided in participatory serious game design-based interventions. Similarly, more meaningful intervention outcomes and creative problem solving has the potential to arise if a healthy amount of group conflict and pressure is managed appropriately within participatory serious game design interventions.

6.3 Strengths, Limitations, and Future Research Directions

The scope of this research has posed many strengths and limitations to the work presented. As such, these strengths and limitations, in combination with the findings presented, may provide direction for future research.

Strengths

1. **Mixed-Methods Approach:** This research utilized a mixed-methods approach to understanding the effectiveness of participatory serious game design-based intervention. As such, it provided better depth of understanding to the research problem as quantitative and qualitative results could be corroborated by one another. It also highlighted the shortcomings of the existing methods and experimental design implemented (i.e., less rich quantitative results, a need for larger sample sizes, etc.), thus providing future researchers with critical areas of improvement for future work.

2. **Accounting for Social Desirability:** In accounting for social desirability, this research took a particularly scrutinous approach to understanding the effectiveness of participatory serious game design-based interventions. In accounting for participants' tendencies to give socially desirable responses as opposed to responses that reflect their true sentiments, a more rigorous estimate of this type of intervention's value was determined. As such, the one-shot solution and time-sensitive nature of such complex societal problem was considered, and a truer indicator of its actual worth for implementation was provided. This approach benefits both researchers who look to build upon this work, and law enforcement institutions who look to implement this intervention.

3. **Reusability:** This research offers a framework for participatory serious game design-based intervention studies that leverages Winn's (2009) scientifically relevant SGD method as its underpinnings. With this framework, any intervention topic of interest can be easily substituted for the topic of "racial bias in policing" in order to conduct similar studies about the effectiveness of participatory serious game design intervention in the context of other social issues. As such, the methods and experimental design used in this research can easily be implemented (with improvement) in the context of another complex societal issue of research interest.

Limitations

1. **Bias in Participation:** In and of itself, participatory data is biased, as volunteers usually hold firmer stances and opinions on the topics in question, and they are eager to add input. As such, a limitation to this study is that bias is likely present in the input data, as all participants wilfully volunteered to engage in the participatory serious game design workshops, for they all had vested interests in racial bias in policing. In order to address this limitation, future research could mobilize participant recruitment through more random selection means.

2. **Supporting Survey Validity:** In creating scientifically sound supporting survey materials for this intervention, questions related to "explicit racism" were taken from pre-existing and validated studies. However, many of these validated surveys were quite outdated. The questions presented as metrics of explicit racism could very well be argued as explicitly racist themselves. While only the most appropriate questions were drawn from previous work when crafting surveys for further validity, the outdated selection of existing content in this domain was limiting. Additionally, in designing the supporting surveys, questions related to constructs of "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" all utilized a 5-point Likert Scale. While the use of this scale made the scoring of surveys easier, such scoring method also served as a limitation, as it made it difficult for respondents to reply objectively. For example, "strongly agree" and "somewhat agree" can become easily muddied in interpretation for different participants. As such, future research should dedicate substantially more time towards developing culturally appropriate and valid supporting survey questions that are designed with more objective scoring metrics.

3. **Sample Size:** In conducting this research, an incredibly small sample size was used, thus posing as a limitation to the validity of the results obtained. Future research should look to involve more participants, and these participants should come from a variety of different demographic backgrounds.

4. **Technological Difficulties**: In compensating for the COVID-19 pandemic, all participatory serious game design workshops were held online. As such, computer literacy greatly impacted participants' experiences with the intervention. In future research, participatory serious game design workshops should be held in an offline and in-person setting to ensure a more inclusive intervention environment.

5. **Inadequate Time:**
   Given the time limitation associated with this study, participatory serious game design workshop sessions were limited to a total of two hours. In providing feedback on the intervention, all participants echoed that the time provided was insufficient, and that future interventions should be held over a longer duration of time with multiple sessions. As such, time was a major limitation to this study. Future research should look to implement these participatory serious game design workshops over longer planning periods. In adopting more long-term oriented timing for this research, additional efforts should be made to investigate the long-term effects that this intervention may have on actual police bias, community fear, community satisfaction, and community faith in law enforcement.

6.4 Research Implications

To the best of our knowledge, this research is the first mixed methods study aimed at investigating the effectiveness of participatory serious game design as a societal intervention method for biased US policing. As such, this study adds several potential scientific contributions to the fields of intervention

science and serious game design, and it also has several potential implications in the context of society and policing.

1. This research contributes to existing scientific evidence that serious games have the potential to provide risk-free spaces, abstracted from reality, where complex societal problems, like racial bias in policing, can be explored in a non-threatening nor confrontational setting. More specifically, this research provides new scientific evidence that participatory serious game design has the potential to effectively be used as an intervention method when difficult intervention topics must be addressed.

2. This research adds to existing scientific evidence that suggests if properly managed, the conflict that can arise from diversity in individual backgrounds and perspectives amongst group members can be beneficial in eliciting create problem solving. In addition, this research also supports existing scientific evidence that echo chambers, polarity, and groupthink can be avoided when multiple diverse perspectives are engaged in problem solving tasks.

3. This research provides further proof in the social sphere and to public policy makers that until intervention efforts to address broken relationships between police and communities involve institutional accountability and proof of long-term commitment, law enforcement efforts to engage with the public will likely be received with caution and interpreted as performative.

4. This research also provides US police departments with a participatory serious game design-based intervention protocol that can be used as a means of COP and public engagement. Additionally, this research also provides institutions related to law enforcement and policing, such as *CPE*, with a participatory means to generate more culturally appropriate and community-deemed relevant serious game designs. In holding participatory serious game design workshop sessions and generating serious game concepts, more culturally appropriate serious games can be developed for future police use.

7

CONCLUSION

Through this research, participatory serious game design as a means for societal intervention was investigated with the main research question,

How can a participatory serious game design process for reduced police bias, involving both communities and police, be developed in the United States as a means for societal intervention?

In addressing this main research question, six sub-questions structured this report and guided this study. This final chapter aims to provide a coherent answer to each of the posed sub questions, and in doing so, this chapter also aims to address the main research question to this work.

As demonstrated in section 2.1 and in addressing SQ*1*, "*How can police bias be defined?*, this study refers to bias as an "inclination or prejudice for or against one person or group, especially in a way considered to be unfair" on the basis of race ("Bias", 2021). While such bias can be both implicit and explicit, this study focuses on racial-based bias that is explicit in nature where such biased sentiment is neither automatic nor subconscious. Explicit racial bias was chosen as the primary focus for this research because often times when acts of discrimination are categorized as implicitly motivated, more lenient repercussions result. As such, in order to highlight the severity of bias in the context of US policing, a "harder" stance on the issue has been taken by focusing on purposeful discrimination. Furthermore, explicit bias was also chosen over implicit bias as the primary focus for this research per the suggestions of Axt (2017), which indicates that the best way to quantify someone's racial bias is to simply ask them explicitly. In the context of US policing, police bias is then defined as explicit bias in the context of policing, where purposeful stereotyping is used to understand community dynamics and potential criminal activity. Given the existing limitations identified by Goff et al. (2016) in section 2.1.3 relevant to measuring this type of police bias, this study "measures" police bias qualitatively by observing the content of subsequently designed serious game concepts and their inclusion of purposeful stereotyping to understand community dynamics and potential criminal activity. In doing so, this research recognizes that while unequal treatment of different demographic groups by the police is not always on the basis of explicit racial bias alone, racial resentment can still impact officers' beliefs, attitudes, and subsequent policing tactics and interactions.

SQ2 posed the question, "*Based on theory and practice, what are successful approaches to involving community members in interventions for reduced police bias that can inform a participatory serious game approach to intervention?".* In addressing this sub-question in section 2.2, a review of the literature pointed to COP being more of a philosophy to policing as opposed to a series of prescribed methods. However, successful approach criteria relevant to "Setting the Stage", "Problem Identification", "Systems Analysis", "Alternative Strategy Formulation", "Continuous Improvement", "Multi-Stakeholder Involvement", "Interdisciplinary Approach", and "Long-Term Orientation" were identified in examining the work of Gaines (1993), Cordner (1999), Rau et al. (2020), and COPS (2012).

SQ3 questioned, "*Which serious game design methods are suitable for participatory serious game design processes between communities and police forces in the United States?"* In answering this sub-question, in section 2.3 different SGD methods were analyzed according to their commonalities with effective COP methods. In considering different SGD methods, the work of Duke & Geurts (2004), Kortmann & Harteveld (2009), Fullerton (2008), Winn (2009), and Abeele et al., (2012) were all considered on the basis of "Setting the Stage", "Problem Identification", "Systems Analysis", "Alternative Strategy Formulation", "Continuous Improvement", "Multi-Stakeholder Involvement", "Interdisciplinary Approach", "Storytelling Emphasis", "Design Approach", and "Technology Platform". Results indicated that the SGD methods of Winn (2009) and Abeele et al. (2012) were most suitable for design processes between communities and police officers in the United States. However, Winn's (2009) method was determined to be best suited. This was due to time limitations of this study, the

readily available resources associated the Winn (2009) method, and the lack of readily available resources for the Abeele et al. (2012) method.

Demonstrated in chapter 3 and in addressing *SQ4, "What methods are suitable for measuring the effectiveness of a police-community serious game design process for reduced police bias?",* a mixed methods approach, involving quantitative elements of inferential and descriptive statistical analysis and qualitative elements of content analysis, sentiment analysis, and micro-interlocutor analysis was used to measure the effectiveness of the police-community serious game design process for reduced police bias. A mixed methods approach to the research was chosen because better depth of understanding to the research could be developed. In using inferential statistics on KAP study data, the effectiveness of the intervention for each individual participant could be measured on the basis of measured KAP constructs. In addition, such quantitative analysis made it possible to gauge how participants' different workshop protocols (i.e., Workshop A and Workshop B) and different demographic characteristics relevant to gender, age, and race could have impacted participants' KAP scores and subsequent experiences with the intervention. Descriptive statistics also provided quantitative insight into the participants' perceived overall effectiveness of the participatory SGD workshop. In using qualitative elements of content analysis, words used by the participants could be counted, compared, and interpreted with context in order to understand how participants experienced different workshop protocols. Insight into the meaningfulness of conversation and types of ideas generated in the separate protocol formats could also developed. In then further understanding the frequently used words in the workshop sessions, sentiment analysis could provide greater insight into the types of sentiment associated with the workshop words. By using these two qualitative techniques together, insight into the workshop protocol effectiveness could be generated. Finally, in also incorporating the qualitative technique of micro-interlocutor analysis in this study, insight into the response rates per different types of group members, response patterns, response characteristics, and response content could also be developed and compared across Workshop A and Workshop B protocols in order to understand the effectiveness of each workshop set-up on the basis of engagement and meaningfulness in responses. In considering quantitative methods for measuring the effectiveness of the police-community serious game design process for reduced police bias, measured KAP survey constructs related "Knowledge about Policing in Communities", "Attitude about Policing in Communities", "Attitude about Explicit Racism", and "Practice about Policing in Communities" served as metrics for the effectiveness of the intervention. Furthermore, participants' self-evaluated perceived overall effectiveness ratings of the participatory SGD workshop also provided a quantitative metric as to how effective the intervention was. In considering qualitative methods for measuring the intervention's effectiveness, observed engagement and optimism amongst participants in workshop sessions served as a means for operationalizing "effectiveness". In addition, the meaningfulness in workshop session dialogue and the designed serious game concepts also served as a metric for intervention effectiveness.

In revisiting *SQ5,* "*What can the results from this study conclude about reducing police bias and improving police-community relations through participatory serious game design-based intervention?"*, chapter 6 of this study has indicated many interesting insights. With regard to reducing police bias, results from this research indicate that a mixed participant SGD workshop protocol, where both community members and law enforcement officers/associates are both present (i.e., similar to that of Workshop B), may have the potential to be more effective at reducing police bias in the long-term. With such workshop set-up, it is possible to engage two "opposing" groups in dialogue about a controversial issue in a less discomforting and more abstract way. Additionally, in further exposing police officers to actual community members and encouraging relationship building, it is possible to mitigate against potential racial resentment and prejudiced attitudes as a result of exposure. Furthermore, as observed by Workshop B protocol sessions, mixed participant SGD workshops have the potential to result in more nuanced, creative, and less accusational serious game designs which could be more effective in teaching police officers about communities, institutional racism, and their own biases during both game design and potential game play post game development. With regard to improving police-community relations through participatory serious game design-based societal intervention, results from this research indicate that current relations between police and BIPOC communities, (specifically black communities) are in a dire state of mistrust. As such, any attempt at intervention between these two groups will likely be received with caution and scepticism. However, in demonstrating that intervention efforts are long-term oriented and not just "throw away" events, more willing engagement could be facilitated. In addition, this research has also demonstrated that police accountability in intervention efforts can exhibit to communities that such efforts to improve police-community relations are serious. In maintaining police accountability throughout interventions, communities may be more willing to welcome police efforts in engagement and relation building.

Finally, in addressing *SQ6*, "*What can the outcomes from this study deduce about participatory serious game design as a means for societal intervention in general?*", section 6.2 of this study indicates that participatory game design has the potential to be an effective intervention method if it is implemented in a way that appropriately caters to the desired audience. For example, in the instance of this research, the workshop facilitator was white and perceived by many participants to be "too European". As such, many participants perceived the facilitator as too disconnected from the issue of racial-based police bias in the US, thus subsequently effecting their perception of the intervention itself. Therefore, a facilitator who may be more embedded in the culture of the participants that are to engage in the intervention may result in a more effective intervention as a result of better participant "buy-in". As observed in this research, in using participatory game design as a means to societal intervention, participant "buy-in" must be facilitated early on to ensure effective engagement. This study also alludes to the fact that participatory game design could be particularly effective as an intervention method when the topic of intervention is taboo, discomforting, or difficult to talk about, as the game-like nature of such intervention can provide an adequate amount of abstraction from reality that helps facilitate less anxiety-inducing dialogue. In a similar vein, the reality-abstracted game-like nature of participatory game design interventions could also be effective at fostering safe and inclusive spaces where all participants can feel able to engage and contribute to discussion, regardless of their backgrounds. Finally, with regard to using participatory game design as an intervention study, this research has exhibited that in ensuring a diversity of participant backgrounds and perspectives, echo chambers, polarity, and groupthink can be avoided in participatory game design-based interventions. Similarly, more meaningful intervention outcomes and creative problem solving has the potential to arise if a healthy amount of group conflict and pressure is managed appropriately within participatory game design interventions.

Thus, in addressing the main research question, it is apparent that societal intervention for complex problems, like that of biased US policing, are difficult to implement effectively. Regardless of the inherent challenges that exist, however, long-term oriented planning and institutional commitments to accountability and change can help facilitate participant "buy-in", optimism, meaningful dialogue, and, effective engagement. Furthermore, while both COP initiatives and serious game play have also been used as an intervention method for improved police practices, participatory serious game design between both communities and police forces may provide diverse individuals with multiple perspectives with a safe space to engage in constructive dialogue. As such, in developing interventions where serious games are to be designed participatively between US communities and police for reduced police bias, safety, inclusion, and diversity should be championed.

REFERENCES

Abeele, V. V., De Schutter, B., Geurts, L., Desmet, S., Wauters, J., Husson, J., van den Audenaeren, L., van Broeckhoven, F., Annema, J.H., Geerts. D.P. (2012). A player-centered, iterative, interdisciplinary and integrated framework for serious game design and development. In: *Serious Games* (82-86). Springer, Cham

Abt, C. C. (1987). *Serious Games*. University Press of America.

Agan, A., & Starr, S. (2017). The Effect of Criminal Records on Access to Employment. *American Economic Review, 107*(5), 560-564.

Aggarwal, R., & Ranganathan, P. (2019). Study designs: Part 4 - Intervention studies. *Perspectives in Clinical Research, 10*(3), 137-139. doi: 10.4103/picr.PICR_91_19

Akcaoglu, M. (2014). Learning problem-solving through making games at the game design and learning summer program. *Education Tech Research Dev, 62,* 583-600. DOI 10.1007/s11423-014-9347-4

Akhgar B., Redhead A., Davey S., Saunders J. (2019) Introduction: Serious Games for Law Enforcement Agencies. In: Akhgar B. (eds) Serious Games for Enhancing Law Enforcement Agencies. Security Informatics and Law Enforcement. Springer, Cham. https://doi.org/10.1007/978-3-030-29926-2_1

Ampatzidou, C., & Gugerell, K. (2019). Participatory game prototyping - balancing domain content and playability in a serious game design for the energy transition. *International Journal of CoCreation in Design and the Arts, 15*(4), 345-360. DOI: 10.1080/15710882.2018.1504084

Ampatzidou, C., Gugerell, K., Constantinescu, T., Devisch, O., Jauschneg, M., Berger, M. (2018). All Work and No Play? Facilitating Serious Games and Gamified Applications in Participatory Urban Planning and Governance. *Urban Planning, 3*(1), 34-36. DOI: 10.17645/up.v3i1.1261

Andrade, C., Menon, V., Ameen, S., & Kumar Praharaj, S. (2020). Designing and Conducting Knowledge, Attitude, and Practice Surveys in Psychiatry: Practical Guidance. *Indian Journal of Psychological Medicine, 42*(5), 478–481. https://doi.org/10.1177/0253717620946111

Arumugam, N. (2019). Knowledge, Attitudes and Practices (KAP) Towards Medicinal Plants Among Malaysian Consumers. *Med Aromat Plants (Los Angeles), 8*, 341. doi: 10.35248/2167-0412.19.8.341

Augoustinos, M., & Every, D. (2007). Contemporary racist discourse: Taboos against racism and racist accusations. In *Language, discourse and social psychology* (pp. 233-254). Palgrave Macmillan, London.

Axt, J. R. (2017). The Best Way to Measure Explicit Racial Attitudes Is to Ask About Them. *Social Psychological and Personality Science, 9*(8), 896-906. doi:10.1177/1948550617728995

Banks, D., Hendrix, J., Hickman, M., & Kyckelhahn, T. (2016). National Sources of Law Enforcement Employment Data. *US Department of Justice*. Retrieved from https://www.bjs.gov/content/pub/pdf/nsleed.pdf

Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.) *Dual-process theories in social psychology* (361-382). Guilford Press, New York.

Bayley, D. H. (2008) Police reform: Who done it? *An International Journal of Research and Policy, 18*(1). DOI: 10.1080/10439460701718518

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science, 16*, 101-105.

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General, 133*, 584-600.

Bhandari, P. (2021). *What is a within-subjects design?*. Retrieved from https://www.scribbr.com/methodology/within-subjects-design/

Bias. (2021). In *Oxford Online Dictionary*. Retrieved from https://www.lexico.com/definition/bias

Braga, A., & Schnell, C. (2013). Evaluating Place-Based Policing Strategies: Lessons Learned from the Smart Policing Initiative in Boston. *Police Quarterly, 16*(3), 339-357. doi:10.1177/1098611113497046

Brown, J. D. Choosing the Right Type of Rotation in PCA and EFA. *JALY & Evaluation SIG Newsletter, 13*(3), 20-25. Retrieved from https://hosted.jalt.org/test/PDF/Brown31.pdf

Boettke, P. J., Palagashvili, L., & Piano, E. E. (2017). Federalism and the police: An applied theory of fiscal attention. *Arizona State Law Journal, 49*(3), 907-934

Bogost, I. (2007). *Persuasive games* (Vol. 5). Cambridge, MA: MIT Press.

Borda, J. L., & Marshall, B. (2020). Creating a space to #SayHerName: Rhetorical stratification in the networked sphere. Quarterly Journal of Speech, 106(2), 133-155. doi:10.1080/00335630.2020.1744182

Borrego, M., Douglas, E. P., & Amelink, C. T. (2009). Quantitative, qualitative, and mixed research methods in engineering education. *Journal of Engineering education, 98*(1), 53-66. https://doi.org/10.1002/j.2168-9830.2009.tb01005.x

Brandt, E., Messeter, J., & Binder, T. (2008). Formatting design dialogues - games and participation. *International Journal of CoCreation in Design and the Art, 4*(1), 51-64, DOI: 10.1080/15710880801905724

Buchanan, K. S., Rau, H., Mulligan, K., Keesee, T., Goff, P. A. (2021). White Supremacy in Policing: How Law Enforcement Agencies Can Respond. *Center for Policing Equity*. Retrieved from https://policingequity.org/images/pdfs-doc/CPE-WhiteSupremacy.pdf

Burruss, G. W., & Giblin, M. J. (2014). Modeling Isomorphism on Policing Innovation: The Role of Institutional Pressures in Adoption Community-Oriented Policing. *Crime & Delinquency, 60*(3), 331-355. https://doi.org/10.1177/0011128709340225

Breen, R. L. (2006). A Practical Guide to Focus-Group Research. *Journal of Geography of Higher Education, 30*(3), 463-475. DOI: 10.1080/03098260600927575

Bryant-Davis, T., Adams, T., Alejandre, A., & Gray, A. A. (2017). The Trauma Lens of Police Violence against Racial and Ethnic Minorities. *Journal of Social Issues, 73*(4), 852-871. https://doi.org/10.1111/josi.12251

Cantone, J. A. (2021). Federal and State Court Cooperation: What is Bias and How is it Measured?. *Federal Judicial Center*. Retrieved from https://www.fjc.gov/content/337736/what-bias-and-how-it-measured

Carruthers, J. (1990). A Rationale for the Use of Semi-structured Interviews. *Journal of Educational Administration, 28*(1). https://doi.org/10.1108/09578239010006046

Chae, D. H., Powell, W. A., Nuru-Jeter, A. M., Smith-Bynum, M. A., Seaton, E. K., Forman, T. A., Turpin, R., & Sellers, R. (2017). The Role of Racial Identify and Implicit Racial Bias in Self-Reporrted Racial Discrimination: Implications for Depression Amongst African American Men. *Journal of Black Psychology, 43*(8), 789-812. doi:10.1177/0095798417690055

Cooper, H., Moore, L., Gruskin, S., & Krieger, N. (2004). Characterizing Perceived Police Violence Implications for Public Health. *AM J Public Health, 94*(7), 1109-1118. doi: 10.2105/ajph.94.7.1109

COPS, (2014). *Community Policing Defined*. US Department of Justice. Retrieved from https://cops.usdoj.gov/RIC/Publications/cops-p157-pub.pdf

COPS. (2018). *Community Survey on Public Safety and Law Enforcemen*t. US Department of Justice. Retrieved from https://www.theiacp.org/sites/default/files/2018-08/cops-w0743-pub.pdf

Coquilhat, J. (2008). *Community Policing: An International Literature Review*. New Zealand Police, Crown Copyright.

Cordner, G. (1999). Elements of community policing. In Gaines, L. K., Cordner, G. W. (Eds.), *Policing perspectives: An anthology* (pp. 137–149). Roxbury.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Creswell, J.W., & Plano Clark, V.L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4) 349-354. doi: 10.1037/h0047358

Curtis, M. A., Garlington, S., & Schottenfeld, L. S. (2013). Alcohol, drug, and criminal history restrictions in public housing. *Cityscape, 15*(3), 37-52.

Dario, L. M., & Crichlow, V. J. (2020). Emergent Dimensions of Community-Oriented Policing. *Crime & Delinquency*, 72-97. https://doi.org/10.1177/1077699019835896

Daumeyer, N. M., Onyeador, I. N., Brown, X., & Richeson, J. A. (2019). Consequences of attributing discrimination to implicit vs explicit bias. *Journal of Experimental Social Psychology, 84*. https://doi.org/10.1016/j.jesp.2019.04.010.

de Bono, E. (1971). *The Use of Lateral Thinking*. Penguin UK.

Decker, S. H., Ortiz, N., Spohn, C., & Hedberg, E. (2015). Criminal stigma, race, and ethnicity: The consequences of imprisonment for employment. *Journal of Criminal Justice, 43*(2), 108–121. https://doi.org/10.1016/j.jcrimjus.2015.02.002

DeGue, S., Fowler, K. A., & Calkins, C. (2016). Deaths Due to Use of Lethal Force by Law Enforcement: Findings From the National Violent Death Reporting System, 17 U.S. States, 2009-2012. *American journal of preventive medicine*, *51*(5 Suppl 3), S173–S187. https://doi.org/10.1016/j.amepre.2016.08.027

De la Hera, T., Jansz, J., Jacobs, J., Schouten, B., Raessens, B., & Kors, M. (2021). *Persuasive Gaming in Context*. Amsterdam: Amsterdam University Press. doi 10.5117/9789463728805_ch01

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo Chambers: Emotional Contagion and Group Polarization on Facebook. *Sci Rep 6*, 37825. https://doi-org.tudelft.idm.oclc.org/10.1038/srep37825

DeSilver, D., Lipka, M., Fahmy, D. (2020). 10 things we know about race and policing in the U.S. *Pew Research Center*. Retrieved from https://www.pewresearch.org/fact-tank/2020/06/03/10-things-we-know-about-race-and-policing-in-the-u-s/

De Soto, W. (2018). Do Police Officers in the USA Protect and Serve All Citizens Equally? *Social Sciences, 7*(10), 190. https://doi.org/10.3390/socsci7100190

DeTombe, D. J. (2001). Compram, a method for handling complex societal problems. *European Journal of Operational Research, 128*(2), 266-281. https://doi.org/10.1016/S0377-2217(00)00070-9

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18. DOI: 10.1037/0022-3514.56.1.5

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology, 48*(6), 1267-1278. https://doi.org/10.1016/j.jesp.2012.06.003

DiAngelo, R. (2018). White fragility: *Why it's so hard for white people to talk about racism*. Beacon Press.

Dicle, M. F., & Dicle, B. (2018). Content analysis: Frequency distribution of words. *The State Journal, 18*(2), 379-386. https://doi.org/10.1177/1536867X1801800205

Djaouti, D., Alvarez, J., Jessel, J.P., & Rampnoux, O. (2011). Origins of Serious Games. DOI: 10.1007/978-1-4471-2161-9_3

Dreyer, B. P., Trent, M., Anderson, A. T., Askew, G. L., Boyd, R., Coker, T. R., Coyne-Beasley, T., Fuentes-Afflick, E., Johnson, T., Mendoza, F., Montoya-Williams, D., Willis, E., Wright, J. L., Berman, S., Berkelhamer, J., Jenkins, R. R., Kraft, C., Palfrey, J., Perrin, J. M., & Stein, F. (2020) The Death of George Floyd: Bending the Arc of History Toward Justice for Generations of Children. *Pediatrics, 146*(3).DOI: https://doi.org/10.1542/peds.2020-009639

Druckman, D., & Ebner, N. (2007). Onstage or behind the scenes? Relative learning benefits of simulation role-play and design. *Simulation Gaming, 39*, 465. DOI: 10.1177/1046878107311377

Duke, R. D., & Geurts, J. (2004). *Policy Games for Strategic Management*. Rozenberg Publishers.

Dukes, K. N., & Khan, K. B. (2017). What Social Science Research Says about Police Violence against Racial and Ethnic Minorities: Understanding the Antecedents and Consequences - An Introduction. *Journal of Social Issues, 73*(4), 690-700. https://doi.org/10.1111/josi.12242

Echo Chamber. (2021). In *Oxford Online Dictionary*. Retrieved from https://www.lexico.com/definition/echo_chamber

Eck, J. (2006). Science, values, and problem-oriented policing: Why problem-oriented policing? In Weisburd, D., Braga, A. A. (Eds). *Police innovation: Contrasting perspectives* (pp. 117-132). Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511489334.006

Edwards, F., Lee, H., & Esposito, M. (2018). Risk of being killed by police use of force in the United States by age, race-ethnicity, and sex. *Proceedings of the National Academy of Sciences, 116*(34) 16793-16798. DOI: 10.1073/pnas.1821204116

Evans, D. N., & Porter, J. R. (2015). Criminal history and landlord rental decisions: A New York quasi-experimental study. *Journal of Experimental Criminology, 11*(1), 21–42. https://doi.org/10.1007/s11292-014-9217-4

Evans, D. N., Szkola, J., & St. John, V. (2019). Going Back to College? Criminal Stigma in Higher Education Admissions in Northeastern U.S. *Critical Criminology, 27*, 291-304. https://doi.org/10.1007/s10612-019-09443-x

Feldman, J. (2015). Public Health and the Policing of Black Lives. *Harvard Public Health Review, 7.*

Ferraro, F., Etzion, D., Gehman, J. (2015). Tackling Grand Challenges Pragmatically: Robust Action Revisited. *Organization Studies, 36*(3). DOI: 10.1177/0170840614563742

Fishbein, M. (1966). The relationships between beliefs, attitudes and behavior. *Cognitive consistency, motivational antecedents and behavioral consequents*, 199-223.

Fontaine, J., Leitson, D., Jannetta, J., & Paddock, E. (2017). *Mistrust and Ambivalence between Residents and the Police: Evidence from Four Chicago Neighborhoods*. Urban Institute. Retrieved from https://www.urban.org/sites/default/files/publication/92316/2017.07.31_legitimacy_brief_finalized_2.pdf

Gaertner, S. L., & Dovidio, J. F. (2000). The aversive form of racism. In J. F. Dovidio & S. Gaertner (Eds.) *Prejudice, discrimination, and racism* (61-89). Academic Press, Orlando.

Gains, L. (1993). Community-Oriented Policing: Management Issues, Concerns, and Problems. *Journal of Contemporary Criminal Justice, 10*(1).

Galletta, A. (2013). *Mastering the semi-structured interview and beyond: From research design to analysis and publication* (Vol. 18). NYU press.

Gamson, W. A. (1971). SIMSOC: Establishing social order in a simulated society. *Simulation & Games, 2*(3), 287-308. DOI: 10.1177/003755007100200302

Gaston, S. (2018). Enforcing Race: A Neighborhood-Level Explanation of Black-White Difference in Drug Arrests. *Crime & Delinquency, 65*(4), 499-526. https://doi-org.tudelft.idm.oclc.org/10.1177/0011128718798566

Gau, J. M. (2010). A longitudinal analysis of citizens' attitudes about police. *Policing, 33*(2), 236-252. doi:10.1108/13639511011044867

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731. DOI 10.1037/0033-2909.132.5.692

George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and Tackling Societal Grand Challenges through Management Research. *Academy of Management Journal, 59*(6). DOI: 10.5465/amj.2016.4007

Geva, R., & Shem-Tov, O. (2002). Setting up Community Policing Centers: Participatory Action Research in Decentralized Policing Services. *Police Practice and Research, 3*(3), 189-200. DOI: 10.1080/15614260220150582

Gianakis, G. A., & Davis, G. J. Reinventing or Repackaging Public Services? The Case of Community-Oriented Policing. *Public Administration Review, 58*(6), 485-498. doi:10.2307/977576

Gibbs, A. (1997). Focus groups. *Social research update, 19*(8), 1-8.

Gilbert, K. L., & Ray, R. (2016). Why Police Kill Black Males with Impunity: Applying Public Health Critical Race Praxis (PHCRP) to Address the Determinants of Policing Behaviors and "Justifiable" Homicides in the USA. *Journal of Urban Health, 93*, 122-140. https://doi.org/10.1007/s11524-015-0005-x

Gill, C., Weisburd, D., Telep, C. W., Vitter, Z., & Bennett, T. (2014). Community-oriented policing to reduce crime, disorder and fear and increase satisfaction and legitimacy among citizens: a systematic review. *Journal of Experimental Criminology, 10*, 399-428. https://doi.org/10.1007/s11292-014-9210-y

Glaser, J. (2014). Suspect race: Causes and consequences of racial profiling. Oxford, UK: Oxford University Press.

Goff, P. A., Lloyd, T., Geller, A., Raphael, S., & Glaser, J. (2016). *The Science of Justice: Race, Arrests, and Police Use of Force*. Center for Policing Equity. Retrieved from https://policingequity.org/images/pdfs-doc/CPE_SoJ_Race-Arrests-UoF_2016-07-08-1130.pdf

Goldstein, H. (1987). Toward community-oriented policing. In D. P. Rosenbaum (Ed.), *The challenge of community policing: Testing the promises* (pp. 263-269). London: Sage.

Gomez, M. B. (2016). Policing, Community Fragmentation, and Public Health: Observations from Baltimore. *Journal of Urban Health, 93*, 154-176. https://doi.org/10.1007/s11524-015-0022-9

Gordon, D. (2020). The police as place-consolidators: The organizational amplification of urban inequality. *Law and Social Inquiry, 45*(1), 1-27. doi:10.1017/lsi.2019.31

Goutille, F. (2009). *Knowledge, Attitudes, and Practices for Risk Education: how to implement KAP surveys*. Handicap International. Retrieved from https://reliefweb.int/sites/reliefweb.int/files/resources/Landmines_How_to_implement_KAP_Surveys.pdf

Govender, D. (2020). Reflections on community-based participation to reduce violent crimes in south africa. *African Security Review*. doi:10.1080/10246029.2020.1830139

Green, C. A., Duan, N., Gibbons, R. D., Hoagwood, K. E., Palinkas, L. A., & Wisdom, J. P. (2016). Approaches to Mixed Methods Dissemination and Implementation Research: Meethods, Strengths, Caveats, and Opportunities. *Administration and policy in mental health, 42*(5), 508–523. https://doi.org/10.1007/s10488-014-0552-6

Greenwalk, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4-27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480. https://doi.org/10.1037/0022-3514.74.6.1464

Grimm, P. (2010). Social Desirability Bias. *Wiley international encyclopedia of marketing*

Hall, J. M., & Carlson, K. (2016). Marginalization: A Revisitation with Integration of Scholarship on Globalization, Intersectionality, Privilege, Microaggressions, and Implicit Bias. *Advances in Nursing Sciences, 39*(3), 200-215. DOI: 10.1097/ANS.0000000000000123

Harteveld, C. (2011). *Triadic game design: Balancing reality, meaning and play*. Springer Science & Business Media.

Harteveld, C., Guimarães, R., Mayer, I. S., & Bidarra, R. (2010). Balancing play, meaning and reality: The design philosophy of LEVEE PATROLLER. *Simulation & Gaming, 41*(3), 316-340. https://doi.org/10.1177/1046878108331237

Helander, M. E., & McNeill Brown, A. (2020). The Public Health Crisis of Law Enforcement's over-Use of Force'. *Lerner Center for Public Health Issues. Issue Brief, 37*.

Hester, N., & Gray, K. (2018). For Black men, being tall increases threat stereotyping and police stops. *PNAS, 115*(11), 2711-2715. https://doi.org/10.1073/pnas.1714454115

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

IBM. (2021a). *KMO and Bartlett's Test*. Retrieved from https://www.ibm.com/docs/en/spss-statistics/26.0.0?topic=detection-kmo-bartletts-test

IBM. (2021b). *Point-biserial correlations in SPSS*. Retrieved from https://www.ibm.com/support/pages/point-biserial-correlations-spss

Insler, M. A., McMurrey, B., McQuoid, A. F. (2019). From broken windows to broken bonds: Militarized police and social fragmentation. *Journal of Economic Behavior & Organization, 163*, 43-62. https://doi.org/10.1016/j.jebo.2019.04.012

Iyengar, S., Messing, S., & Hahn, K. S. (2011). Explicit and implicit racial attitudes: A test of their convergent and predictive validity. In *APSA 2011 Annual Meeting Paper*.

Jacobs, R. S. (2018). Play to win over: Effects of persuasive games. *Psychology of Popular Media Culture, 7*(3), 231–240. https://doi.org/10.1037/ppm0000124

Jacobs, R. S., Jansz, J., & de la Hera CondePumpido, T. (2017). 10 The Key Features of Persuasive Games. *New perspectives on the social aspects of digital gaming: Multiplayer 2*.

Jacobson, C. K. (1985). Resistance to affirmative action: Self-Interest or racism? *Journal of Conflict Resolution, 29*, 306–329.

Javed, A., & Kohda, Y. (2020). Facilitation Engineering: A realization of Dense Learning Through Facilitation Embedded in Serious Game. *2020 6th International Conference on Information Management (ICIM)*, London, UK, 306-313. DOI: 10.1109/ICIM49319.2020.244716

Jick, T.D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly, 24*(4).

Johnson, R.B., Onwuegbuzie, A.J., Turner, L.A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1, 112-133. DOI: 10.1177/1558689806298224

Joshi, A., Kale, S., Chandel, S., Pal, D. K. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology, 7*(4), 396-403.

Kahn, K. B., & Martin, K. D. (2016). Policing and Race: Disparate Treatment, Perceptions, and Policy Responses. *Social Issues and Policy Review, 10*(1), 82-121. https://doi.org/10.1111/sipr.12019

Ke, F. (2014). An implementation of design-based learning through creating educational computer games: A case study on mathematics learning during design and computing. *Computers & Education, 73*, 26-39. https://doi.org/10.1016/j.compedu.2013.12.010

Kerley, K. R., & Benson M. L. (2000). Does Community-Oriented Policing Help Build Stronger Communities? *Police Quarterly, 3*(1). doi:10.1177/1098611100003001002

Kim, D., & Li, M. (2021). Digital storytelling: facilitating learning and identity development. *Journal of Computers in Education, 8*, 33-61. DOI: https://doi-org.tudelft.idm.oclc.org/10.1007/s40692-020-00170-9

Kinder, D. R., & Sanders, L. M. (1996). *Divided by Color: Racial Politics and Democractic Ideas*. Chicago: The University of Chicago Press

Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology, 40*(3), 414-431. DOI: 10.1037/0022-3514.40.3.414

Kocak, D. (2018). On The Spatial-temporal Diffusion of Community based Policing from Japan to Peninsula Southeast Asia: The Case of Timor-Leste. *Journal of Intervention and Statebuilding, 13*(1), 25-40. DOI: 10.1080/17502977.2018.1516390

Kortmann, R., & Harteveld, C. (2009). Agile game development: lessons learned from software engineering. In *Learn to game, game to learn; the 40th conference ISAGA*.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.

Krueger, R. A. (2014). *Focus groups: A practical guide for applied research*. Sage publications.

Laamarti, F., Eid, M., El Saddik, A. (2014). An Overview of Serious Games. *International Journal of Computer Games Technology, 2014*. DOI: 10.1155/2014/358152

Lawson, B. (1980). How Designers Think. Eastview Editions.

Leasure, P. (2019). Securing private housing with a criminal record. *Journal of Offender Rehabilitation, 58*(1), 30-49. DOI: 10.1080/10509674.2018.1549182

Lee, J., & Choi, Y. (2019). Effects of network heterogeneity on social media on opinion polarization among South Koreans: Focusing on fear and political orientation. *International Communication Gazette, 82*(2), 119-139. https://doi-org.tudelft.idm.oclc.org/10.1177/1748048518820499

Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016). Blacks' Death Rate Due to Circulatory Disease is Positively Related to Whites' Explicit Racial Bias: A Nationwide Investigation Using Project Implicit. *Psychological Science, 27*(10), 1299-1311. https://doi.org/10.1177/0956797616658450

Linssen, J., Theune, M., & Bruijnes, M. (2014). Beyond simulations: serious games for training interpersonal skills in law enforcement. In *Social Simulation Conference*. Retrieved from https://ddd.uab.cat/pub/poncom/2014/127944/ssc14_a2014a16iENG.pdf

Louise Barriball, K., & While, A. (1994). Collecting Data using a semi-structured interview: a discussion paper. *Journal of advanced nursing, 19*(2), 328-335. https://doi.org/10.1111/j.1365-2648.1994.tb01088.x

Lukosch, H., van Ruijven, T., Verbraeck, A. (2012). The participatory design of a simulation training game. *Proceedings of the 2012 Winter Simulation Conference (WSC)*, Berlin, 2012, pp. 1-11, doi: 10.1109/WSC.2012.6465218.

Lund, A., & Lund, M. (2018). *Paired t-test using Stata*. Retrieved from https://statistics.laerd.com/stata-tutorials/paired-t-test-using-stata.php

Maguire, E. R., & Mastrofski, S. D. (2000). Patterns of Community Policing in the United States. *Police Quarterly, 3*(1). doi: 10.1177/1098611100003001001

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science, 17*, 944-948. https://doi.org/10.1111/j.1467-9280.2006.01809.x

McBride, M. F., Lambert, K. F, Huff, E. S., Theoharides, K. A., Field, P., & Thompson, J. R. (2017). Increasing the effectiveness of participatory scenario development through codesign. *Ecology and Society, 22*(3):16. https://doi.org/10.5751/ES-09386-220316

McCabe, A., & Halog, A. (2018). Exploring the potential of participatory systems thinking techniques in progressing SLCA. *The International Journal of Life Cycle Assessment, 23*, 739-750. https://doi-org.tudelft.idm.oclc.org/10.1007/s11367-016-1143-4

McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In Dovidio, J. F., Gaertner, S. L. (Eds.), *Prejudice, Discrimination, and Racism* (91–125). San Diego, CA: Academic Press.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology, 37*(5), 435-442. https://doi.org/10.1006/jesp.2000.1470

McGregor, A. (2016). Politics, Police Accountability, and Public Health: Civilian Review in Newark, New Jersey. *J Urban Health, 93*(1), 141-153. doi: 10.1007/s11524-015-9998-4

Mildner, P., & Mueller, F. (2016) Design of Serious Games. In: Dörner R., Göbel S., Effelsberg W., Wiemeyer J. (eds) *Serious Games* (pp. 57-82). Springer, Cham. https://doi.org/10.1007/978-3-319-40612-1_3

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access*, *8*, 131662-131682. doi: 10.1109/ACCESS.2020.3009626

Muleme, J., Kankya, C., Ssempebwa, J. C., Mazeri, S., & Muwonge, A. (2017). A Framework for Integrating Qualitative and Quantitative Data in Knowledge, Attitude, and Practice Studies: A Case Study of Pesticide Usage in Eastern Uganda. *Frontiers in public health, 5*, 318. https://doi.org/10.3389/fpubh.2017.00318

Nemeth, C. (1986). Differential contributions of majority and minority influence. *Psychological Review, 93*, 23 – 32.

Nemeth, C., & Wachtler, J. (1983). Creative problem solving as a result of majority vs. minority influence. *European Journal of Social Psychology, 13*, 45 – 55.

Norges Miljo-Og Biovitenskaplige Universitet. (2015). *Community-Based Policing and Post-Conflict Police Reform*. Retrieved from https://cordis.europa.eu/project/id/653909

Obasogie, O. K., & Newman, Z. (2017). Police Violence, Use of Force Policies, and Public Health. *American Journal of Law & Medicine. 43*(2-3), 279-295. https://doi.org/10.1177/0098858817723665

O'Brien, T. C., Tyler, T. R., & Meares, T. L. (2020). Building popular legitimacy with reconciliatory gestures and participation: A community-level model of authority. *Regulation and Governance, 14*(4), 821-839. doi:10.1111/rego.12264

Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, An. G. (2009). A Qualitative Framework for Collecting and Analyzing Focus Group Research. *International Journal of Qualitative Methods, 8*(3), 1-21.

Ørngreen, R., & Levinsen, K. (2017). Workshops as a Research Methodology. *Electronic Journal of E-learning, 15*(1), 70-81.

Pager, D. (2007). *Marked: Race, Crime, and Finding Work in an Era of Mass Incarceration*. Chicago, IL: University of Chicago Press.

Pager, D. (2003). The mark of a criminal record. *American Journal of Sociology, 108*(5), 937–975. https://doi.org/10.1086/374403

Pardede, P. (2019). Mixed Methods Research Designs in EFL.

Peters, V., & Van de Westelaken, M. (2014). Simulation Games-A Concise Introduction to the Design Process. *Samenspraak Advies Nijmegen*.

Philippas, D. (2014) Analysis of Covariance (ANCOVA). In: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi-org.tudelft.idm.oclc.org/10.1007/978-94-007-0753-5_82

Pierce, M. W., Runyan, C. W., & Bangdiwala, S. I. (2014). The use of criminal history information in college admissions decisions. *Journal of School Violence, 13*(4), 359–376. https://doi.org/10.1080/15388220.2013.870061

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour, 4*, 736-745. https://doi-org.tudelft.idm.oclc.org/10.1038/s41562-020-0858-1

Price, P. C., Jhangiani, R., & Chiang, I. A. (2015). Research methods in psychology-2nd Canadian Edition. BC Campus. Retrieved from https://opentextbc.ca/researchmethods/chapter/some-basic-null-hypothesis-tests/

PyPI. (2021). *Clean-text 0.4.0*. Retrieved from https://pypi.org/project/clean-text/

Rake, D. E. (1987). Crime Control and Police-Community Relations: A Cross-Cultural Comparison of Tokyo, Japan, and Santa Ana, California. *The Annals of the American Academy of Political and Social Science, 494*, 148-154. Retrieved November 20, 2020, from http://www.jstor.org/stable/1045715

Rau, H., L., Subramanian, A., Simmons, S. J., & Goff, P. A. (2020). *Toolkit for Equitable Public Safety*. Center for Policing Equity & PolicyLink. Retrieved from https://policingequity.org/what-we-do/research/toolkit-for-equitable-public-safety

Ribeiro, C., Fernandes, J., Lourenço, A., Borbinha, J., & Pereira, J. (2012). Using serious games to teach business process modeling and simulation. In P*roceedings of the International Conference on Modeling, Simulation and Visualization Methods (MSV)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). Retrieved from http://worldcomp-proceedings.com/proc/p2012/MSV2948.pdf

Ritterfield, U., Cody, M., & Vorderer, P. (2009). *Serious Games: Mechanisms and Effects*. Routledge

Robinson, M. (2002). Justice Blind? Ideals and Realities of American Criminal Justice. *National Criminal Justice Reference Service*. Retrieved from https://www.ncjrs.gov/App/abstractdb/AbstractDBDetails.aspx?id=188577

Rohe, W. M., Adams, R. E., & Arcury, T. A. (1997). *Community oriented policing: What it is, why it works, how to get started*. Center for Urban and Regional Studies, University of North Carolina at Chapel Hill. Retrieved from https://www.ojp.gov/pdffiles1/Digitization/169113NCJRS.pdf

Rosenbaum, D. P., & Lurigio, A. (1994). An Inside Look at Community Policing Reform: Definitions, Organizational Changes, and Evaluation Findings. *Crime & Delinquency, 40*(3), 299-314. https://doi.org/10.1177/0011128794040003001

Ross, A., & Willson, V. L. (2017). Paired samples T-test. In *Basic and advanced statistical tests*, (pp. 17-19). Brill Sense. https://doi.org/10.1007/9789463510868_005

Ruteere, M., & Pomeerolle, M. E. (2003). Democratizing Security or Decentralizing Repression? The ambiguities of community policing in Kenya. *African Affairs, 102*(409), 587-604. https://doi.org/10.1093/afraf/adg065

Sârbescu, P., Costea, I., & Rusu, S. (2012). Psychometric properties of the Marlowe-Crowne Social Desirability Scale in a Romanian sample. *Procedia - Social and Behavioral Sciences, 33*, 707-711. https://doi.org/10.1016/j.sbspro.2012.01.213

Samson, F. L. (2015). Racial resentment and smoking. *Social Science & Medicine, 126*, 164-168. https://doi.org/10.1016/j.socscimed.2014.12.033

Schneider, E. C., Agee, C., & Chronopoulos, T. (2020). Dirty work: Police and community relations and the limits of liberalism in postwar philadelphia. *Journal of Urban History, 46*(5), 961-979. doi:10.1177/0096144217705497

Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology, 6,* 474. doi: 10.3389/fpsyg.2015.00474

Schoonenboom, J., & Johnson, R. B. (2017). How to Construct a Mixed Methods Research Design. *Kolner Zeitschrift fur Soziologie und Sozialpsychologie, 69*(Suppl 2), 107–131. https://doi.org/10.1007/s11577-017-0454-1

Scott, J. D. (2002). Assessing the Relationship Between Police-Community Coproduction and Neighborhood-Level Social Capital. *Journal of Contemporary Criminal Justice, 18*(2). doi:10.1177/1043986202018002003

Sears, D. O. (1988). Symbolic racism. In Katz, P. A., Taylor, D. A. (Eds.), *Eliminating racism: Profiles in controversy* (53–84). New York, NY: Plenum.

Seitz, R. J., & Angel, H. F. (2020). Belief formation - A driving force for brain evolution. *Brain and Cognition, 140*, 105548. https://doi.org/10.1016/j.bandc.2020.105548.

Selmi, M. (2018). The paradox of implicit bias and a plea for a new narrative. *Ariz. St. LJ, 50*, 193.

Sentiment Analysis. (2021). In *Oxford Online Dictionary*. Retrieved from https://www.lexico.com/definition/sentimentanalysis

Simonsen, J., & Robertson, T. (2012). *Routledge International Handbook of Participatory Design*. Routledge, 2012

Smith, C., Felderhof, L., & Bosch, J. H. (2008). Adaptive management: making it happen through participatory systems analysis. *Systems Research and Behavioral Science, 24*(6), 567-587. https://doi-org.tudelft.idm.oclc.org/10.1002/sres.835

Sinclair, J., & Irani, T. (2013). ADVOCACY ADVERTISING FOR BIOTECHNOLOGY: The Effect of Public Accountability on Corporate Trust and Attitude Toward the Ad. *Journal of Advertising, 34*(3), 59-73. DOI: 10.1080/00913367.2005.10639203

Solomon, A. L. (2012). In search of a job: Criminal records as barriers to employment. *NIJ Journal, 270*, 42-51.

Sorace, S., Quercia, E., La Mattina, E., Patrikakis, C. Z., Bacon, L., Loukas, G., & Mackinnon, L. (2018). Serious Games: An Attractive Approach to Improve Awareness. In: Leventakis G., Haberfeld M. (eds) Community-Oriented Policing and Technological Innovations. SpringerBriefs in Criminology. Springer, Cham. doi: 10.1007/978-3-319-89294-8_1

Sorbring, E., Bolin, A., & Ryding, J. (2015). Game-Based Intervention - A Technical Tool for Social Workers to Combat Adolescent Dating Violence. *Advances in Social Work, 16*(1). 125-139. DOI: https://doi.org/10.18060/18260

Spencer, K. B., Charbonneau, A. K., Glaser, J. (2016). Implicit Bias and Policing. *Social and Personality Psychology Compass, 10*(1). DOI: 10.1111/spc3.12210

Sunstein, C. R. (2006). The law of group polarization. *Journal of political philosophy 10*, 175–195.

Suprun, E., Sahin, O., Stewart, R. A., Panuwatwanich, K., & Shcherbachenko, Y. (2018). An Integrated Participatory Systems Modelling Approach: Application to Construction Innovation. *Systems, 6*(3), 33. https://doi.org/10.3390/systems6030033

Thurman, Q. C. (1995). Community policing: The police as a community resource. In P. Adams & K. Nelson (Eds.), *Reinventing human services: Community-and family-centered practices* (pp. 122-145). New York: Aldine De Gruyter.

Toch, H. (2008). Police officers as change agents in police reform. *Policing and Society, 18*(1). doi: 10.1080/10439460701718575

Treece, E.W., & Treece, J.W. (1986). *Elements of Research in Nursing* 4th edn. C.V. Mosby, St Louis

Troyer, L., & Youngreen, R. (2009). Conflict and Creativity in Groups. *Journal of Social Issues, 65*(2), 409-427.

Uggen, C., Vuolo, M., Lageson, S., Ruhland, E., & Whitham, H. K. (2014). The Edge of Stigma: An Experimental Audit of the Effects of Low-Level Criminal Records on Employment. *Criminology, 52*(4), 627-654. https://doi.org/10.1111/1745-9125.12051

ul Haq, N., Hassali, M.A., Shafie, A.A., Saleem, F., Farooqui, M., & Aljadhey, H. (20120). A cross sectional assessment of knowledge, attitude and practice towards Hepatitis B among healthy population of Quetta, Pakistan. *BMC Public Health 12*, 692. https://doi.org/10.1186/1471-2458-12-692

Ungar, M., & Desmond Arias, E. (2012). Reassessing community-oriented policing in Latin America. *Policing and Society, 22*(1), 1-13. DOI: 10.1080/10439463.2011.597856

Unzueta, M. M., & Lowery, B. S. (2008). Defining racism safely: The role of self-image maintenance on white Americans' conceptions of racism. *Experimental Social Psychology, 44*(6), 1491-1497. https://doi.org/10.1016/j.jesp.2008.07.011

van der Velden, M., Mörtberg, C., Van den Hoven, J., Vermaas, P. E., & Van de Poel, I. (2014). Participatory design and design for values. *Development, 11*(3), 215-236.

Van Dijk, T. A. (1992). Denying racism: Elite discourse and racism. *Discourse and Society, 3*(1), 87-118.

Vos, N., van der Meijden, H., & Denessen, E. (2011). Effects of constructing versus playing an educational game on student motivation and deep learning strategy use. *Computers & Education, 56*(1), 127-137. https://doi.org/10.1016/j.compedu.2010.08.013

Walker, S. (2001). Police accountability. *Belmont, CA: Wadsworth*, 199-214.

Warr, A., & O'Neill, E. (2006). The effect of group composition on divergent thinking in an interaction design activity. In *Proceedings of the 6th conference on Designing Interactive systems* (pp. 122-131).

Weisburd, D., Majmundar, M. K., Aden, H., Braga, A., Bueermann, J., Cook, P. J., . . . Tyler, T. (2019). Proactive policing: A summary of the report of the national academies of sciences, engineering, and medicine. Asian Journal of Criminology, 14(2), 145-177. doi:10.1007/s11417-019-09284-1

Weissman, M., Rosenthal, A., Warth, P., Wolf, E., & Messina-Yauchzy, M. (2010). Reconsidered: The use of criminal history records in college admissions. New York: Center for Community Alternatives. Retrieved http://www.communityalternatives.org/pdf/Reconsidered-criminal-hist-recs-in-college-admissions.pdf.

Weng, J. (2019). *NLP Text Preprocessing: A Practical Guide and Template*. Retrieved from https://towardsdatascience.com/nlp-text-preprocessing-a-practical-guide-and-template-d80874676e79

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76*, 92-104.

Winn, B. M. (2009). The design, play, and experience framework. In *Handbook of research on effective electronic gaming in education* (pp. 1010-1024). IGI Global.

Worthy, D. A., Markman, A. B., & Maddox, W. T. (2009). What is pressure? Evidence for social pressure as a type of regulatory focus. *Psychonomic Bulletin & Review, 16*, 344-349. https://doi-org.tudelft.idm.oclc.org/10.3758/PBR.16.2.344 result of being under additional pressure

Wycoff, M. A. (1991). The Benefits of Community Policing: Evidence and Conjecture. In Green, J. R., & Mastrofski, S. D. (Eds). *Community Policing: Rhetoric or Reality* (pp. 103-120). New York: Praeger

Ying, M. (2020). *NLP: How does NLTK.Vader Calculate Sentiment?*. Retrieved from https://medium.com/ro-data-team-blog/nlp-how-does-nltk-vader-calculate-sentiment-6c32d0f5046b

Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Emotional dynamics in the age of misinformation. *PloS one, 10*(9), e0138740. https://doi.org/10.1371/journal.pone.0138740

APPENDIX A

This appendix outlines the Knowledge, Attitude, and Practice (KAP) survey development process and the validated KAP survey questions used during both Workshop A and Workshop B. First, Andrade et al.'s (2020) guidelines for KAP survey development have been outlined with reference to its application in this research. Then, the specifics of the KAP survey validation process used for this study are put forth. Finally, all validated KAP questions used in the final validated survey are outlined in a concise table (Table A9).

A.1 KAP Survey Development Process

With the use of Andrade et al.'s (2020) five step guideline, a KAP survey has been developed to generate insight into the knowledge, attitudes, and practices of study participants towards their community, policing, and racism. Specifically, KAP survey constructs have been developed relevant to "Attitude about Explicit Racism", "Attitude about Policing in Communities", "Practice about Policing in Communities", and "Knowledge about Policing in Communities". In order to develop the structured questionnaire for these constructs, a need for the survey was identified, the target population was defined, questions were prepared, potential answers to questions were drafted, and survey scoring was also outlined according to the work of Andrade et al.'s (2020) and Joshi et al. (2015). While Andrade et al.'s (2020) technique for KAP survey validation was considered, alternative survey validation steps were taken given limitations to time and resources for this study. Details of the KAP survey development process for this study have been outlined below. Table A1 outlines the first draft of developed survey questions, the sources in which questions were drafted from, the associated question scoring system, the associated answers, the associated hypothesized construct, and a brief explanation as to how each question addresses the proposed hypothesized construct.

1. Topic Validation & Selection of the Target Population

According to Andrade et al. (2020), "a good research question addresses a felt need" (p. 479). Therefore, for a topic to be validated in terms of genuinely calling for a KAP study, a need for intervention must exist (Andrade et al., 2020). Similarly, a valid "target population for a KAP survey selects itself because it is the population in which the need exists" (Andrade et al., 2020, p. 479). Given Andrade et al.'s (2020) guidelines and as outlined by section 1.1.1, it is evident that police-community relations and biased policing in the US serves as a valid topic for a KAP study. Additionally, seeing as law enforcement officers express bias which translates into bias policing practices that communities are subjected to, the selected target population of both law enforcement officers and community members can be considered valid for this analysis. Both law enforcement officers and community members are key stakeholders involved in the case of police-community relations and biased US policing, and as such, they both hold a need that must be addressed with a KAP study.

2. Preparation of Questions

In order to prepare KAP survey questions, substantial thought must be put into question framing (Andrade et al., 2020). Survey "questions" can be posed as actual formal questions, or questions can be posed as statements in which respondents must decide how much they agree or disagree with the statement's contents (Andrade et al., 2020). In order to frame KAP survey questions, the level of expected knowledge of the target population must be defined (Andrade et al., 2020). However, assessing knowledge is only relevant to the extent in which it influences both attitudes and practice (Andrade et al., 2020). Therefore, it is critical to determine which facts the target population must be aware of so that they can form sound attitudes and act with competence (Andrade et al., 2020). In testing the target population's knowledge, questions about general facts, myths, and misconceptions must be posed (Andrade et al., 2020). Questions relevant to attitude and practice should be posed in a similar manner (Andrade et al., 2020). Notably, questions should not be framed in such a way that is too obvious or too difficult for respondents to answer (Andrade et al., 2020).

Given Andrade et al.'s (2020) guidelines for question preparation, a general baseline of knowledge in which the average US citizen should have about their community, policing, and racism has been determined as the level of expected knowledge for the target population of law enforcement officers and community members. This baseline level of knowledge has been formed with a US citizen average level of civic engagement and access to national news in mind. In expanding upon this notion of expected knowledge, questions relevant to attitudes and practices for study have also been drafted in accordance with Andrade et al.'s (2020) guidelines. In addition to Andrade et al.'s guidelines for question

preparation, questions have been drafted explicitly per the recommendations of Axt (2017). Additionally, following the serious game design workshop intervention, KAP survey #2 also includes workshop activity related questions relevant to constructs of "structure & flow", "usability", and "team communication" to gauge the influence that such elements may have had on survey results (Table A2).

3. Preparing Options for Answers

When preparing options for answers, it is critical to ensure that respondents are not forced into selecting options that they may not truly endorse (Andrade et al., 2020). As such, it is crucial to always include a "I Don't Know" option in responses for knowledge-based questions (Andrade et al., 2020). Simultaneously, however, it is also important to ensure that too many option responses are not provided, so to not overwhelm respondents nor complicate survey scoring (Andrade et al., 2020). For this study, these guidelines have been considered for the drafting of option answers.

4. Scoring the Questionnaire

KAP surveys should not result in a total score, but instead they should result in subscale scores for each construct (Andrade et al., 2020). Additionally, in scoring the KAP survey, it can be assumed that all questions are of equal difficulty, therefore each question's answer is considered with equal weighting (Andrade et al., 2020). In scoring knowledge-based questions, 1 is assigned to properly answered questions and 0 is assigned to answers in which were answered incorrectly or as "I Don't Know" (Andrade et al., 2020). According to Andrade et al. (2020) and similarly, attitude-based questions which express positive attitude are scored as 1, and attitude-based questions that express negative attitude are scored as -1. Practice-based questions can be scored with similar reasoning to attitude-based questions (Andrade et al., 2020).

In order to score KAP survey results for this study, scoring guidelines according to Andrade et al. (2020) have been adhered to for knowledge-based questions. However, a 5-point Likert scale has instead been used for scoring both attitude and practice based questions. For questions that utilize the 5-point Likert scale in this study, a response of "1" correlates to "strongly disagree", a response of "2" correlates to "disagree", a response of "3" correlates to "neutral", a response of "4" correlates to "agree", and a response of "5" correlates to "strongly agree". Likert scale questions were scored according to guidelines put forth by Joshi et al. (2015).

Table A1: Knowledge, attitude, and practice questions relative to racism, community, and policing used in the KAP surveying portion of this research. Questions were drawn from previously validated questionnaires and checklists from the works of Iyengar et al. (2011), Kinder & Sanders (1996), and COPS (2018). Studies by Fontaine et al. (2017), Rohe et al. (1997), Buchanan et al. (2021), Edwards, Lee, & Esposito (2018), Pierson et al. (2020), Gaston (2018), and Goutille (2009) also informed question content and formatting.

#	Hypothesized Construct	Question	Scoring	Correct Answer	Source	Explanation of how the Question Measures the Hypothesized Construct
Q1_1	Attitude about Explicit Racism	I feel warmly towards Black people	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how warm/cold the participant feels towards Black people
Q1_2	Attitude about Explicit Racism	I feel warmly towards white people	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how warm/cold the participant feels towards white people
Q1_3	Attitude about Explicit Racism	Black people are lazy	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how hard working/lazy the participant perceives Black people
Q1_4	Attitude about Explicit Racism	Black people are peaceful	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how violent/peaceful the participant perceives Black people
Q1_5	Attitude about Explicit Racism	Black people are dependent	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how self-reliant/dependent the participant perceivesBblack people
Q1_6	Attitude about Explicit Racism	Black people are closed to interact with people of different backgrounds	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how open/closed the participant perceives Black people to be towards people from different backgrounds
Q3_1	Attitude about Explicit Racism	Over the past few years, Black people have gotten less than they deserve	Likert Scale (1-5)	----	Kinder & Sanders, 1996	Explicitly measures racism on the basis of the participant's opinions related to the realities of systemic oppression
Q3_2	Attitude about Explicit Racism	Irish people, Italian people, Jewish people, Vietnamese people, and other minority groups in the United States overcame prejudice and worked their way up. Black people should do the same without any special favors	Likert Scale (1-5)	----	Kinder & Sanders, 1996	Explicitly measures racism on the basis of the participant's opinions related to the realities of systemic oppression
Q3_3	Attitude about Explicit Racism	It is really a matter of some people not trying hard enough; If Black people would only try harder they could be just as well off as white people	Likert Scale (1-5)	----	Kinder & Sanders, 1996	Explicitly measures racism on the basis of the participant's opinions related to the realities of systemic oppression
Q3_4	Attitude about Explicit Racism	Generations of slavery and discrimination have created conditions that make it difficult for Black people to work they way out of the lower class.	Likert Scale (1-5)	----	Kinder & Sanders, 1996	Explicitly measures racism on the basis of the participant's opinions related to the realities of systemic oppression
Q2_1	Attitude about policing in communities	The police are legitimate and they are justified to exercise their power to maintain order and solve crimes. For this reason, people should obey and follow the rules and laws that the police enforce.	Likert Scale (1-5)	----	Fontaine et al., 2017; COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to police legitimacy in communities
Q2_2	Attitude about policing in communities	Police officers are approachable and easy to talk to.	Likert Scale (1-5)	----	COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to how approachable police officers are to communities
Q2_3	Attitude about policing in communities	Police officers show concern for community members.	Likert Scale (1-5)	----	COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to police officers' perceived concern for community members
Q2_4	Attitude about policing in communities	Police officers are responsive to the concerns of community members.	Likert Scale (1-5)	----	COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to police officers' responsiveness towards the concerns of community members
Q2_5	Attitude about policing in communities	Police officers work with community members to solve local problems.	Likert Scale (1-5)	----	COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to police-citizen co-production efforts
Q4	Knowledge about policing in communities	Community-oriented policing (COP) can be broadly defined as police officers and residents working together to solve local problems.	True/ False/ I Don't Know	TRUE	Rohe et al., 1997; COPS, 2018	Measures a participant's level of knowledge about policing and communities on the basis of what COP is
Q5	Knowledge about policing in communities	Community-oriented policing (COP) is a top-down approach to policing.	True/ False/ I Don't Know	FALSE	Rohe et al., 1997	Measures a participant's level of knowledge about policing and communities on the basis of what COP is
Q6	Knowledge about policing in communities	While community-oriented policing (COP) can address community-specific issues, COP can also be used to solve long-term endemic problems, such as poverty and unemployment.	True/ False/ I Don't Know	FALSE	Rohe et al., 1997; COPS, 2018; Goutille, 2009	Measures a participant's level of knowledge about policing and communities on the basis of what COP is and what its limitation are
Q7	Knowledge about policing in communities	Historically, law enforcement has been entangled with white supremacy and associated hate groups.	True/ False/ I Don't Know	TRUE	Buchanan et al., 2021	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q8	Knowledge about policing in communities	Racial disparities in public safety still exist today.	True/ False/ I Don't Know	TRUE	Center for Policing Equity, 2021	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q9	Knowledge about policing in communities	Folks that are Black, Indigenous, and People of Color (BIPOC) are less likely to experience police-involved harm than white people.	True/ False/ I Don't Know	FALSE	Edwards, Lee, & Esposito, 2018	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q10	Knowledge about policing in communities	White people are more likely to be stopped by the police than folks that are Black, Indigenous, and People of Color (BIPOC).	True/ False/ I Don't Know	FALSE	Pierson et al., 2020	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q11	Knowledge about policing in communities	In some communities, residents do not trust the police to treat them fairly.	True/ False/ I Don't Know	TRUE	Rohe et al., 1997	Measures a participant's level of knowledge about policing and communities on the basis of the historic distrust between different communities and police officers in which has been fueled by systemic oppression
Q12	Knowledge about policing in communities	Some police officers believe that they are disliked by some communities, and these police officers express a low degree of confidence towards these communities to come to their aid during a crisis.	True/ False/ I Don't Know	TRUE	Fontaine et al., 2017; Rohe et al., 1997; COPS, 2018	Measures a participant's level of knowledge about policing and communities on the basis of the historic distrust between different communities and police officers in which has been fueled by systemic oppression
Q13	Knowledge about policing in communities	In communities where folks are predominantly Black, Indigenous, and People of Color (BIPOC), laws are less strictly enforced by police than in communities where residents are predominantly white.	True/ False/ I Don't Know	FALSE	Gaston, 2018	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q2_6	Practice about policing in communities	I fear the police	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much they fear police
Q2_7	Attitude about policing in communities	I trust the police.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much they trust police
Q2_8	Practice about policing in communities	I trust community members and their efforts to collaborate with the police.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much they trust community members and efforts to foster collaboration with police
Q2_9	Attitude about policing in communities	I have confidence in the police to address community concerns.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much confidence that they have in police
Q2_10	Practice about policing in communities	I have confidence in communities to offer aid to police for addressing community concerns.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much confidence they have in communities to offer aid to the police in order to address community concerns
Q2_11	Attitude about policing in communities	I am comfortable communicating with the police.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are communicating with police
Q2_12	Attitude about policing in communities	I am comfortable communicating with community members within organizations that collaborate with the police.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards how comfortable they are communicating with community members and organizations that collaborate with police
Q2_13	Practice about policing in communities	I would contact the police for help.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are reaching out to the police for help
Q2_14	Practice about policing in communities	I would contact community members within organizations that collaborate with the police for help with addressing community-specific problems	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are reaching out for help to community members and organizations that collaborate with the police
Q2_15	Practice about policing in communities	I would participate in a community-oriented policing (COP) initiative locally if invited.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how willing they would be to participate in a local COP initiative

Table A2: KAP survey #2 serious game design process questions relative to structure & flow, usability, and team communication by Winn (2009)

#	Construct	Question	Scoring	Correct Answer	Source
1	Structure & Flow	The steps in the serious game conceptual design workshop process follow a logical sequence	Likert Scale (1-5)	----	Winn, 2009
2	Structure & Flow	The serious game conceptual design workshop process is adequate for designing game concepts	Likert Scale (1-5)	----	Winn, 2009
3	Structure & Flow	The included activities in the serious game conceptual design workshop are effective	Likert Scale (1-5)	----	Winn, 2009
4	Usability	The serious game conceptual design workshop process is clear to follow and apply	Likert Scale (1-5)	----	Winn, 2009
5	Usability	The serious game conceptual design workshop can be useful for designing games in teams	Likert Scale (1-5)	----	Winn, 2009
6	Usability	The serious game conceptual design workshop does not need to be simplified	Likert Scale (1-5)	----	Winn, 2009
7	Usability	The language used in the brainstorming cards is clear	Likert Scale (1-5)	----	Winn, 2009
8	Team Communication	The serious game conceptual design workshop allows all team members the opportunity to provide their input	Likert Scale (1-5)	----	Winn, 2009
9	Team Communication	The brainstorming cards effectively help to generate ideas	Likert Scale (1-5)	----	Winn, 2009
10	Team Communication	After using the serious game conceptual design workshop process, I think that I have a better understanding of the various components of game design	Likert Scale (1-5)	----	Winn, 2009
11	Team Communication	There is value in using the serious game conceptual design workshop process in a multidisciplinary game design team	Likert Scale (1-5)	----	Winn, 2009

A.2 KAP Survey Validation Process

According to Andrade et al. (2020), a KAP survey must be validated after preparation via face and content validation. With such techniques, the survey is validated to see if it serves its purpose as intended, and it is also validated to ensure that the content of the survey is appropriate (Andrade et al., 2020). In doing so, 3-5 experts must independently rate each survey question as either satisfactory or unsatisfactory, and then the survey must be revised accordingly before recirculating the survey for further expert review (Andrade et al., 2020). Additionally, through administering the survey to a small number of volunteers in the target population before commencing the KAP study, insight into survey completion difficulty can be gained and suggestions for improvement can be incorporated into the KAP survey before study survey administration (Andrade et al., 2020).

While Andrade et al. (2020) provide extensive guidance for KAP survey validation, these techniques have been altered for the sake of this study. Firstly, questions were drawn from previously validated questionnaires and checklists from the works of Iyengar et al. (2011), Kinder & Sanders (1996), and COPS (2018) in order to draft the initial KAP survey related to racism, community, and policing. Studies by Fontaine et al. (2017), Rohe et al. (1997), Buchanan et al. (2021), Edwards, Lee, & Esposito (2018), Pierson et al. (2020), Gaston (2018), and Goutille (2009) also informed KAP survey question structure and content. The KAP survey was then reviewed by a single law enforcement and community expert from the *Center for Policing Equity (CPE)* for face and content validation. Then, 49 volunteers from the target population took the initially drafted KAP survey in a pilot test. The results of these pilot test surveys were analyzed in two parts in order to reduce the number of questions per hypothesized construct and ensure survey validity. First, Part 1, consisting of three sub-parts, focuses on both face validity relative to Attitude and Practice type questions. Part 1.1 outlines the initial identification of underlying components via Principal Component Analysis (PCA), and it also highlights the internal consistency within identified components via Cronbach's Alpha. With the reduced number of questions, Part 1.2 further investigates the factor loadings per identified component via a second round of PCA analysis, and the scale reliability per said component is also investigated once more with Cronbach's Alpha. Finally, in Part 1.3 the hypothesized constructs per question are compared with the actual question component grouping to confirm the final naming of the constructs to be tested with the survey. Then, in Part 2 face validity relative to Knowledge type questions is outlined via exam item analysis results involving point-biserial correlation testing. The final results and question selections from both Part 1 and Part 2 were lastly used as input for the final

Part 1 – Attitude & Practice Questions Validity

1.1 Initial Principal Component Analysis (PCA) & Cronbach's Alpha

First, with all test survey results relevant to hypothesized constructs of "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" dimension reduction was performed using principal component analysis (PCA) in *SPSS*. Seeing as all of these

construct-type questions used the same ordinal Likert Scale, they could be analyzed together. Given that the survey aimed to cover three hypothesized constructs, the analysis was extracted with a fixed number of three factors (i.e., components) to extract. Using guidelines outlined by Brown (2009), an initial oblique rotation (i.e., direct oblimin) was tested, and results yielded factor correlations below 0.32. Therefore, instead an orthogonal rotation method (varimax) was used to carry out further analysis, and as indicated by Brown (2009) it was assumed that the factors in the analysis were uncorrelated.

As advised by IBM (2021a), in order to confirm the suitability of the data for structure detection, both the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) and the Bartlett's Test of Sphericity were considered. As outlined by IBM (2021a), KMO indicates "the proportion of variance in the variables that might be caused by underlying factors". Results from this analysis indicated an adequate KMO value of 0.639 (i.e., KMO > 0.5). As additionally outlined by IBM (2021a), Bartlett's Test of Sphericity measured whether or not structure detection is possible given the relatedness of variables, thus proving this metric relevant for the analysis at hand. Results indicated an adequate value for such metric (i.e., $p < 0.05$). As such, dimension reduction was confirmed appropriate for this data set.

With the associated Rotated Component Matrix (Table A3), it was observed how different survey questions loaded to three different components respectively. Only factor loadings with a value greater than 0.3 were considered relevant for the analysis. Then, in order to confirm the question coherency and the reliability of each component, Cronbach's Alpha was calculated for all components. Considering the 10 questions which loaded to Component 1, analysis for Cronbach's alpha yielded $\alpha = 0.904$. Given the 8 questions which loaded to Component 2, analysis for Cronbach's alpha yielded $\alpha = 0.735$. Both of these alpha values were considered sufficient for measuring internal consistency. Finally, considering the four questions which loaded to Component 3, analysis for Cronbach's alpha yielded an inadequate $\alpha = 0.462$.

From this initial analysis, three questions for each component were selected and considered for a narrower version of the original KAP survey draft. These questions were selected based on the magnitude of their factor loadings, where questions with larger factor loadings were selected. Table A4 outlines the analyzed questions, the component number in which they loaded to, and the associated factor loadings. Questions highlighted in orange were selected from Component 1, questions highlighted in yellow were selected from Component 2, and questions highlighted in red were selected from Component 3 to be used in the next iteration of a PCA with a smaller draft survey.

Table A3: Rotated Component Matrix outcome from the initial Principal Component (PCA) analysis on all survey draft responses

Rotated Component Matrix			
	Component		
	1	2	3
I trust the police.	0.862	-0.112	-0.182
I have confidence in the police to address community concerns.	0.845	-0.085	0.092
I am comfortable communicating with the police.	0.789	0.048	-0.244
Police officers show concern for community members.	0.773	0.075	-0.088
I would contact the police for help.	0.719	0.011	0.211
Police officers work with community members to solve local problems.	0.718	-0.021	0.238
The police are legitimate and they are justified to exercise their power to maintain order and solve crimes. For this reason, people should obey and follow the rules and laws that the police enforce.	0.713	-0.085	-0.156
Police officers are responsive to the concerns of community members.	0.650	0.020	-0.162
Police officers are approachable and easy to talk to.	0.597	-0.026	-0.348
Irish reverse coding	-0.575	0.505	0.338
try harder reverse coding	-0.554	0.442	0.340
I trust community members and their efforts to collaborate with the police.	0.505	0.063	0.150
Generations of slavery and discrimination have created conditions that make it difficult for Black people to work their way out of the lower class.	-0.500	0.313	0.424
Over the past few years, Black people have gotten less than they deserve	-0.409	0.198	0.191
Black people are lazy	0.165	0.128	0.069
I feel warmly towards white people	0.037	0.789	-0.028
I feel warmly towards Black people	-0.389	0.736	-0.081
Black people are peaceful	-0.079	0.639	-0.085
Open reverse coding	0.061	0.505	0.048
I am comfortable communicating with community members within organizations that collaborate with the police.	0.394	0.465	0.149
Black people are dependent reverse coding	-0.006	0.316	0.195
I fear the police reverse coding	0.257	0.146	-0.775
I would contact community members within organizations that collaborate with the police for help with addressing community-specific problems.	-0.007	0.284	0.611
I have confidence in communities to offer aid to police for addressing community concerns.	0.337	0.101	0.609
I would participate in a community-oriented policing (COP) initiative locally if invited.	-0.083	-0.031	0.572

Extraction Method: Principle Component Analysis
Rotation Method: Varimax with Kaiser Normalization
a. Rotation converged in 5 iterations

Table A4: Outline of PCA dimension reduction analyzed questions where questions highlighted in orange were selected from component 1, questions highlighted in yellow were selected from component 2, and questions highlighted in red were selected from component 3 to be used in the next iteration of a PCA with a smaller draft survey.

#	Hypothesized Construct	Question	Scoring	Correct Answer	Source	Explanation of how the Question Measures the Hypothesized Construct
Q1_1	Attitude about Explicit Racism	I feel warmly towards Black people	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how warm/cold the participant feels towards Black people
Q1_2	Attitude about Explicit Racism	I feel warmly towards white people	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how warm/cold the participant feels towards white people
Q1_3	Attitude about Explicit Racism	Black people are lazy	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how hard working/lazy the participant perceives Black people
Q1_4	Attitude about Explicit Racism	Black people are peaceful	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how violent/peaceful the participant perceives Black people
Q1_5	Attitude about Explicit Racism	Black people are dependent	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how self-reliant/dependent the participant perceives black people
Q1_6	Attitude about Explicit Racism	Black people are closed to interact with people of different backgrounds	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how open/closed the participant perceives Black people to be towards people from different backgrounds
Q3_1	Attitude about Explicit Racism	Over the past few years, Black people have gotten less than they deserve	Likert Scale (1-5)	----	Kinder & Sanders, 1996	Explicitly measures racism on the basis of the participant's opinions related to the realities of systemic oppression
Q3_2	Attitude about Explicit Racism	Irish people, Italian people, Jewish people, Vietnamese people, and other minority groups in the United States overcame prejudice and worked their way up. Black people should do the same without any special favors	Likert Scale (1-5)	----	Kinder & Sanders, 1996	Explicitly measures racism on the basis of the participant's opinions related to the realities of systemic oppression
Q3_3	Attitude about Explicit Racism	It is really a matter of some people not trying hard enough; If Black people would only try harder they could be just as well off as white people	Likert Scale (1-5)	----	Kinder & Sanders, 1996	Explicitly measures racism on the basis of the participant's opinions related to the realities of systemic oppression
Q3_4	Attitude about Explicit Racism	Generations of slavery and discrimination have created conditions that make it difficult for Black people to work they way out of the lower class.	Likert Scale (1-5)	----	Kinder & Sanders, 1996	Explicitly measures racism on the basis of the participant's opinions related to the realities of systemic oppression
Q2_1	Attitude about policing in communities	The police are legitimate and they are justified to exercise their power to maintain order and solve crimes. For this reason, people should obey and follow the rules and laws that the police enforce.	Likert Scale (1-5)	----	Fontaine et al., 2017; COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to police legitimacy in communities
Q2_2	Attitude about policing in communities	Police officers are approachable and easy to talk to.	Likert Scale (1-5)	----	COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to how approachable police officers are to communities
Q2_3	Attitude about policing in communities	Police officers show concern for community members.	Likert Scale (1-5)	----	COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to police officers' perceived concern for community members
Q2_4	Attitude about policing in communities	Police officers are responsive to the concerns of community members.	Likert Scale (1-5)	----	COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to police officers' responsiveness towards the concerns of community members
Q2_5	Attitude about policing in communities	Police officers work with community members to solve local problems.	Likert Scale (1-5)	----	COPS, 2018	Measures the attitudes that a participant has about policing and communities relevant to police-citizen co-production efforts
Q4	Knowledge about policing in communities	Community-oriented policing (COP) can be broadly defined as police officers and residents working together to solve local problems.	True/ False/ I Don't Know	TRUE	Rohe et al., 1997; COPS, 2018	Measures a participant's level of knowledge about policing and communities on the basis of what COP is
Q5	Knowledge about policing in communities	Community-oriented policing (COP) is a top-down approach to policing.	True/ False/ I Don't Know	FALSE	Rohe et al., 1997	Measures a participant's level of knowledge about policing and communities on the basis of what COP is
Q6	Knowledge about policing in communities	While community-oriented policing (COP) can address community-specific issues, COP can also be used to solve long-term endemic problems, such as poverty and unemployment.	True/ False/ I Don't Know	FALSE	Rohe et al., 1997; COPS, 2018; Goutille, 2009	Measures a participant's level of knowledge about policing and communities on the basis of what COP is and what its limitation are
Q7	Knowledge about policing in communities	Historically, law enforcement has been entangled with white supremacy and associated hate groups.	True/ False/ I Don't Know	TRUE	Buchanan et al., 2021	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q8	Knowledge about policing in communities	Racial disparities in public safety still exist today.	True/ False/ I Don't Know	TRUE	Center for Policing Equity, 2021	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q9	Knowledge about policing in communities	Folks that are Black, Indigenous, and People of Color (BIPOC) are less likely to experience police-involved harm than white people.	True/ False/ I Don't Know	FALSE	Edwards, Lee, & Esposito, 2018	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q10	Knowledge about policing in communities	White people are more likely to be stopped by the police than folks that are Black, Indigenous, and People of Color (BIPOC).	True/ False/ I Don't Know	FALSE	Pierson et al., 2020	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q11	Knowledge about policing in communities	In some communities, residents do not trust the police to treat them fairly.	True/ False/ I Don't Know	TRUE	Rohe et al., 1997	Measures a participant's level of knowledge about policing and communities on the basis of the historic distrust between different communities and police officers in which has been fueled by systemic oppression
Q12	Knowledge about policing in communities	Some police officers believe that they are disliked by some communities, and these police officers express a low degree of confidence towards these communities to come to their aid during a crisis.	True/ False/ I Don't Know	TRUE	Fontaine et al., 2017; Rohe et al., 1997; COPS, 2018	Measures a participant's level of knowledge about policing and communities on the basis of the historic distrust between different communities and police officers in which has been fueled by systemic oppression
Q13	Knowledge about policing in communities	In communities where folks are predominantly Black, Indigenous, and People of Color (BIPOC), laws are less strictly enforced by police than in communities where residents are predominantly white.	True/ False/ I Don't Know	FALSE	Gaston, 2018	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q2_6	Practice about policing in communities	I fear the police	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much they fear police
Q2_7	Attitude about policing in communities	I trust the police.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much they trust police
Q2_8	Practice about policing in communities	I trust community members and their efforts to collaborate with the police.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much they trust community members and efforts to foster collaboration with police
Q2_9	Attitude about policing in communities	I have confidence in the police to address community concerns.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much confidence that they have in police
Q2_10	Practice about policing in communities	I have confidence in communities to offer aid to police for addressing community concerns.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much confidence they have in communities to offer aid to the police in order to address community concerns
Q2_11	Attitude about policing in communities	I am comfortable communicating with the police.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are communicating with police
Q2_12	Attitude about policing in communities	I am comfortable communicating with community members within organizations that collaborate with the police.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards how comfortable they are communicating with community members and organizations that collaborate with police
Q2_13	Practice about policing in communities	I would contact the police for help.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are reaching out to the police for help
Q2_14	Practice about policing in communities	I would contact community members within organizations that collaborate with the police for help with addressing community-specific problems	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are reaching out for help to community members and organizations that collaborate with the police
Q2_15	Practice about policing in communities	I would participate in a community-oriented policing (COP) initiative locally if invited.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how willing they would be to participate in a local COP initiative

1.2 Final Principal Component Analysis (PCA) & Cronbach's Alpha

Then, with the selected questions from Part 1.1 (outlined in Table A4), dimension reduction via PCA was performed once more in *SPSS*. Again, the analysis was extracted with a fixed number of three factors (i.e., components). With these three extracted components, nearly 70% the variability was explained. Once more, an initial oblique rotation (i.e., direct oblimin) was tested, and results yielded factor correlations below 0.32. Therefore, instead an orthogonal rotation method (varimax) was used to carry out further analysis.

Results yielded KMO = 0.624 and an adequate significance value for the Bartlett's Test of Sphericity (i.e., $p < 0.05$), thus confirming the suitability of the data for structure detection. With the associated Rotated Component Matrix (Table A5), it was once again observed how different survey questions loaded to three different components respectively. All factor loadings were confirmed to be sufficient in magnitude. Then, in order to confirm the question coherency and the reliability of each component once more, Cronbach's Alpha was calculated for all components.

Considering the 3 questions which loaded to Component 1, analysis for Cronbach's alpha yielded $\alpha = 0.861$. Given the 3 questions which loaded to Component 2, analysis for Cronbach's alpha yielded $\alpha = 0.745$. Both of these alpha values were considered sufficient for measuring internal consistency. Finally, considering the 3 questions which loaded to Component 3, analysis for Cronbach's alpha yielded an $\alpha = 0.528$. While most guidelines recommend a necessary Cronbach's Alpha's coefficient of 0.6 or greater, Component 3 was still considered somewhat internally consistent for the sake of this study given the associated logistical constraints and inability to redraft and distribute relevant survey questions. However, the low scoring value was noted, and the associated limitations of such implications were considered upon analysis of survey results. Table A6 below outlines the analyzed and final survey questions, the component number in which they loaded to, and the associated factor loadings. Questions highlighted in orange are associated with Component 1, questions highlighted in yellow are associated with Component 2, and questions highlighted in red are associated with Component 3.

Table A5: Rotated Component Matrix outcome from the final Principal Component (PCA) analysis on Part 1.1 selected survey draft responses

Rotated Component Matrix			
	Component		
	1	2	3
I trust the police.	0.910	-0.049	-0.129
I have confidence in the police to address community concerns.	0.881	-0.117	0.233
I am comfortable communicating with the police.	0.826	-0.043	-0.105
I feel warmly towards white people	0.095	0.883	0.107
I feel warmly towards Black people	-0.351	0.790	0.026
Black people are peaceful	-0.021	0.748	0.017
I would contact community members within organizations that collaborate with the police for help with addressing community-specific problems.	-0.072	0.183	0.738
I would participate in a community-oriented policing (COP) initiative locally if invited.	-0.166	-0.115	0.714
I have confidence in communities to offer aid to police for addressing community concerns.	0.375	0.104	0.685

Extraction Method: Principle Component Analysis
Rotation Method: Varimax with Kaiser Normalization
a. Rotation converged in 5 iterations

Table A6: Outline of PCA dimension reduction analyzed and final survey questions where questions highlighted in orange are associated with Component 1, questions highlighted in yellow are associated with Component 2, and questions highlighted in red are associated with Component 3

#	Hypothesized Construct	Question	Scoring	Correct Answer	Source	Explanation of how the Question Measures the Hypothesized Construct
Q1_1	Attitude about Explicit Racism	I feel warmly towards Black people	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how warm/cold the participant feels towards Black people
Q1_2	Attitude about Explicit Racism	I feel warmly towards white people	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how warm/cold the participant feels towards white people
Q1_4	Attitude about Explicit Racism	Black people are peaceful	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how violent/peaceful the participant perceives Black people
Q2_7	Attitude about policing in communities	I trust the police.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much they trust police
Q2_9	Attitude about policing in communities	I have confidence in the police to address community concerns.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much confidence that they have in police
Q2_10	Practice about policing in communities	I have confidence in communities to offer aid to police for addressing community concerns.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much confidence they have in communities to offer aid to the police in order to address community concerns
Q2_11	Attitude about policing in communities	I am comfortable communicating with the police.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are communicating with police
Q2_14	Practice about policing in communities	I would contact community members within organizations that collaborate with the police for help with addressing community-specific problems	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are reaching out for help to community members and organizations that collaborate with the police
Q2_15	Practice about policing in communities	I would participate in a community-oriented policing (COP) initiative locally if invited.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how willing they would be to participate in a local COP initiative

1.3 A Comparison of Hypothesized Constructs & Actual Components

Finally, with three different questions selected per component for the final KAP survey, the hypothesized constructs were relevant to attitude and practice were revisited. Given that the selected questions for Component 2 were all hypothesized to be "Attitude about Explicit Racism", this construct was selected to represent this cluster of questions in the final KAP survey. While all questions clustered as both Component 1 and Component 3 were originally hypothesized to be related to "Practice about Policing in Communities", further inspection resulted in apparent differences in these highlighted questions. While originally hypothesized to be related to "Practice about Policing in Communities", the Component 1 questions all distinctly pointed to attitudes that individuals' have about police officers in their communities. As such, the Component 1 construct has been re-labelled as "Attitude about Policing in Communities". Similarly, the Component 3 questions all distinctly pointed towards community involvement and practice in local policing. As such, the Component 3 construct has been determined to be the true "Practice about Policing in Communities" construct, for it points to individuals' behaviour and practice about policing. Once components were labelled with constructs, further validation of label appropriateness was conducted by a law enforcement and community expert from *CPE*. Here, it was determined that all posed questions were appropriate in terms of face validity.

Part 2 – Knowledge Questions Validity

In order to determine the appropriateness of questions relevant to "Knowledge about Policing in Communities", exam item analysis involving point-biserial correlation testing was conducted in *SPSS*. Seeing as all questions relevant to "Knowledge about Policing in Communities" was scored dichotomously as either "correct" or "incorrect", such hypothesized construct had to be analyzed separately from the other hypothesized constructs of "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" which all used the same ordinal Likert Scale. In doing so and as advised by IBM (2021b), the special case of the Pearson product moment correlation was applied to dichotomous question variables and a continuous total score variable.

In order to determine the most appropriate knowledge-based questions to include in the final survey, the relationship between total score on all knowledge-based questions on a scale from 0-100 and score per each individual question (correct = 1 and incorrect = 0) was assessed. The output from this analysis is displayed below in Table A7. Results indicated that *Q7* (r_{pb} = .665, p < 0.01), *Q8* (r_{pb} = .631, p < 0.01, and *Q10* (r_{pb} = .590, p < 0.01) were best at discriminating between pilot survey participants who were highly knowledgeable about policing in communities from pilot survey participants who were not knowledgeable about policing in communities. As such and as indicated with green in Table A8, these three questions were chosen to represent the "Knowledge about Policing in Communities" construct for the validated survey. This question selection was further validated for content appropriateness by a law enforcement and community expert from *CPE*. Here, it was determined that all posed questions were appropriate in terms of face validity.

Table A7: Exam item analysis with point-biserial correlation testing output correlations

		Correlations										
		Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Total Score
Q4	Pearson Correlation	1	0.275	-0.084	0.259	0.169	0.110	0.130	0.164	.320*	.343*	.574**
	Sig. (2-tailed)		0.056	0.565	0.072	0.244	0.454	0.374	0.260	0.025	0.016	0.000
	N	49	49	49	49	49	49	49	49	49	49	49
Q5	Pearson Correlation	0.275	1	-0.160	0.207	0.052	0.092	0.181	0.111	0.148	-0.016	.397**
	Sig. (2-tailed)	0.056		0.271	0.153	0.724	0.528	0.212	0.448	0.309	0.913	0.005
	N	49	49	49	49	49	49	49	49	49	49	49
Q6	Pearson Correlation	-0.084	-0.160	1	0.073	0.111	-0.061	-0.146	0.062	0.248	-0.131	0.120
	Sig. (2-tailed)	0.565	0.271		0.620	0.446	0.675	0.318	0.675	0.086	0.369	0.410
	N	49	49	49	49	49	49	49	49	49	49	49
Q7	Pearson Correlation	0.259	0.207	0.073	1	.361*	.509**	.303*	0.271	0.142	.377**	.665**
	Sig. (2-tailed)	0.072	0.153	0.620		0.011	0.000	0.035	0.060	0.329	0.008	0.000
	N	49	49	49	49	49	49	49	49	49	49	49
Q8	Pearson Correlation	0.169	0.052	0.111	.361*	1	.446**	.491**	.552**	.450**	0.102	.631**
	Sig. (2-tailed)	0.244	0.724	0.446	0.011		0.001	0.000	0.000	0.001	0.486	0.000
	N	49	49	49	49	49	49	49	49	49	49	49
Q9	Pearson Correlation	0.110	0.092	-0.061	.509**	.446**	1	.670**	.410**	0.072	0.170	.555**
	Sig. (2-tailed)	0.454	0.528	0.675	0.000	0.001		0.000	0.003	0.621	0.242	0.000
	N	49	49	49	49	49	49	49	49	49	49	49
Q10	Pearson Correlation	0.130	0.181	-0.146	.303*	.491**	.670**	1	.612**	0.132	0.182	.590**
	Sig. (2-tailed)	0.374	0.212	0.318	0.035	0.000	0.000		0.000	0.367	0.211	0.000
	N	49	49	49	49	49	49	49	49	49	49	49
Q11	Pearson Correlation	0.164	0.111	0.062	0.271	.552**	.410**	.612**	1	0.248	0.194	.576**
	Sig. (2-tailed)	0.260	0.448	0.675	0.060	0.000	0.003	0.000		0.085	0.182	0.000
	N	49	49	49	49	49	49	49	49	49	49	49
Q12	Pearson Correlation	.320*	0.148	0.248	0.142	.450**	0.072	0.132	0.248	1	0.199	.589**
	Sig. (2-tailed)	0.025	0.309	0.086	0.329	0.001	0.621	0.367	0.085		0.171	0.000
	N	49	49	49	49	49	49	49	49	49	49	49
Q13	Pearson Correlation	.343*	-0.016	-0.131	.377**	0.102	0.170	0.182	0.194	0.199	1	.530**
	Sig. (2-tailed)	0.016	0.913	0.369	0.008	0.486	0.242	0.211	0.182	0.171		0.000
	N	49	49	49	49	49	49	49	49	49	49	49
Total Score	Pearson Correlation	.574**	.397**	0.120	.665**	.631**	.555**	.590**	.576**	.589**	.530**	1
	Sig. (2-tailed)	0.000	0.005	0.410	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	N	49	49	49	49	49	49	49	49	49	49	49

*. Correlation is significant at the 0.05 level (2-tailed)
**. Correlation is significant at the 0.01 level (2-tailed)

Table A8: Final selected "Knowledge about Policing in Communities" questions for validated survey

#	Hypothesized Construct	Question	Scoring	Correct Answer	Source	Explanation of how the Question Measures the Hypothesized Construct
Q4	Knowledge about policing in communities	Community-oriented policing (COP) can be broadly defined as police officers and residents working together to solve local problems.	True/ False/ I Don't Know	TRUE	Rohe et al., 1997; COPS, 2018	Measures a participant's level of knowledge about policing and communities on the basis of what COP is
Q5	Knowledge about policing in communities	Community-oriented policing (COP) is a top-down approach to policing.	True/ False/ I Don't Know	FALSE	Rohe et al., 1997	Measures a participant's level of knowledge about policing and communities on the basis of what COP is
Q6	Knowledge about policing in communities	While community-oriented policing (COP) can address community-specific issues, COP can also be used to solve long-term endemic problems, such as poverty and unemployment.	True/ False/ I Don't Know	FALSE	Rohe et al., 1997; COPS, 2018; Goutille, 2009	Measures a participant's level of knowledge about policing and communities on the basis of what COP is and what its limitation are
Q7	Knowledge about policing in communities	Historically, law enforcement has been entangled with white supremacy and associated hate groups.	True/ False/ I Don't Know	TRUE	Buchanan et al., 2021	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q8	Knowledge about policing in communities	Racial disparities in public safety still exist today.	True/ False/ I Don't Know	TRUE	Center for Policing Equity, 2021	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q9	Knowledge about policing in communities	Folks that are Black, Indigenous, and People of Color (BIPOC) are less likely to experience police-involved harm than white people.	True/ False/ I Don't Know	FALSE	Edwards, Lee, & Esposito, 2018	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q10	Knowledge about policing in communities	White people are more likely to be stopped by the police than folks that are Black, Indigenous, and People of Color (BIPOC).	True/ False/ I Don't Know	FALSE	Pierson et al., 2020	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q11	Knowledge about policing in communities	In some communities, residents do not trust the police to treat them fairly.	True/ False/ I Don't Know	TRUE	Rohe et al., 1997	Measures a participant's level of knowledge about policing and communities on the basis of the historic distrust between different communities and police officers in which has been fueled by systemic oppression
Q12	Knowledge about policing in communities	Some police officers believe that they are disliked by some communities, and these police officers express a low degree of confidence towards these communities to come to their aid during a crisis.	True/ False/ I Don't Know	TRUE	Fontaine et al., 2017; Rohe et al., 1997; COPS, 2018	Measures a participant's level of knowledge about policing and communities on the basis of the historic distrust between different communities and police officers in which has been fueled by systemic oppression
Q13	Knowledge about policing in communities	In communities where folks are predominantly Black, Indigineous, and People of Color (BIPOC), laws are less strictly enforced by police than in communities where residents are predominantly white.	True/ False/ I Don't Know	FALSE	Gaston, 2018	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems

A.3 Final KAP Survey Questions

After analysing, selecting, and validating Knowledge, Attitude, and Practice (KAP) questions related to racism, community, and policing, Table A9 below summarizes the final selection of validated questions for KAP surveying in this study.

Table A8: All final selected questions for validated KAP survey

#	Hypothesized Construct	Question	Scoring	Correct Answer	Source	Explanation of how the Question Measures the Hypothesized Construct
Q1_1	Attitude about Explicit Racism	I feel warmly towards Black people	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how warm/cold the participant feels towards Black people
Q1_2	Attitude about Explicit Racism	I feel warmly towards white people	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how warm/cold the participant feels towards white people
Q1_4	Attitude about Explicit Racism	Black people are peaceful	Likert Scale (1-5)	----	Iyengar et al., 2011	Explicitly measures racism on the basis of how violent/peaceful the participant perceives Black people
Q7	Knowledge about policing in communities	Historically, law enforcement has been entangled with white supremacy and associated hate groups.	True/ False/ I Don't Know	TRUE	Buchanan et al., 2021	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q8	Knowledge about policing in communities	Racial disparities in public safety still exist today.	True/ False/ I Don't Know	TRUE	Center for Policing Equity, 2021	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q10	Knowledge about policing in communities	White people are more likely to be stopped by the police than folks that are Black, Indigenous, and People of Color (BIPOC).	True/ False/ I Don't Know	FALSE	Pierson et al., 2020	Measures a participant's level of knowledge about policing and communities on the basis of the historic and systemic racism that has been intertwined with police systems
Q2_7	Attitude about policing in communities	I trust the police.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much they trust police
Q2_9	Attitude about policing in communities	I have confidence in the police to address community concerns.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much confidence that they have in police
Q2_10	Practice about policing in communities	I have confidence in communities to offer aid to police for addressing community concerns.	Likert Scale (1-5)	----	COPS, 2018; Goutille, 2009	Measures a participant's practice/behaviour towards policing in communities relevant to how much confidence they have in communities to offer aid to the police in order to address community concerns
Q2_11	Attitude about policing in communities	I am comfortable communicating with the police.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are communicating with police
Q2_14	Practice about policing in communities	I would contact community members within organizations that collaborate with the police for help with addressing community-specific problems	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how comfortable they are reaching out for help to community members and organizations that collaborate with the police
Q2_15	Practice about policing in communities	I would participate in a community-oriented policing (COP) initiative locally if invited.	Likert Scale (1-5)	----	COPS, 2018	Measures a participant's practice/behaviour towards how willing they would be to participate in a local COP initiative

APPENDIX B

This appendix outlines the original 33 question Marlowe-Crowne Social Desirability Scale (MC-SDS) for self-reported concern for social approval. The short version (13 questions) of this scale, developed by Sârbescu, Costea, & Rusu (2012), was used to measure all participants' self-reported concern for social approval before workshop activities commenced. In gathering this information from participants, issues for social desirability can be accounted for, and better interpretation of experimental results is ensured. Specifics of each question are outlined below in Table B1, and short version questions are highlighted in green.

Table B1: Marlowe-Crowne Social Desirability Scale (MC-SDS) question specifics as outlined by Crowne & Marlowe (1960) and shortened by Sârbescu, Costea, & Rusu (2012) in green

Question #	Question	Scoring	Answer in Which Indicates High Social Desirability
1	Before voting, I thoroughly investigate the qualifications of all candidates.	True/False	TRUE
2	I never hesitate to go out of my way to help someone in trouble.	True/ False	TRUE
3	It is sometimes hard for me to go on with my work if I am not encouraged.	True/ False	FALSE
4	I have never intensely disliked someone.	True/ False	TRUE
5	On occasions I have had doubts about my ability to succeed in life.	True/ False	FALSE
6	I sometimes feel resentful when I don't get my way.	True/ False	FALSE
7	I am always careful about my manner of dress.	True /False	TRUE
8	My table manners at home are as good as when I eat out in a restaurant.	True /False	TRUE
9	If I could get into a movie without paying and be sure I was not seen, I would probably do it.	True /False	FALSE
10	True /False	True /False	FALSE
11	I like to gossip at times.	True /False	FALSE
12	There have been times when I felt like rebelling against people in authority even though I knew that they were right.	True /False	FALSE
13	No matter who I'm talking to, I'm always a good listener.	True /False	TRUE
14	I can remember "playing sick" to get out of something.	True /False	FALSE
15	There have been occasions when I have taken advantage of someone.	True /False	FALSE
16	I'm always willing to admit it when I make a mistake.	True /False	TRUE
17	I always try to practice what I preach.	True /False	TRUE
18	I don't find it particularly difficult to get along with loudmouthed, obnoxious people.	True /False	TRUE
19	I sometimes try to get even rather than forgive and forget.	True /False	FALSE
20	When I don't know something I don't mind at all admitting it.	True /False	TRUE
21	I am always courteous, even to people who are disagreeable.	True /False	TRUE
22	At times I have really insisted on having things my own way.	True /False	FALSE
23	There have been occasions when I felt like smashing things.	True /False	FALSE
24	I would never think of letting someone else be punished for my wrong-doings.	True /False	TRUE
25	I never resent being asked to return a favor.	True /False	TRUE
26	I have never been irked when people expressed ideas very different from my own.	True /False	TRUE
27	I never make a long trip without checking the safety of my car.	True /False	TRUE
28	There have been times when I was quite jealous of the good fortune of others.	True /False	FALSE
29	I have almost never felt the urge to tell someone off.	True /False	TRUE
30	I am sometimes irritated by people who ask favors of me.	True /False	FALSE
31	I have never never felt that I was punished without cause.	True /False	TRUE
32	I sometimes think when people have a misfortune they only got what they deserved.	True /False	FALSE
33	I have never deliberately said something that hurt someone's feelings.	True /False	TRUE

APPENDIX C

This appendix outlines the Observation Protocol used by *CPE* appointed observers during workshop and focus group observation. This protocol was curated via Onwuegbuzie et al.'s (2009) qualitative framework.

Observation Protocol

Observation protocol for both 1) serious game design workshop activities and 2) the semi-structured interview question session is outlined below. Most important to this research is to capture NONVERBAL data about each participant. You can jot down field notes about the overall flow of the workshop, **but focus on body language, overall agreement/disagreement, and response patterns**. The specifics of each of these elements is briefly outlined below.

C.1 SGD Activity

C.1.1 SGD Activity Field Notes

- Here, include notes on what is said and the overall flow of the workshop
- Focus on nonverbal data, such as individual participants' **overall attitude** and **body language** (i.e., does someone look disgruntled, upset, frustrated, focused, etc.)
- Focus on overall **response patterns** - i.e., does a particular demographic (e.g., young white female [y-w-f]) seem to be dominating the conversation? Are there some people who are less enthusiastic to participate?
- **Your notes here……..**

C.2 Semi-Structured Interview Question Session

C.2.1 Semi-Structured Interview Questions Field Notes

- Here, include notes on what is said and the overall flow of the workshop, but primarily focus on filling out the provided **Observation Table**

- **Your notes here……..**

C.2.2 Observation Table

Successfully taking notes in the Observation Table is critical to ensuring good research results. To help make sure this procedure is clear, an example Observation Table has been filled out below with example data in yellow highlighted text. Comments have also been made for brief explanation of the recorded information. Use this example as a guideline for taking notes in the actual Observation Table below. A **KEY** outlines the type of information to be recorded in the Observation Table. "**Demographics**", "**Response**", and "**Agreement**" related info is CRITICAL. Taking notes on "**Additional Indicators**" is helpful, but secondary.

With the Observation Table, use the provided keys below (or create your own coded shorthand) to make note of:

- Participant demographics (e.g., young white female, y-w-f)
- Order in which the participants responded to the question
- Agreement/disagreement for each participant on what is said during each question discussion
- Whether or not a participant responded at all or multiple times to a question
- Any additional notes on speech speed, speech tone, body language, etc. (see the key below for examples)

Example Observation Table Note Taking

Table C1: Example observation table for note taking

Question		Member 1 JD y-w-f	Member 2	Member 3	Member 3	Member 4
EXAMPLE	1. What was the biggest challenge for your team when designing your serious game concept? How did you overcome this challenge?	*1	*2 SE	NR X	*3 D FFF PPP	

Actual Observation Table Note Taking

Table C2: Actual observation table note taking

Question		Member 1	Member 2	Member 3	Member 3	Member 4
Debrief	1. What was the biggest challenge for your team when designing your serious game concept? How did you overcome this challenge?					
Debrief	2. What was the biggest learning moment for you and your team while designing your serious game concept? What made this event so impactful?					
Debrief	3. How did you support one another to achieve success throughout the serious game design process?					
Debrief	4. How do you feel that your experience today relates to the real-world?					
Debrief	5. Is there anything else you would like to say about your experience participating in these workshops?					
Evaluation	6. How do you think the involvement of both community members and law enforcement officers in these workshops impacted the outcome of your serious game concept?					
Evaluation	7. Do you think that the serious game conceptual design workshop encouraged participants to openly explore aspects of racial bias in policing? Please explain.					
Evaluation	8. Do you think that serious game design based workshops could be effective in improving police-community relationships and improving overall community satisfaction with police? Why?					
Evaluation	9. How could these serious game conceptual design workshops be improved for better police-community relationship outcomes?					

APPENDIX D

Validated group semi-structured interview questions drawn and adapted from: Gomez's (2016) study on policing, community fragmentation, and public health; COPS (2018) community survey on public safety and law enforcement; and Winn's (2009) game development process evaluation survey.

Group semi-structured interviews were used to collect qualitative data about the effectiveness of a police-community serious game design process for reduced police bias. Notably, given the sensitive nature of racism and policing, only written notes about the question response content were made in recording group semi-structured interviews. These notes were then used for analysis. In administering these interviews, *SQ4* was addressed via qualitative means. More specifically, however, the group semi-structured interview questions included in this research have been drafted in such a way to address *SQ5* and *SQ6* from a qualitative perspective. Results of this research activity have led to conclusions about reducing police bias and improving police-community relations through participatory serious game design. In addition, conclusions about participatory serious game design as a means for societal intervention in general have been deduced from information collected during this research activity. While the qualitative data collected with group semi-structured interview questions is enriching in and of itself, this data has also been used as a means to enhance survey data. Group semi-structured interviews about participants' experience with the serious game design intervention for reduced police bias have been used to further understand the KAP study results regarding community, policing, and racism. Triangulating interview data with focus group observations has also been used to further explain workshop observed phenomena, providing an even deeper understanding of the success of the serious game design intervention. Thus, the interview questions used in this analysis have been crafted to further develop an understanding of how participants experienced the serious game conceptual design activities as a means for COP intervention. In drafting semi-structured interview questions, *SQ4* is directly addressed, qualitative insight into *SQ5* and *SQ6* is generated, and an overall understanding of the *Main Research Question* is developed (see Figure C1).

Table D1 outlines the group semi-structured interview questions used in this study, the research questions in which the group semi-structured interview questions relate to, and how the posed interview questions contribute to addressing the related research questions.

Main Research Question

How can a participatory serious game design process for reduced police bias, involving both communities and police, be developed in the United States as a means for societal intervention?

Sub-Research Question

SQ1: *How can police bias be defined?*

SQ2: *Based on theory and practice, what are successful approaches to involving community members in interventions for reduced police bias that can inform a participatory serious game approach to intervention?*

SQ3: *Which serious game design methods are suitable for participatory serious game design processes between communities and police forces in the United States?*

SQ4: *What methods are suitable for measuring the effectiveness of a police-community serious game design process for reduced police bias?*

SQ5: *What can the results from this study conclude about reducing police bias and improving police-community relations through participatory serious game design-based intervention?*

SQ6: *What can the outcomes from this study deduce about participatory serious game design as a means for societal intervention in general?*

Figure D1: Highlighted research questions addressed with semi-structured interviews

Table D1: Outlined group semi-structured interview questions, related research questions, and an explanation as to how each interview question contributes to addressing the identified research question

	Semi-Structured Interview Question	Related Research Question	Explanation of Semi-Structured Interview Question Contribution
1	How do you think the involvement of both community members and law enforcement officers in these workshops could impact the outcome of a serious game concept?	*SQ5; SQ6; Main Research Question*	This group semi-structured interview question has been drafted with the intent of understanding how participants perceive the effectiveness of involvement of both community members and law enforcement officers/associates in the same SGD workshop. In posing this question, the utility of SGD workshops for actual game concept generation and as a COP method can be developed. Insight generated with this question can also provide information about using participatory SGD for interventions in general.
2	Do you think that the serious game conceptual design workshop encouraged participants to openly explore aspects of racial bias in policing? Please explain.	*SQ5; SQ6; Main Research Question*	With this group semi-structured interview question, the participants' perceived effectiveness of the SGD workshop to encourage participants to openly explore aspects of racial bias in policing is identified. In gathering this information, conclusions can be generated about how the workshop environment encouraged/discouraged discourse about community-police relations and racial bias in policing. Insight generated with this question can also provide information about using participatory SGD for interventions in general.
3	Do you think that serious game design based workshops could be effective in improving police-community relationships and improving overall community satisfaction with police? Why?	*SQ5; SQ6; Main Research Question*	This group semi-structured interview question has been drafted with the intent of understanding how the SGD workshop contributed to improving police-community relations. By directly asking the interviewees whether or not they believe that SGD activities can be used as effective means for improved police-community relationships and overall community satisfaction with police, the potential success of SGD workshops for intervention in COP environments can be generated. Insight generated with this question can also provide information about using participatory SGD for interventions in general.
4	How could these serious game conceptual design workshops be improved for better police-community relationship outcomes?	*SQ5; SQ6; Main Research Question*	This group semi-structured interview question can generate insight into current limitations of the SGD approach to COP intervention. It can also provide recommendations for future work. In asking interviewees this question, the shortcomings of SGD workshops for improved police-community relations can be identified. Simultaneously, insight into how to better develop SGD activities between communities and police for societal intervention can also be made. Insight generated with this question can also provide information about using participatory SGD for interventions in general.

APPENDIX E

This appendix outlines Winn's (2009) Design, Play, and Experience (DPE) Framework and the Art of Serious Game Design method, supporting SGD activity resources, and the workshop session descriptions. First, the DPE Framework and the Art of Serious Game Design method are introduced. Then, supporting SGD activity resources are outlined. Finally, taking into account both Winn's (2009) SGD framework and method and the supporting SGD activity resources, the workshop session description is defined.

E.1 Design, Play, and Experience (DPE) Framework and the Art of Serious Game Design Method by Winn (2009)

Originally conceptualized by Winn (2009) as the Design, Play, and Experience (DPE) Framework for serious games, this work was later adapted into The Art of Serious Game Design method for participatory game design processes. With this method, facilitated workshops guide interdisciplinary teams of participants towards serious game conceptualization and design with brainstorming and prototyping activities (Winn, 2009). This method highlights the importance of balancing four interconnected and equally important elements: 1. Learning; 2. Storytelling; 3. Gameplay; and 4. User Experience (Winn, 2009). Within each of these three elements (i.e., 1. Learning; 2. Storytelling; 3. Gameplay; and 4. User Experience), symbolic meaning is attributed to designer/player concepts of: 1. Design Story; 2. Play; and 3. Experience. Further explanation of each method element and designer/player concept can be found in Table 4. A visual representation of this SGD design methodology and all of its contributing elements is also depicted in Figure 5. In actually implementing this method in practice, interdisciplinary workshop teams take on steps of 1. Brainstorming 1; 2. Brainstorming 2; and 3. Paper Prototyping.

Table E1: The Art of Serious Game Design method element explanations adapted from Winn, 2009

Concept	Explanation
Learning	"refers to the content to be learned by players through the game with specific and measurable learning outcomes" (p. 7)
Storytelling	"refers to the background story of the game and includes a description of the character(s), the setting, and the ultimate goal of the game" (p. 7)
Gameplay	"refers to the way in which the player interacts with the game, or with other players (if a multiplayer game). It encapsulates the type of activity (e.g., puzzle, trivia, etc.) found in the game" (p. 7)
User Experience	"refers to the player's emotions and attitudes while playing the game, as well as how the player interacts with the game" (p. 7)
Design	"symbolizes the designer's "story," or all the elements that the designer introduces to the game that will allow the player to play the game" (p. 7)
Play	"symbolizes the "mediated experience" between the player and design input through play, and is influenced by the design and the player's background brought into the play experience" (p. 7)
Experience	"symbolizes the varying play experiences that players can have depending on their backgrounds, as well as the choices and actions made during game play" (p. 7)

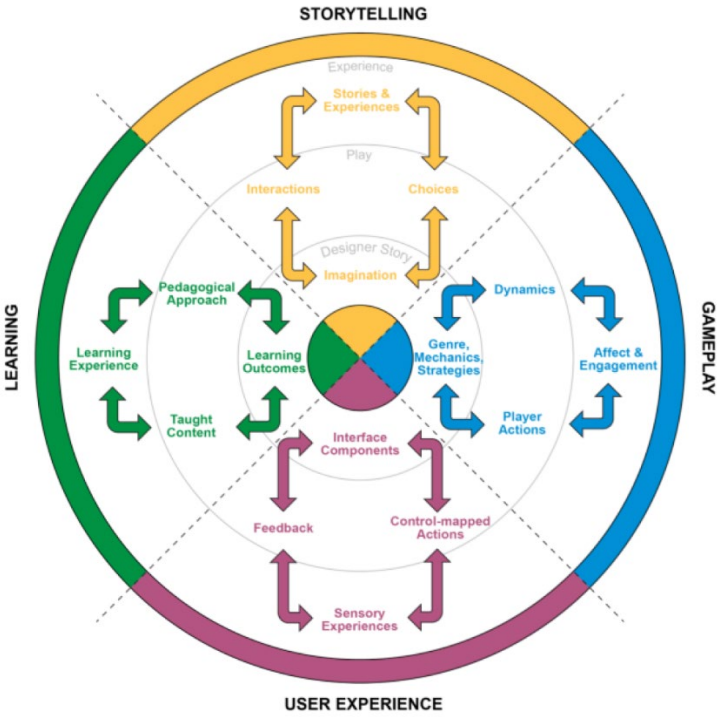


Figure E1: The Art of Serious Game Design methodology circle as adapted from the Design, Play, and Experience (DPE) Framework by Winn, 2009

1. Brainstorming 1

During the first step of Winn's (2009) approach, workshop participants that are deemed as "subject matter experts" are guided through a brainstorming session with the use of ideation cards to determine gameplay elements. During this first session, the goal is to have the participants answer as many ideation card questions as possible, so to jumpstart the serious game concept (Winn, 2009). During this brainstorming process, one team member is tasked with the responsibility of drawing each ideation card from the deck and reading its content aloud to the rest of the team (Winn, 2009). As a team, the participants then discuss their answers to the ideation card as a group until they have agreed upon a final response (Winn, 2009). Final responses are then marked on a colour coordinated sticky note and mapped onto the methodology circle (Figure 5).

2. Brainstorming 2

Similar to step 1 of this method, step 2 involves further game concept brainstorming (Winn, 2009). However, during this session, additional participants join the "subject matter experts" in game ideation (Winn, 2009). During this session, the "subject matter experts" first brief the new participants about the outcome of their Brainstorming 1 session (Winn, 2009). Then, a new set of ideation cards are used in a similar fashion to that of step 1, and team responses are ultimately mapped once more to the methodology circle (Figure 5). Ultimately, Brainstorming 2 is an iteration of Brainstorming 1, and completion of this step results in a concrete game concept (Winn, 2009).

3. Paper Prototyping

After the concrete game concept has been determined, technological considerations for the actual developed game are first discussed before paper prototyping commences (Winn, 2009). During paper prototyping, workshop participants come together to draw out the game concept, as it would appear in a developed game interface (Winn, 2009). Group debriefing follows this process (Winn, 2009).

E.2 Supporting SGD Resources

Supporting SGD resources taken from Winn's (2009) DPE Framework and Art of Serious Game Design method are outlined below. Group debriefing focus group questions and semi-structured interview questions relevant to evaluation have been adapted from: Krueger's (2014) focus group guide; Breen's (2006) practical guide to focus group research; and Winn's (2009) game development process evaluation survey and are also listed. Winn's (2009) The Art of Serious Game Design methodology circle, brainstorming (1 and 2) cards, and glossary have been used according to the prescribed procedure in D1.

E.2.1 Methodology Circle

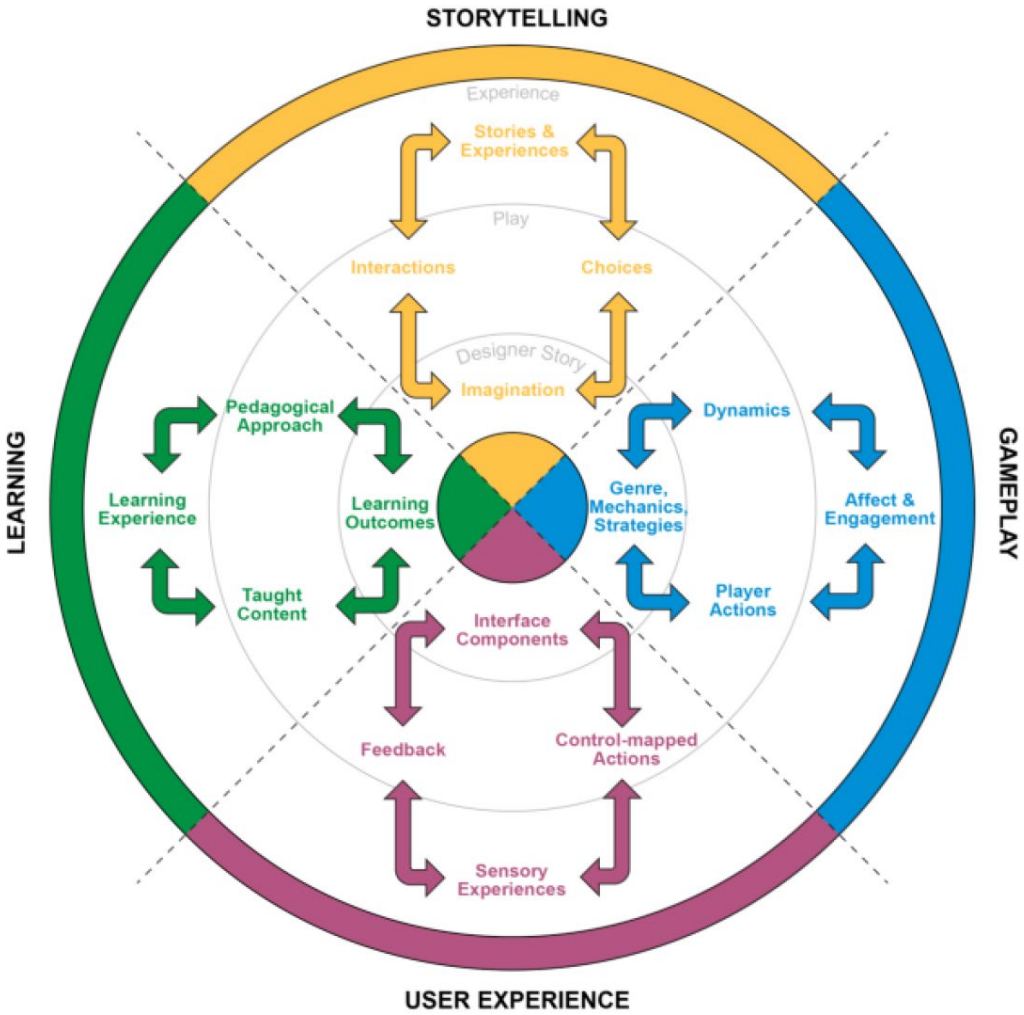
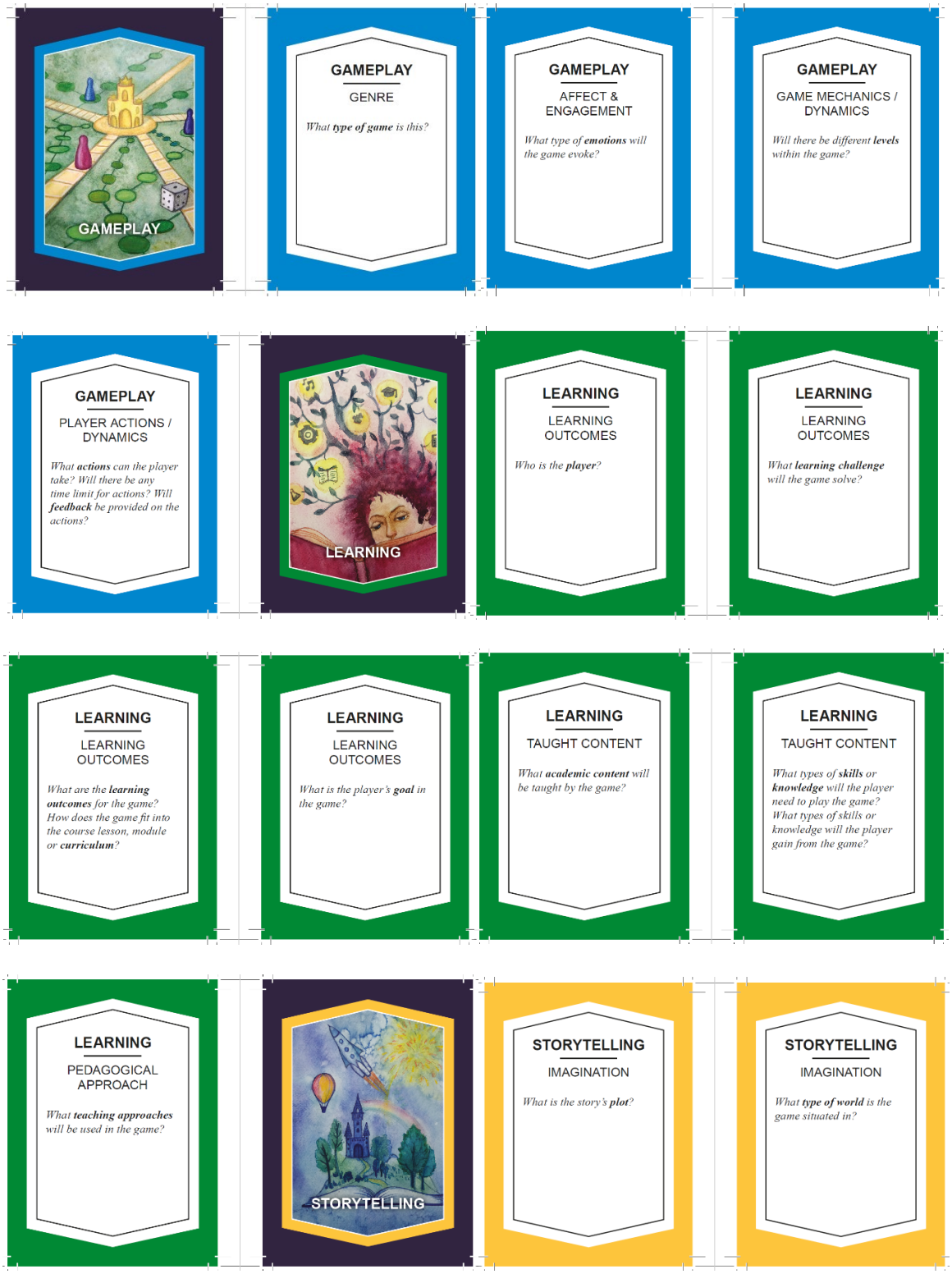


Figure E2: The Art of Serious Game Design methodology circle as adapted from the Design, Play, and Experience (DPE) Framework by Winn, 2009

E.2.2 Brainstorming 1 Cards

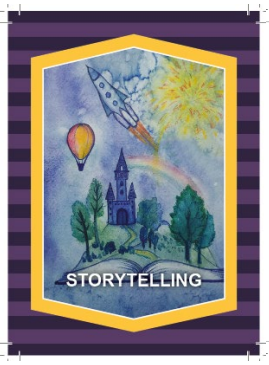


E.2.3 Brainstorming 2 Cards

LEARNING

TAUGHT CONTENT

*What types of **skills** or **knowledge** will the player need to play the game? What types of skills or knowledge will the player gain from the game?*



STORYTELLING

IMAGINATION

*What does the **game's world** look like? What is the world's name and setting? What are the world's rules and limitations? What do your **senses** tell you?*

STORYTELLING

IMAGINATION

*What **backstory** is the player provided with? How will this be communicated?*

STORYTELLING

IMAGINATION

*Who is the main character? Who are the secondary characters? What are their names, missions and **background characteristics**?*

STORYTELLING

IMAGINATION

What type of actions can the main character take? What roles do the secondary characters have?

STORYTELLING

INTERACTIONS

*Who or what does the player **interact** with during the game? What do these characters or objects look like? What do they do?*

STORYTELLING

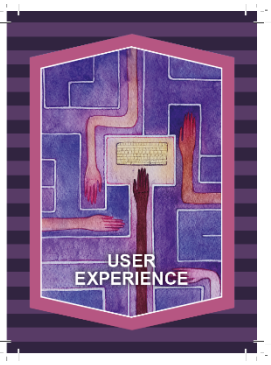
CHOICES

Describe 2-3 scenarios within the game: what problems or challenges are the characters and objects involved with? What are the possible actions or solutions in each scenario? What is the ideal solution?

STORYTELLING

EMERGING STORIES

*How will the player's **stories unfold** in the game?*



USER EXPERIENCE

CONTROL-MAPPED ACTIONS

Can the player pause, stop or save their progress during the game?

USER EXPERIENCE

FEEDBACK

How will the player share their game results with other players?

USER EXPERIENCE

FEEDBACK

How is the player's progress in the game communicated?

USER EXPERIENCE

INTERFACE COMPONENTS

Is this a 2D or 3D game? Is this a photorealistic, cartoon or abstract game?

USER EXPERIENCE

INTERFACE COMPONENTS

How and when will audio be used in the game (e.g. narration, music, etc.)?

USER EXPERIENCE

INTERFACE COMPONENTS

*Will there be **minimaps**, menus or scores in the game? Where will these be located and when will they be presented to the player?*

E.2.4 Group Semi-Structured Interviewing: Debrief and Evaluation Questions

Following the design activities, workshop participants finally partook in a focus group debriefing and group semi-structured interview relevant to evaluation. During this session, participants spoke of their experience with the SGD workshop for police-community intervention. Part 1 (Debriefing) of this session consisted of questions which shed light onto what the participants learned from the workshop for their own practice. Part 2 (Evaluation) of the session consisted of questions which investigated the value of serious game design as an intervention method. Debrief and evaluation questions were drawn and adapted from: Krueger's (2014) focus group guide; Breen's (2006) practical guide to focus group research; and Winn's (2009) game development process evaluation survey. All posed questions are outlined below in Table D2.

Table E2: Group semi-structured interviewing: debrief and evaluation questions for workshops

Part		Question
1	Debrief	1. What was the biggest challenge for your team when designing your serious game concept? How did you overcome this challenge?
1	Debrief	2. What was the biggest learning moment for you and your team while designing your serious game concept? What made this event so impactful?
1	Debrief	3. How did you support one another to achieve success throughout the serious game design process?
1	Debrief	4. How do you feel that your experience today relates to the real-world?
1	Debrief	5. Is there anything else you would like to say about your experience participating in these workshops?
2	Evaluation	6. How do you think the involvement of both community members and law enforcement officers in these workshops impacted the outcome of your serious game concept?
2	Evaluation	7. How do you think the involvement of both community members and law enforcement officers in these workshops impacted overall police-community relationships?
2	Evaluation	8. Do you think that the serious game conceptual design workshop encouraged participants to openly explore aspects of racial bias in policing? Please explain.
2	Evaluation	9. Do you think that serious game design based workshops could be effective in improving police-community relationships and improving overall community satisfaction with police? Why?
2	Evaluation	10. How could these serious game conceptual design workshops be improved for better police-community relationship outcomes?

E.2.5 Glossary

360-degree video: an immersive video where the player is presented with a panoramic view of a location, mimicking the experience of being physically present in that location.

Academic content: subject-specific material that is based on a given curriculum.

Accessible: ensuring the player is provided with sufficient means to interact with the game, e.g., providing visual cues for players with hearing impairments or auditory cues or high contrast for those with visual impairments.

Actions: the things the player does during gameplay.

Background characteristics: characteristics, such as gender, appearance, occupation, age, etc., as well as any strengths, weaknesses, dreams and achievements the characters may have.

Backstory: the background story of the game. This can be communicated using text, video or audio, through dialog with cut scenes, interactions with objects in the world, etc. It can be provided at the beginning or throughout the game.

Curriculum: the planned academic content to be taught in a course of study.

Characters: actors in the game. The main character is the character controlled by the player and the secondary characters are those encountered in the game.

Emotions: the feelings and sensations evoked in the player during gameplay, e.g., a sense of achievement, control, power, weakness, adventure, etc.

Feedback: the information that tells the player how they are doing in the game. It can be communicated either after the player performs an action or at particular intervals in the game (e.g., telling the player how to perform actions or displaying their status levels, giving hints, etc.).

Game controls: any devices used to perform an action or navigate within a digital game, e.g., mouse, keyboard or joystick (simple or with haptic feedback). In mobile devices, game control activities are performed by touch.

Game's world: the real or fictional setting where the game takes place, i.e., location (indoors, outdoors, on earth, in space, etc.). Games typically have multiple settings.

Gameplay data: any information about the player's behavior that is collected during play.

Goal: the ultimate aim or objective of the game.

Interact: verbal or non-verbal communication with characters and objects such as tools, doors, vehicles, etc.

Knowledge: understanding and learning of content (e.g., facts, concepts, information).
Learning challenge: typical difficulties encountered by the learner during the learning process.

Learning outcomes: statements that describe what a learner should know or be able to do within a defined learning context. Please see the handout on Bloom's Taxonomy of Learning for a list of action verbs and activities that can help to formulate learning outcomes.

Levels: particular sections in the game and include predefined tasks that the player should complete in order to progress through the game.

Mini-maps: visual cues that are provided to the player to show their location within the game's world.

Motivates: strategies built into the game to engage the player. For example, the difficulty may be adjusted as the game progresses, the player may be provided with powers or power ups, etc.

Player: the game's participant (or a target audience). For example, in serious games, this could be a group of high school, undergraduate, postgraduate or continuing education students.

Plot: the defined narrative of the game, e.g., the events unfolding in the game that lead to the game's goal. There are usually a main plot and subplots within a game.

Progress reports: data that the game generates to let the player know how they are doing.

Senses: the five senses and the associated emotional reactions that are evoked in the player. For example, when considering the senses in the storytelling, is there any background sound specific to this world? What is the weather and temperature? How does the air smell?

Skills: the player's ability to do something well.

Story: refers to the narrative of events in the game. There could be one or multiple stories in the game. The actions and choices of the player can lead to different stories. The way stories unfold could be communicated through a narrative at the beginning, segments across game levels, or after a specific task or milestone has been achieved.

Teaching approaches: the specific pedagogical strategies included in the game, e.g., problem-based and experiential learning.

Type of game: a predefined category that best matches the game, e.g., a fantasy game, a shooter game, a puzzle, etc.

Type of world: the historical and social contexts of the game's world, e.g., a prehistoric world, an underwater kingdom, a fantasy world, etc.

E.3 Workshop Session Description

Workshop session activities were carried out by all workshop participants in a collaborative manner. While Paper Prototyping activities were carried out as described in D1, procedures for carrying out Brainstorming 1 and Brainstorming 2 activities were altered slightly due to time restrictions in sessions. As such, a facilitator was tasked with the responsibility of drawing each ideation card from the deck and reading its content aloud to the group. Then, all group participants (i.e., not just "subject matter experts") partook in Brainstorming 1 activities. Then, utilizing only the ideation cards which did not overlap with Brainstorming 1 ideation cards, a facilitator was tasked once more with the responsibility of drawing the new Brainstorming 2 ideation cards from the deck and reading the contents aloud to the group. Similar to Brainstorming 1, all group participants (i.e., not just "subject matter experts") then also partook in Brainstorming 2 activities. Workshop activities pertaining to preliminary surveying and a brief introduction were held in a 30-minute Session 1. SGD related workshop activities, additional surveying, and focus group debriefing were conducted over 90 minutes in Session 2. A session breakdown for the workshop can be found below in Table E3.

Table E3: Session breakdown for workshop A

Workshop Session Breakdown	
Session	**SGD Activity**
Session 1	Social desirability survey
	KAP survey #1
	Introduction to community-oriented policing and serious game design
Session 2	Brainstorming 1
	Brainstorming 2
	Paper prototyping
	Focus group discussion (debriefing)
	KAP survey #2
	Group semi-structured interview (evaluation)

In order to conduct workshops, *Zoom* video conferencing software and *Conceptboard*, a collaborative online whiteboard, was used. Workshop participants interacted through voice and video via a *Zoom* breakout room as they simultaneously followed Winn's (2009) workshop protocol on *Conceptboard*. The online *Conceptboard* workshop space included Winn's methodology circle, Brainstorming 1 ideation cards, Brainstorming 2 ideation cards, and online materials for creative interaction, such as virtual sticky notes, pens, pencils, and markers. Screenshots of the workshop *Conceptboard* space are illustrated below in Figures E3-E7.

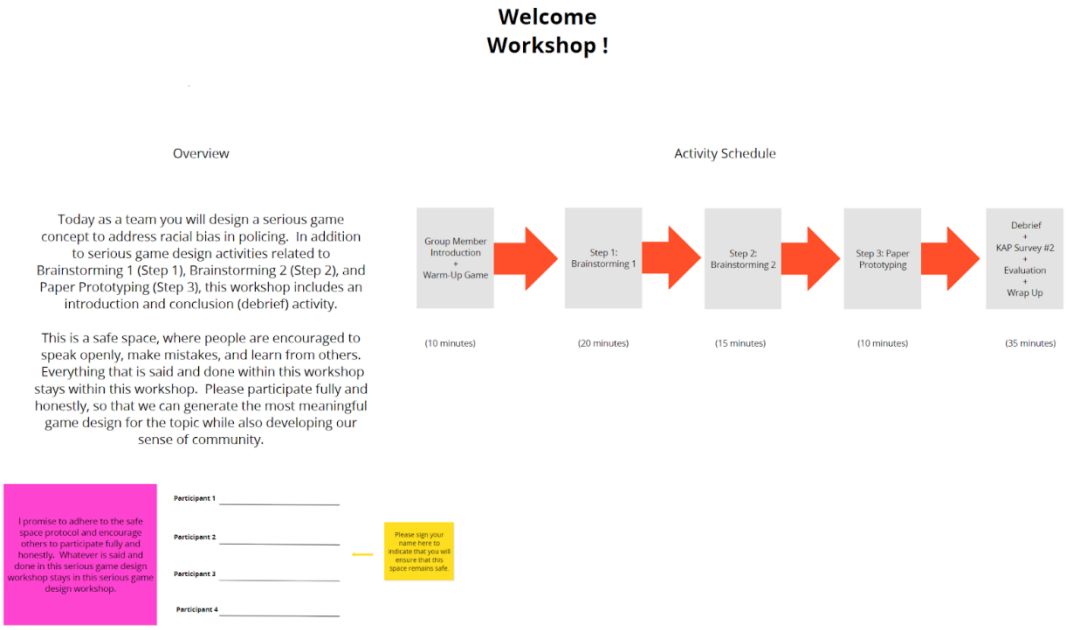


Figure E3: Welcome Section on *Conceptboard* space for workshops

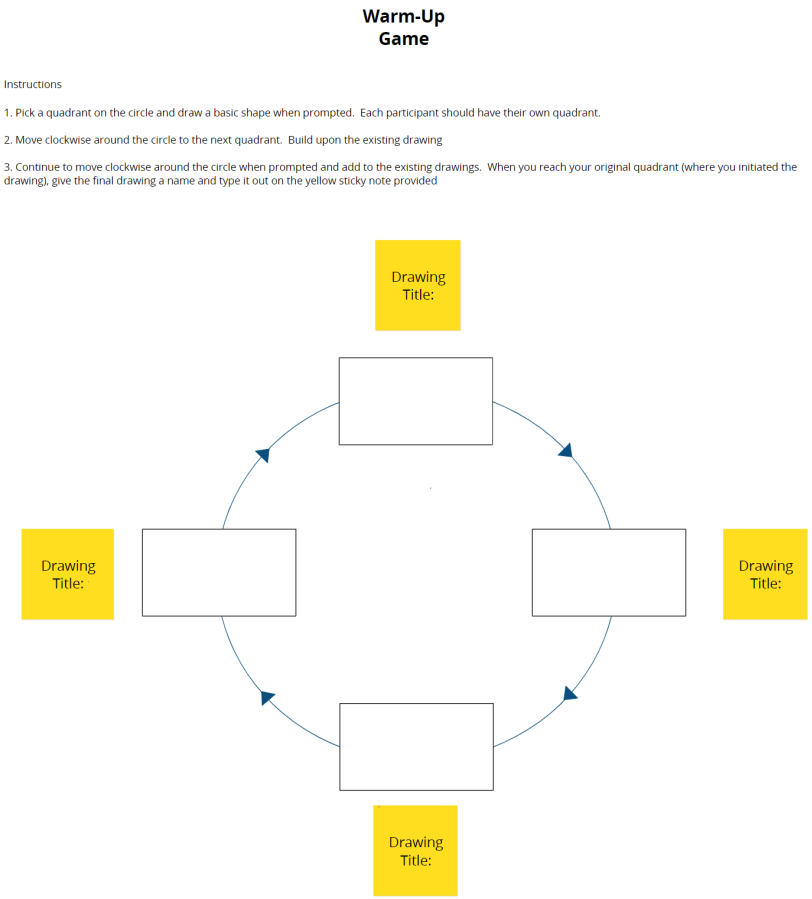


Figure E4: Warm-Up Game Section on *Conceptboard* space for workshops

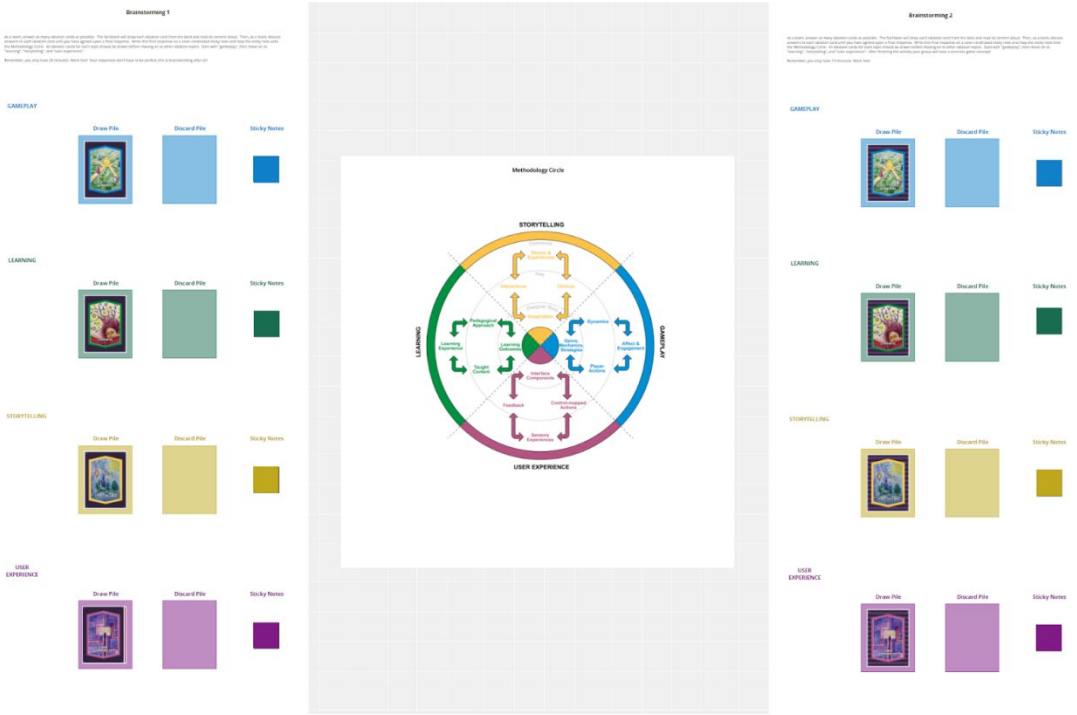


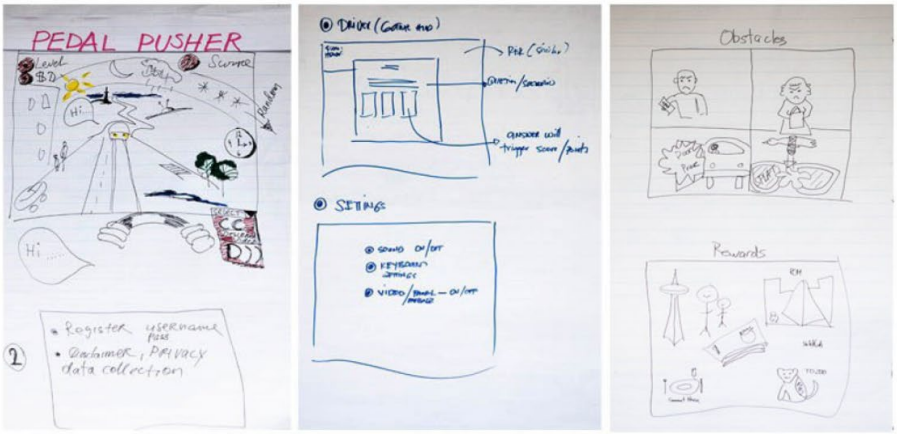
Figure E5: Brainstorming 1, Methodology Circle, and Brainstorming 2 Sections on *Conceptboard* space for workshops

Paper Prototyping

Come together as a group to draw out the game concept as it would appear as a finished product. You can write an accompanying description to the game concept as well!

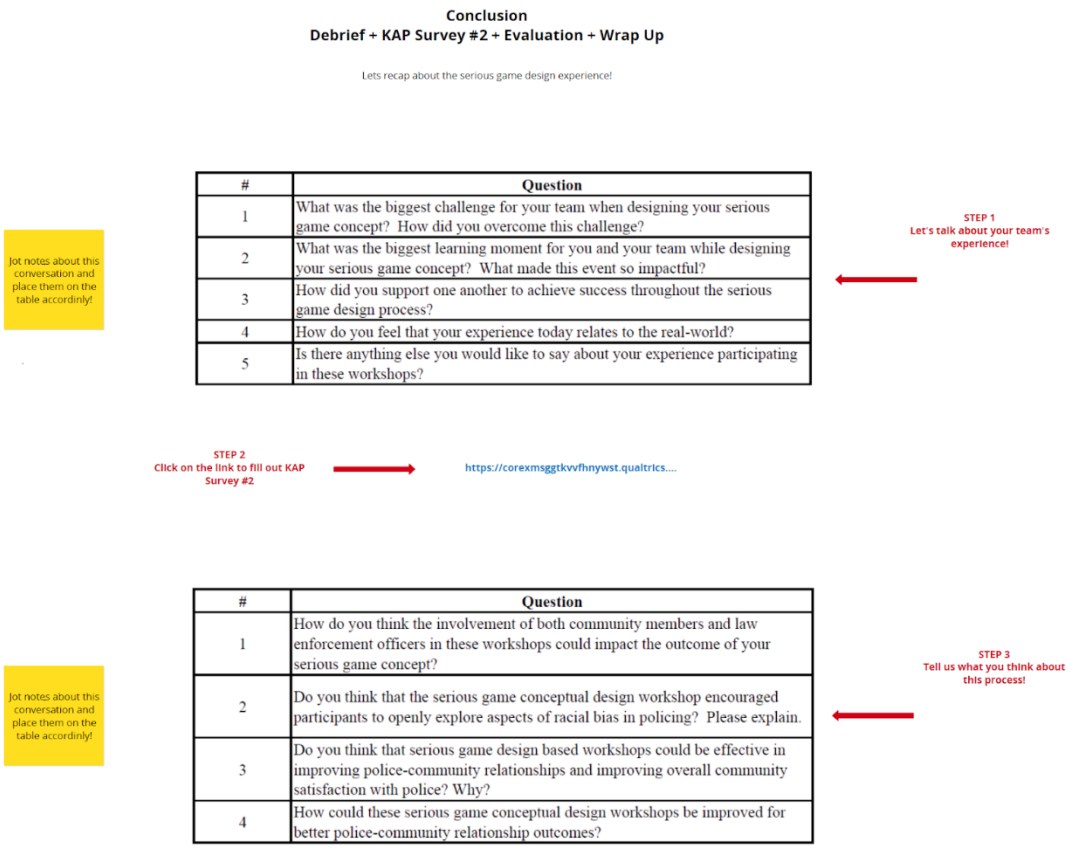
Remember, you only have 10 minutes!

Example



Group Paper Protoype

Figure E6: Paper Prototyping Section on *Conceptboard* space for workshops



#	Question
1	What was the biggest challenge for your team when designing your serious game concept? How did you overcome this challenge?
2	What was the biggest learning moment for you and your team while designing your serious game concept? What made this event so impactful?
3	How did you support one another to achieve success throughout the serious game design process?
4	How do you feel that your experience today relates to the real-world?
5	Is there anything else you would like to say about your experience participating in these workshops?

#	Question
1	How do you think the involvement of both community members and law enforcement officers in these workshops could impact the outcome of your serious game concept?
2	Do you think that the serious game conceptual design workshop encouraged participants to openly explore aspects of racial bias in policing? Please explain.
3	Do you think that serious game design based workshops could be effective in improving police-community relationships and improving overall community satisfaction with police? Why?
4	How could these serious game conceptual design workshops be improved for better police-community relationship outcomes?

Figure E7: Conclusion (Debrief + KAP Survey #2 + Evaluation + Wrap Up) Section on *Conceptboard* space for workshops

APPENDIX F

This appendix outlines the specifics of the within-subjects: paired sampled t-test used within the quantitative component of this research. Related equations and necessary assumptions are put forth. Additionally, the application of this analysis in the context of this research is also described.

As described by Ross & Willson (2017, p. 17) "a paired samples t-test compares the mean of two matched groups of people or cases, or compares the mean of a single group, examined at two different points in time". With this study design, repeated measures can be taken from each participant as they participate in a treatment condition so that the changes within the same participants can be studied over time (Bhandari, 2021).

As indicated by Price, Jhangiani, & Chiang (2015), in conducting a paired samples t-test, first the null hypothesis (H_o) (F.1) and alternative hypothesis (H_1) (F.2) are declared where:

$$H_o: \mu_d = 0 \qquad \text{(F.1)}$$

$$H_1: \mu_d \neq 0 \qquad \text{(F.2)}$$

The standard error (*SE*) (F.3) and average score difference (μ_d) are then used to calculate the test statistic (t) (F.4) where:

$$SE = \frac{s_d}{\sqrt{n}} \qquad \text{(F.3)}$$

$$t = \frac{\mu_d}{\frac{s_d}{\sqrt{n}}} \qquad \text{(F.4)}$$

With the calculated degrees of freedom (*df*) (F.5) and a pre-selected significance level of $\alpha = 0.05$, the test statistic (t) is then compared to a t value from the t-distribution. If the test statistic (t) is greater than the critical value of t from the t-distribution, then the null hypothesis (H_o) can be rejected, and there is statistical evidence that means of the matched measures differ.

$$df = n - 1 \qquad \text{(F.5)}$$

In conducting a paired samples *t*-test, four crucial assumptions must be met to ensure valid results. Firstly, it is assumed that an interval or ratio level is used to measure the dependent variable (Lund & Lund, 2018). Thus, it is assumed that the dependent variable is continuous (Lund & Lund, 2018). Additionally, it is assumed that the distribution of the differences of the dependent variable is approximately normally distributed and that significant outliers are not present in the differences of the dependent variable (Lund & Lund, 2018). Finally, it is also assumed that two categorical "related groups" or "matched pairs" make up the independent variable for analysis (Lund & Lund, 2018).

In this study, the quantitative data pertaining to constructs of "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" was analyzed from a within-subjects perspective with a paired samples t-test in order to gauge the effectiveness of the intervention for each individual participant. With the paired samples t-tests, the means of each construct measured in KAP #1 (pre SGD workshop intervention) and KAP #2 (post SGD workshop intervention) were compared for the same participant on the basis of their unique "Random ID". As such, multiple paired sample t-tests were conducted with the "Random ID" serving as the independent variable and each of the KAP related construct scores pertaining to "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" serving as the dependent variable for subsequent testing. The *Python* programming language was used in accordance with the *SciPy.stats* module for all analysis. In conducting this analysis, it was determined whether or not statistical evidence existed to support the alternative hypothesis that the mean difference between the paired construct observations from KAP #1 and KAP #2 were significantly different from zero.

APPENDIX G

This appendix outlines the specifics of the between-subjects: analysis of covariance (ANCOVA) used within the quantitative component of this research. Related equations and necessary assumptions are put forth. Additionally, the application of this analysis in the context of this research is also described.

As described by Philippas (2014, p.157) the analysis of covariance (ANCOVA) is a technique that "analyzes grouped data having a response (dependent variable) and two or more predictor variables (covariates) where at least one of them in continuous (i.e., quantitative, scaled) and one of them is categorical (i.e., nominal, non-scaled)". With ANCOVA, it is possible to "model the response of a variable as a linear function of predictor(s), with the coefficients of the line varying among different groups" (Philippas, 2014, pp. 157-158). In ANCOVA, the covariates are included as statistical controls, so to explain the dependent variable variation, reduce error variation, and increase statistical power (Philippas, 2014). Ultimately, this analysis can be used with between-subjects classification designs, where participants are sampled from mutually-exclusive populations (e.g., male, female), in order to generate an understanding of how the sampled responses (i.e., dependent variable) could be affected by participants' different characteristics (i.e., covariates) (Schneider, Avivi-Reich, & Mozuraitis, 2015).

As indicated by Philippas (2014), in conducting an ANCOVA analysis, the null hypothesis (H_o) (G.6) and alternative hypothesis (H_1) (G.7) are declared where:

$$H_o: \mu_1 = \mu_2 = \ldots = \mu_k \quad \text{(G.6)}$$

$$H_1: \mu_i \neq \mu_j, i \neq j \; for \; at \; least \; one \; pair \; of \; (i.j) \quad \text{(G.7)}$$

Then, four steps of calculations are required where in step one, an analysis of variance is considered, in steps two and three, an analysis of covariance is considered, and in step four, steps one, two, and three are brought together. In doing so calculations are made for: the sum of squared deviates (*SS*) (total (*SST*) (G.8), within (*SSW*) (G.9), and between groups (*SSB*) (G.10)) values for the dependent variable of interest; the sum of squared deviates (*SS*) (total (*SSTX*) (G.11) and within groups (*SSWX*) (G.12)) values for the covariates whose effects upon the dependent variable is of interest; and the sum of co-deviates (*SC*) (total (*SCT*) (G.13) and within groups (*SCW*) (G.14)) values for the covariance of covariates and dependent variable where:

$$SST = \sum_{j=1,..k} \sum_{i=1,..nj} (y_{ij} - y)^2 \quad \text{(G.8)}$$

$where;$
$the \; overall \; mean, \; y = \sum_{j=1,..k} \sum_{i=1,..nj} (\frac{y_{ij}}{n})$
$n = \sum_{j=1,..k} (n_j)$

$$SSW = \sum_{j=1,..k} \sum_{i=1,..nj} (y_{ij} - y_j)^2 \quad \text{(G.9)}$$

$$SSB = \sum_{j=1,..k} \sum_{i=1,..nj} (y_j - y)^2 \quad \text{(G.10)}$$

$where;$
$y_j \; is \; an \; estimate \; of \; mean \; for \; each \; group \; j$

$$SST_x = \sum_{j=1,..k} \sum_{i=1,..nj} (x_{ij} - x)^2 \quad \text{(G.11)}$$

$$SSW_x = \sum_{j=1,..k} \sum_{i=1,..nj} \left(x_{ij} - x_j\right)^2 \quad \text{(G.12)}$$

$$SCT = \sum_{j=1,..k} \sum_{i=1,..nj} \left(x_{ij} y_{ij} - xy\right)^2 \quad \text{(G.13)}$$

$$SCW = \sum_{j=1,..k} \sum_{i=1,..nj} \left(x_{ij} y_{ij} - x_j y_j\right)^2 \quad \text{(G.14)}$$

Then, mean square values are calculated and the between-groups mean squares (MS'_B) and the adjusted within-groups mean square (MS'_W) are then used to calculate the test statistic for ANCOVA (F) (G.15) where:

$$F = \frac{MS'_B}{MS'_W} \quad \text{(G.15)}$$

Finally, with a pre-selected significance level of $\alpha = 0.05$, the test statistic for ANCOVA (F) is then compared to an f value from the f-distribution. If the statistic for ANCOVA (F) is greater than the critical value of f from the f-distribution, then the null hypothesis (H_o) can be rejected, and there is statistical evidence that a significant relationship exists.

In conducting ANCOVA analysis, nine crucial assumptions must be met to ensure valid results. First, it is assumed that an interval or ratio level is used to measure the dependent variable and covariate variable(s) (Lund & Lund, 2018). Thus, it is assumed that the dependent variable and the covariate variable(s) are continuous (Lund & Lund, 2018). Second, it is assumed that the independent variable consists of at least two categorical and independent groups (Lund & Lund, 2018). Independence of observations is also assumed (Lund & Lund, 2018). Additionally, it is assumed that significant outliers are not present in the data (Lund & Lund, 2018). For each category of the independent variable, the residuals are assumed to be approximately normally distributed (Lund & Lund, 2018). Variances and regression slopes are assumed homogeneous (Lund & Lund, 2018). Homoscedasticity is assumed present (Lund & Lund, 2018). Finally, at each level of the independent variable, the covariate is assumed to be related linearly to the dependent variable (Lund & Lund, 2018).

In this study, the quantitative data pertaining to constructs of "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" was analyzed from a between-subjects perspective with the analysis of covariance (ANCOVA) in order to gauge how participants' different workshop protocols (i.e., Workshop A and Workshop B) and different demographic characteristics relevant to gender, age, and race could have impacted participants' experiences with the SGD workshop intervention. As such, the effectiveness of the intervention for different groups of participants (on the basis of workshop protocol, gender, age, and race. With ANCOVA, the means of each construct measured in KAP #1 (pre SGD workshop intervention) were used for the covariate variable. The means of each construct measured in KAP #2 (post SGD workshop intervention) were used for the dependent response variable. Multiple ANCOVA tests were conducted where workshop protocol, gender, age, and race each served as the independent factor variable. The *Python* programming language was used in accordance with the *SciPy.stats* module for all analysis. In conducting this analysis, it was determined whether or not statistical evidence existed to support the alternative hypothesis that a significant relationship exists between participants' different demographic characteristics and how they experienced the SGD workshop intervention as a result of their KAP #2 mean construct scores (post SGD workshop intervention).

APPENDIX H

This appendix outlines the results of the between-subjects: analysis of covariance (ANCOVA) in tabular form.

H.1 Between-Subjects: Analysis of Covariance (ANCOVA) Results

The quantitative data pertaining to constructs of "Knowledge about Policing in Communities", "Attitude about Explicit Racism", "Attitude about Policing in Communities", and "Practice about Policing in Communities" was analyzed from a between-subjects perspective with the analysis of covariance (ANCOVA) in order to gauge how participants' different workshop protocols (i.e., Workshop A and Workshop B) and different demographic characteristics relevant to gender, age, and race could have impacted participants' experiences with the SGD workshop intervention. As such, the effectiveness of the intervention for different groups of participants (on the basis of workshop protocol, gender, age, and was investigated. With ANCOVA, the means of each construct measured in KAP #1 (pre SGD workshop intervention) were used for the covariate variable. The means of each construct measured in KAP #2 (post SGD workshop intervention) were used for the dependent response variable. Multiple ANCOVA tests were conducted where workshop protocol, gender, age, and race each served as the independent factor variable. Notably, the calculated "social desirability score" for each participant was applied to both KAP #1 data and KAP #2 before subsequent t-testing. An alpha level (α) of 0.05 was used for all tests.

H.1.1 Knowledge about Policing in Communities

H.1.1.1 Gender

A one-way ANCOVA test was conducted to determine a statistically significant difference between gender identities on post SGD workshop intervention "Knowledge about Policing in Communities" controlling for pre SGD workshop intervention "Knowledge about Policing in Communities". As outlined in Table H1, results from this analysis indicate that there is no significant effect of gender identity on post SGD workshop intervention "Knowledge about Policing in Communities" after controlling for pre SGD workshop intervention "Knowledge about Policing in Communities", F (2, 33) = 0.187, p = .830.

In interpreting Table H1 and subsequent tables in this appendix, it should be noted that the p-values associated with the independent factor variables should be of main concern (e.g., "Uncorrected p-value (p-unc)" relevant to gender). Focus should not be given to the "Uncorrected p-value (p-unc)" relevant to the covariate variable. Essentially, in instances where covariate variables indicate statistical significance with $p > .05$, it can be interpreted that the covariate variable significantly adjusts the dependent response variable. In such instances, it can be understood that the covariate variable is a significant predictor of the dependent response variable. For example, in considering Table H1, it can be understood that there is no significant effect of gender identity on post SGD workshop intervention "Knowledge about Policing in Communities", although higher pre SGD workshop intervention scores for "Knowledge about Policing in Communities" were associated with higher post SGD workshop intervention scores for "Knowledge about Policing in Communities".

Furthermore, in considering the "Source" related to Residuals, it is important to note that this row represents the variance that is not explained by the explanatory variables (i.e., the covariate variable and the dependent response variable). As such, while the "Sum of Squares (SS)" and the "Degrees of freedom (DF)" can be calculated, it is standard in output for ANCOVA tests to observe "NaN" values for columns related to "F-value (F)", "Uncorrected p-value (p-unc)", and "Partial eta-squared (np2)".

Table H1: One-way ANCOVA test results where "gender", the mean "Knowledge about Policing in Communities" pre intervention construct score, and the mean "Knowledge about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
gender	0.007339	2	0.186887	0.830409	0.011200
Mean "Knowledge about Policing in Communities" (pre SGD intervention)	0.372852	1	18.988371	0.000121	0.365243
Residual	0.647981	33	NaN	NaN	NaN

H.1.1.2 Age

A one-way ANCOVA test was conducted to determine a statistically significant difference between age groups on post SGD workshop intervention "Knowledge about Policing in Communities" controlling for pre SGD workshop intervention "Knowledge about Policing in Communities". As outlined in Table H2, results from this analysis indicate that there is a significant effect of age group on post SGD workshop intervention "Knowledge about Policing in Communities" after controlling for pre SGD workshop intervention "Knowledge about Policing in Communities", F (3, 32) = 4.30, p = .012.

Table H2: One-way ANCOVA test results where "age", the mean "Knowledge about Policing in Communities" pre intervention construct score, and the mean "Knowledge about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
age	0.188371	3	4.303022	0.011686	0.287449
Mean "Knowledge about Policing in Communities" (pre SGD intervention)	0.421939	1	28.915462	0.000007	0.474682
Residual	0.466949	32	NaN	NaN	NaN

In further investigating the significant effect of age group on post SGD workshop intervention "Knowledge about Policing in Communities", posteriori (i.e., post hoc) analysis was conducted first with a Tamhane T2 test. It can be observed in Table H3 that results from this analysis suggest that there could be different distributions of data between age groups of 25-34 years old and 18-24 years old when considering post SGD intervention "Knowledge about Policing in Communities". As also observed in Table H3, differences in data distributions could not be calculated for the age group 55-64 as there was only 1 participant who fell within this age demographic.

Table H3: Tamhane T2 post hoc test results for investigating which specific age group means for the "Knowledge about Policing in Communities" post intervention construct score differ from others

Age Groups	35-44 years old	25-34 years old	18-24 years old	55-64 years old
35-44 years old	1.000000	0.031257	0.811278	NaN
25-34 years old	0.031257	1.000000	0.301448	NaN
18-24 years old	0.811278	0.301448	1.000000	NaN
55-64 years old	NaN	NaN	NaN	1.0

In further investigating these outcomes, descriptive statistics indicated that there was a slight increase from 0.94 to 0.97 in the measured mean "Knowledge about Policing in Communities" for pre and post SGD workshop intervention levels for all 11 participants aged between 18-24 years old. Further descriptive statistics indicated that there was a slight decrease in the mean of this construct score from pre SGD workshop intervention levels to post SGD workshop intervention levels from 0.94 to 0.88 respectively for all 22 participants aged between 25 and 34 years old. Finally, or participants aged

between 35 and 44 years old, the mean "Knowledge about Policing in Communities" construct score increased slightly from 0.89 to 1.0 for pre SGD workshop intervention scoring and post SGD workshop intervention scoring when considering all 3 participants in this age range.

H.1.1.3 Race

A one-way ANCOVA test was conducted to determine a statistically significant difference between racial identities on post SGD workshop intervention "Knowledge about Policing in Communities" controlling for pre SGD workshop intervention "Knowledge about Policing in Communities". As outlined in Table H4, results from this analysis indicate that there is no significant effect of racial identity on post SGD workshop intervention "Knowledge about Policing in Communities" after controlling for pre SGD workshop intervention "Knowledge about Policing in Communities" F (1, 34) = 2.57, p = .118.

Table H4: One-way ANCOVA test results where "race", the mean "Knowledge about Policing in Communities" pre intervention construct score, and the mean "Knowledge about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
race	0.046038	1	2.569093	0.118221	0.070253
Mean "Knowledge about Policing in Communities" (pre SGD intervention)	0.325090	1	18.141137	0.000153	0.347924
Residual	0.609282	34	NaN	NaN	NaN

H.1.1.4 Workshop Protocol

A one-way ANCOVA test was conducted to determine a statistically significant difference between workshop protocols on post SGD workshop intervention "Knowledge about Policing in Communities" controlling for pre SGD workshop intervention "Knowledge about Policing in Communities". As outlined in Table H5, results from this analysis indicate that there is no significant effect of workshop protocol on post SGD workshop intervention "Knowledge about Policing in Communities" after controlling for pre SGD workshop intervention "Knowledge about Policing in Communities" F (1, 34) = 1.79, p = .190.

Table H5: One-way ANCOVA test results where "protocol", the mean "Knowledge about Policing in Communities" pre intervention construct score, and the mean "Knowledge about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
protocol	0.032754	1	1.788794	0.189952	0.049982
Mean "Knowledge about Policing in Communities" (pre SGD intervention)	0.270244	1	14.758728	0.000508	0.302689
Residual	0.622567	34	NaN	NaN	NaN

H.1.2 Attitude about Explicit Racism

H.1.2.1 Gender

A one-way ANCOVA test was conducted to determine a statistically significant difference between gender identities on post SGD workshop intervention "Attitude about Explicit Racism" controlling for pre SGD workshop intervention "Attitude about Explicit Racism". As outlined in Table H6, results from this analysis indicate that there is no significant effect of gender identity on post SGD workshop

intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Explicit Racism", F (2, 33) = 0.443, p = .646.

Table H6: One-way ANCOVA test results where "gender", the mean "Attitude about Explicit Racism" pre intervention construct score, and the mean "Attitude about Explicit Racism" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
gender	0.052273	2	0.443070	6.458260e-01	0.026150
Mean "Attitude about Explicit Racism" (pre SGD intervention)	25.725883	1	436.106268	1.360828e-20	0.929653
Residual	1.946668	33	NaN	NaN	NaN

H.1.2.2 Age

A one-way ANCOVA test was conducted to determine a statistically significant difference between age groups on post SGD workshop intervention "Attitude about Explicit Racism" controlling for pre SGD workshop intervention "Attitude about Explicit Racism". As outlined in Table H7, results from this analysis indicate that there is no significant effect of age group on post SGD workshop intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Explicit Racism", F (3, 32) = 0.190, p = .903.

Table H7: One-way ANCOVA test results where "age", the mean "Attitude about Explicit Racism" pre intervention construct score, and the mean "Attitude about Explicit Racism" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
age	0.034916	3	0.189627	9.026824e-01	0.017467
Mean "Attitude about Explicit Racism" (pre SGD intervention)	19.926198	1	324.658841	2.579210e-18	0.910278
Residual	1.964026	32	NaN	NaN	NaN

H.1.2.3 Race

A one-way ANCOVA test was conducted to determine a statistically significant difference between racial identities on post SGD workshop intervention "Attitude about Explicit Racism" controlling for pre SGD workshop intervention "Attitude about Explicit Racism". As outlined in Table H8, results from this analysis indicate that there is no significant effect of racial identity on post SGD workshop intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Explicit Racism", F (1, 34) = 2.11, p = .155.

Table H8: One-way ANCOVA test results where "race", the mean "Attitude about Explicit Racism" pre intervention construct score, and the mean "Attitude about Explicit Racism" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
race	0.117006	1	2.113882	1.551380e-01	0.058534
Mean "Attitude about Explicit Racism" (pre SGD intervention)	27.126211	1	490.075817	8.957390e-22	0.935124
Residual	1.881936	34	NaN	NaN	NaN

H.1.2.4 Workshop Protocol

A one-way ANCOVA test was conducted to determine a statistically significant difference between workshop protocols on post SGD workshop intervention "Attitude about Explicit Racism" controlling for pre SGD workshop intervention "Attitude about Explicit Racism". As outlined in Table H9, results from this analysis indicate that there is no significant effect of workshop protocol on post SGD workshop intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Explicit Racism" $F(1, 34) = 0.000028$, $p = .996$.

Table H9: One-way ANCOVA test results where "protocol", the mean "Attitude about Explicit Racism" pre intervention construct score, and the mean "Attitude about Explicit Racism" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
protocol	0.000002	1	0.000028	9.958421e-01	8.105470e-07
Mean "Attitude about Explicit Racism" (pre SGD intervention)	25.810062	1	439.003783	5.137116e-21	9.281190e-01
Residual	1.998940	34	NaN	NaN	NaN

H.1.3 Attitude about Policing in Communities

H.1.3.1 Gender

A one-way ANCOVA test was conducted to determine a statistically significant difference between gender identities on post SGD workshop intervention "Attitude about Policing in Communities" controlling for pre SGD workshop intervention "Attitude about Policing in Communities". As outlined in Table H10, results from this analysis indicate that there is no significant effect of gender identity on post SGD workshop intervention "Attitude about Policing in Communities" after controlling for pre SGD workshop intervention "Attitude about Policing in Communities", $F(2, 33) = 0.00756$, $p = .925$.

Table H10: One-way ANCOVA test results where "gender", the mean "Attitude about Policing in Communities" pre intervention construct score, and the mean "Attitude about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
gender	0.000537	2	0.007558	9.924721e-01	0.000458
Mean "Attitude about Policing in Communities" (pre SGD intervention)	5.001578	1	140.834235	1.885780e-13	0.810164
Residual	1.171960	33	NaN	NaN	NaN

H.1.3.2 Age

A one-way ANCOVA test was conducted to determine a statistically significant difference between age groups on post SGD workshop intervention "Attitude about Policing in Communities" controlling for pre SGD workshop intervention "Attitude about Policing in Communities". As outlined in Table H11, results from this analysis indicate that there is no significant effect of age group on post SGD workshop intervention "Attitude about Policing in Communities" after controlling for pre SGD workshop intervention "Attitude about Policing in Communities", $F(3, 32) = 1.071$, $p = .375$.

Table H11: One-way ANCOVA test results where "age", the mean "Attitude about Policing in Communities" pre intervention construct score, and the mean "Attitude about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
age	0.308281	3	1.071399	3.750201e-01	0.091276
Mean "Attitude about Policing in Communities" (pre SGD intervention)	9.064294	1	94.506156	4.504876e-11	0.747048
Residual	3.069191	32	NaN	NaN	NaN

H.1.3.3 Race

A one-way ANCOVA test was conducted to determine a statistically significant difference between racial identities on post SGD workshop intervention "Attitude about Policing in Communities" controlling for pre SGD workshop intervention "Attitude about Policing in Communities". As outlined in Table H12, results from this analysis indicate that there is no significant effect of racial identity on post SGD workshop intervention "Attitude about Explicit Racism" after controlling for pre SGD workshop intervention "Attitude about Policing in Communities", $F(1, 34) = 0.0106$, $p = .919$.

Table H12: One-way ANCOVA test results where "race", the mean "Attitude about Policing in Communities" pre intervention construct score, and the mean "Attitude about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
race	0.001050	1	0.010569	9.187205e-01	0.000311
Mean "Attitude about Policing in Communities" (pre SGD intervention)	11.016641	1	110.935717	3.038036e-12	0.765413
Residual	3.376422	34	NaN	NaN	NaN

H.1.3.4 Workshop Protocol

A one-way ANCOVA test was conducted to determine a statistically significant difference between workshop protocols on post SGD workshop intervention "Attitude about Policing in Communities" controlling for pre SGD workshop intervention "Attitude about Policing in Communities". As outlined in Table H13, results from this analysis indicate that there is no significant effect of workshop protocol on post SGD workshop intervention "Attitude about Policing in Communities" after controlling for pre SGD workshop intervention "Attitude about Policing in Communities" $F(1, 34) = 0.533$, $p = .470$.

Table H13: One-way ANCOVA test results where "protocol", the mean "Attitude about Policing in Communities" pre intervention construct score, and the mean "Attitude about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
protocol	0.052165	1	0.533369	4.701966e-01	0.015445
Mean "Attitude about Policing in Communities" (pre SGD intervention)	10.851922	1	110.956804	3.030469e-12	0.765447
Residual	3.325306	34	NaN	NaN	NaN

H.1.4 Practice about Policing in Communities

H.1.4.1 Gender

A one-way ANCOVA test was conducted to determine a statistically significant difference between gender identities on post SGD workshop intervention "Practice about Policing in Communities"

controlling for pre SGD workshop intervention "Practice about Policing in Communities". As outlined in Table H14, results from this analysis indicate that there is no significant effect of gender identity on post SGD workshop intervention "Practice about Policing in Communities" after controlling for pre SGD workshop intervention "Practice about Policing in Communities $F(2, 33) = 0.195$, $p = .824$.

Table H14: One-way ANCOVA test results where "gender", the mean "Practice about Policing in Communities" pre intervention construct score, and the mean "Practice about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
gender	0.030771	2	0.194809	8.239311e-01	0.011669
Mean "Practice about Policing in Communities" (pre SGD intervention)	20.126644	1	254.840548	4.402654e-17	0.885353
Residual	2.606254	33	NaN	NaN	NaN

H.1.4.2 Age

A one-way ANCOVA test was conducted to determine a statistically significant difference between age groups on post SGD workshop intervention "Practice about Policing in Communities" controlling for pre SGD workshop intervention "Practice about Policing in Communities". As outlined in Table H15, results from this analysis indicate that there is no significant effect of age group on post SGD workshop intervention "Practice about Policing in Communities" after controlling for pre SGD workshop intervention "Practice about Policing in Communities", $F(3, 32) = 0.299$, $p = .826$.

Table H15: One-way ANCOVA test results where "age", the mean "Practice about Policing in Communities" pre intervention construct score, and the mean "Practice about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
age	0.071819	3	0.298639	8.261030e-01	0.027235
Mean "Practice about Policing in Communities" (pre SGD intervention)	17.474799	1	217.991673	7.752646e-16	0.871996
Residual	2.565206	32	NaN	NaN	NaN

H.1.4.3 Race

A one-way ANCOVA test was conducted to determine a statistically significant difference between racial identities on post SGD workshop intervention "Practice about Policing in Communities"" controlling for pre SGD workshop intervention "Practice about Policing in Communities"". As outlined in Table H16, results from this analysis indicate that there is no significant effect of racial identity on post SGD workshop intervention "Practice about Policing in Communities"" after controlling for pre SGD workshop intervention "Practice about Policing in Communities"", $F(1, 34) = 0.0449$, $p = .833$.

Table H16: One-way ANCOVA test results where "race", the mean "Practice about Policing in Communities" pre intervention construct score, and the mean "Practice about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
race	0.003479	1	0.044914	8.334280e-01	0.001319
Mean "Practice about Policing in Communities" (pre SGD intervention)	21.776685	1	281.144583	5.209840e-18	0.892113
Residual	2.633546	34	NaN	NaN	NaN

H.1.4.4 Workshop Protocol

A one-way ANCOVA test was conducted to determine a statistically significant difference between workshop protocols on post SGD workshop intervention "Practice about Policing in Communities" controlling for pre SGD workshop intervention "Practice about Policing in Communities". As outlined in Table H17, results from this analysis indicate that there is no significant effect of workshop protocol on post SGD workshop intervention "Practice about Policing in Communities" after controlling for pre SGD workshop intervention "Practice about Policing in Communities" $F(1, 34) = 0.385$, $p = .53$.

Table H17: One-way ANCOVA test results where "protocol", the mean "Practice about Policing in Communities" pre intervention construct score, and the mean "Practice about Policing in Communities" post intervention construct score serves as the factor variable, covariate, and response variable respectively

Source	Sum of Squares (SS)	Degrees of freedom (DF)	F-value (F)	Uncorrected p-value (p-unc)	Partial eta-squared (np2)
protocol	0.029541	1	0.385199	5.389721e-01	0.011202
Mean "Practice about Policing in Communities" (pre SGD intervention)	20.245337	1	263.986838	1.353834e-17	0.885901
Residual	2.607484	34	NaN	NaN	NaN