Risk Assessment and Management of Group-Based Violence

by

Alana N. Cook

M.S. (Clinical Psychology), Pacific University, 2010
B.A. (Psychology), University of British Columbia, 2008

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Psychology
Faculty of Arts and Social Sciences

© Alana N. Cook 2014

SIMON FRASER UNIVERSITY
Summer 2014

All rights reserved. However, in accordance with the Copyright Act of Canada, this work may be reproduced, without authorization, under the conditions for “Fair Dealing.” Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
Approval

Name: Alana N. Cook
Degree: Doctor of Philosophy (Psychology)
Title: Risk Assessment and Management of Group-Based Violence

Examining Committee:

**Chair:** Robert McMahon
Professor

---

Stephen D. Hart
Senior Supervisor
Professor

Raymond R. Corrado
Supervisor
Professor
School of Criminology

P. Randall Kropp
Supervisor
Adjunct Professor

Ronald Roesch
Supervisor
Professor

Garth Davies
Internal Examiner
Associate Professor
School of Criminology

Mario Scalora
External Examiner
Associate Professor
Department of Psychology
University of Nebraska-Lincoln

---

Date Defended/Approved: April 16, 2014
Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the non-exclusive, royalty-free right to include a digital copy of this thesis, project or extended essay[s] and associated supplemental files (“Work”) (title[s] below) in Summit, the Institutional Research Repository at SFU. SFU may also make copies of the Work for purposes of a scholarly or research nature; for users of the SFU Library; or in response to a request from another library, or educational institution, on SFU’s own behalf or for one of its users. Distribution may be in any form.

The author has further agreed that SFU may keep more than one copy of the Work for purposes of back-up and security; and that SFU may, without changing the content, translate, if technically possible, the Work to any medium or format for the purpose of preserving the Work and facilitating the exercise of SFU’s rights under this licence.

It is understood that copying, publication, or public performance of the Work for commercial purposes shall not be allowed without the author’s written permission.

While granting the above uses to SFU, the author retains copyright ownership and moral rights in the Work, and may deal with the copyright in the Work in any way consistent with the terms of this licence, including the right to change the Work for subsequent purposes, including editing and publishing the Work in whole or in part, and licensing the content to other parties as the author may desire.

The author represents and warrants that he/she has the right to grant the rights contained in this licence and that the Work does not, to the best of the author’s knowledge, infringe upon anyone’s copyright. The author has obtained written copyright permission, where required, for the use of any third-party copyrighted material contained in the Work. The author represents and warrants that the Work is his/her own original work and that he/she has not previously assigned or relinquished the rights conferred in this licence.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2013
Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010
Abstract

Group-based violence (GBV) may be defined as actual, attempted, or threatened physical injury that is deliberate and nonconsensual, perpetrated by one or more individuals whose decisions and behaviour are influenced by a group to which they currently belong or with which they are affiliated. Although GBV represents a serious challenge to professionals around the world tasked with protecting public safety, there is lack of systematic, evidence-based procedures to aid decision-making. This dissertation reports the development and evaluation of a new set of structured professional judgment (SPJ) guidelines for assessing and managing GBV, called the Multi-level Guidelines (MLG; Cook, Hart, & Kropp, 2013). The first part of the dissertation describes the development of the MLG based on a Campbell Collaboration review and expert feedback. The MLG was structured according to an ecological model of GBV comprising 20 risk factors in four nested domains: Individual, Individual-Group, Group, and Group-Societal. The second part of the dissertation reports on an evaluation of the MLG in two samples of criminal justice and mental health professionals who completed training and rated case studies. Consistent with predictions, the results of the evaluation indicated that professionals who completed the training: (1) reported significant increases in their confidence, competence, and knowledge concerning the assessment and management of GBV significantly; (2) appraised the MLG to be useful for their practice; and (3) made judgments concerning the presence of risk factors, as well as the nature and level of risks posed, with a degree of reliability comparable to that reported in evaluations of other SPJ guidelines. The professionals also provided feedback for improving the MLG. Overall, the findings suggest the MLG may aid decisions about GBV made by professionals working with diverse problems in a wide range of settings.

Keywords: Risk Assessment; Risk Management; Group-Based Violence; Multi-level Guidelines; Terrorism; Organized Crime
Dedication

To my mom. Without her love, strength, and bravery my completion of a Doctor of Philosophy would not have been possible.
Acknowledgements

On the first day of my PhD, Stephen Hart asked me what I wanted to do while at SFU. I told Stephen an idea I had, an *if I could do anything* idea for a project. He simply responded: “that is the one I want to hear!” Four years later with the teaching, mentorship, and leadership of Stephen that project is complete and continues to develop and spur thinking, theory development, and practice globally.

I feel incredibly grateful and privileged to have had the opportunity to work with Stephen these past four years. He not only has significantly influenced my thinking, but he is hands down the best teacher I have ever had – and that is some steep competition! Time and time again he has encouraged me to challenge myself intellectually, be innovative in our field, and to have fun doing it! I only hope his preposterous definition of graduate (or gradual?) school is correct, and that I continue to have the privilege of working with him for years to come.

Stephen is one of many phenomenal colleagues at (or affiliated with) SFU I have been fortunate to work with. Together they created an environment for me that fostered learning, growth, and support in all aspects of my graduate training. I would like to thank my doctoral thesis committee: Ray Corrado for telling me at my proposal defence that this project was impossible, which only further inspired me to successfully complete it. Randy Kropp for your thoughtful guidance on this project and teaching me exceptional forensic assessment skills that are the foundation of my clinical practice. Ron Roesch for your vision for the field and for taking me along for the ride. I would also like to acknowledge Lisa Brown, Deb Connolly, Kevin Douglas (no, I am not going to law school now!), Jeanne LeBlanc, Bob McMahon, Hendre Viljoen, Jodi Viljoen and all of the wonderful friends I made through SFU, especially Michelle Collins, Laura Guy, Bree Layden, Yan Lim, Jen Storey, Diane Strub, Simone Viljoen, Kelly Watt, Cathy Wilson and all the members of Stephen’s lab.

I would also like to thank my students, colleagues, supervisors, teachers, and mentors throughout my graduate training, including Gwyn Amat, Kate Bousfleid, Steve Burton, Gary Chaimowitz, Lisa Christensen, Sophia Grewal, Michelle Guyton (you started this!),
Lauren Kois, Mini Mamak, Heather Moulden, and all the law enforcement members I have been privileged work with across Canada.

This project was supported by numerous funding sources: the Social Sciences and Humanities Research Council of Canada (SSHRC), Public Safety Canada Kanishka Project Contribution, ProActive ReSolutions, and a M.D. Angus and Associates Graduate Fellowship in Psychology. I also received considerable support from the SFU library services who help me access numerous sources required for my review of the literature. I would like to thank the experts who reviewed this project, and especially Reid Meloy, John Monahan, and Mario Scalora who continue to provide valuable feedback.

Lastly, I would like to thank my friends and family who saw me through the challenges and triumphs of my graduate training. I would like to extended special thanks to my parents, for supporting me to every way possible, to my brother, for always keeping me on my toes, to my cousins, aunts and uncles, friends, and family friends. It took a village to get me though, but we did it!

The final person I would like to thank has been the unsung pillar of my success in graduate school, Jason. Jason has whole heartedly supported me on this journey, regardless of what city, province, or country I have been in or what monumental task I faced. He is far too modest about the appreciation he deserves for all that he has done to contribute to my success over the past six years. With all my love, thank-you.
# Table of Contents

1. **Introduction** .................................................................................................................. 1
   1.1. Group-Based Violence ................................................................................................. 1
   1.1.1. Nature of the Problem ............................................................................................ 2
   1.2. Risk Assessment .......................................................................................................... 3
       1.2.1. Individual Risk Assessment of Group-Based Violence ...................................... 4
       1.2.2. The Application of Structure Professional Judgment ......................................... 6

2. **Results, Part 1: Development of the Multi-Level Guidelines** ................................. 12
   2.1. Campbell Collaboration Methodology .......................................................................... 12
       2.1.1. Data screening ..................................................................................................... 14
       2.1.2. Data management ............................................................................................... 15
   2.2. Results .......................................................................................................................... 15
       2.2.1. Data coding ........................................................................................................ 15
       2.2.2. Scope of the existing evidence ........................................................................... 16
   2.3. Discussion ..................................................................................................................... 17
       2.3.1. Group-Based Violence Risk Factors .................................................................... 17
       2.3.2. The Multi-Level Guidelines (MLG) .................................................................... 19

3. **Results, Part 2: Evaluation of the Multi-Level Guidelines** .................................... 23
   3.1. Evaluation of Risk Assessment Guidelines .................................................................... 23
   3.2. Method ......................................................................................................................... 28
       3.2.1. Participants .......................................................................................................... 28
       3.2.2. Description of the MLG Training ...................................................................... 31
           Materials ..................................................................................................................... 32
           Utility Questionnaires ............................................................................................ 32
           MLG Worksheets .................................................................................................... 33
           Cases ....................................................................................................................... 33
       3.2.3. Procedure ............................................................................................................ 35
       3.2.4. Statistical Analysis ............................................................................................. 37
   3.3. Results .......................................................................................................................... 39
       3.3.1. Utility ................................................................................................................. 39
List of Tables

Table 2.1. Inclusion criteria for the Campbell Collaboration Review.......................... 13
Table 2.2. Countries represented in Campbell Collaboration review.......................... 17
Table 2.3. Group-Based Violence Risk Factors by Domain........................................ 21
Table 3.1. Demographics of Course Participants, N = 46........................................... 29
Table 3.2. Experience of course participants with widely used threat assessment instruments, N = 46.............................................................. 31
Table 3.3. Pre- and post- training self-ratings questions for general and GBV............. 33
Table 3.4. Characteristics of the eleven standardized and sanitized cases .................. 35
Table 3.5. Mean pre- and post- training self-ratings for general and GBV, N = 46........ 40
Table 3.6. Participants indicating intended use of the MLG with various forms of GBV, N = 46................................................................. 41
Table 3.7. Presence of the Multi-level Guidelines (MLG) factors: Distribution and Interrater Reliability of Ratings ................................................... 45
Table 3.8. Presence of the Multi-level Guidelines (MLG) factors: Distribution, Structural Reliability, and Interrater reliability of Numerical scores ........ 46
Table 3.9. Multi-level Guidelines (MLG) Conclusory Opinions: Distribution and Interrater Reliability of Ratings .................................................. 46

List of Figures

Figure 2.1. Nested model of GBV risk factors............................................................ 18
Figure 3.1. Content Analysis Results........................................................................... 43
List of Acronyms

GBV     Group-Based Violence
MLG     Multi-level Guidelines

Glossary

Group-based violence  the actual, attempted, or threatened physical injury of other people that is deliberate and nonconsensual, perpetrated by individuals whose decisions and behaviour are influenced by a group to which they currently belong or with which they are affiliated (Cook, Hart, & Kropp, 2013).
Chapter 1.

Introduction

In September 2010 the first of the ‘Toronto 18’, Ali Mohamed Dirie, became eligible for parole following conviction for terrorism-related offenses. The Toronto 18 is the name given to a group of individuals who were arrested in June 2006 for plotting a violent overthrow of the Government of Canada. Typically, a risk assessment would be employed to aid decision-making about Mr. Dirie’s parole. However, the risk factors typically considered in decisions about parole of correctional offenders (e.g., history of violence, psychopathology; R. v. Gaya, 2008) were not present in many of the Toronto 18 cases. For this reason, researchers have cautioned that traditional risk assessment tools have limited applicability when used with terrorists, whose actions are often influenced by or take place in the context of groups (Monahan, 2012; Roberts & Horgan, 2008). Group dynamics have unique and powerful influences on an individual’s behaviour (Pynchon & Borum, 1999). This is illustrated clearly in the words of an alleged leader of the Toronto 18: “Our mission’s greater, whether we get arrested, killed, or tortured, our mission is greater than just individuals” (R. v. Gaya, 2008, p. 4).

1.1. Group-Based Violence

Violence is defined as actual, attempted, or threatened physical injury of an individual that is deliberate and nonconsensual (Douglas, Hart, Webster, & Belfrage, 2013; see also World Health Organization [WHO], 2002). One subtype of violence is group-based violence, which includes such things as violence by terrorist and organized criminal groups (WHO, 2002).

For the purpose of this dissertation, I will define group-based violence (GBV) as the actual, attempted, or threatened physical injury of other people that is deliberate and
nonconsensual, perpetrated by individuals whose decisions and behaviour are influenced by a group to which they currently belong or with which they are affiliated (Cook, Hart, & Kropp, 2013). In this definition, a group refers to an identifiable collective of two or more individuals who have some stable pattern of associations based to some extent on shared attitudes, norms, values, or goals. This definition captures all major types of collective violence identified by the WHO (2002), as well as many other forms of violence discussed in the literature, such as honour-based violence.

1.1.1. Nature of the Problem

There is an increasing need to understand GBV. GBV represents a daily challenge for professionals around the world tasked with protecting public safety, and its frequency may be increasing (Dahlberg & Krug, 2002; Pynchon & Borum, 1999; WHO, 2002). This is particularly true for two common forms of GBV, violence committed by criminal and terrorist organizations. Transnational organized crime was identified by the United Nations Office on Drugs and Crime (UNODC, 2010a) as a key target due to the major threat to security and development and challenges to democracy internationally, but particular in West Africa. Although no comprehensive international data is recorded on organized crime, the UNODC (2010b) indicates that organized crime is prominent internationally and that increased homicide rates in the Americas (Guatemala, Venezuela, Jamaica, Belize, Trinidad and Tobago, and Honduras) may be linked to organized crime, drug trafficking and gang activity. Trend data are available for some counties. For example, the number of street gangs identified by law enforcement in Canada has been steadily increasing since 2006 (Criminal Intelligence Service Canada, 2010). Similar trends are indicated in the United States of America (US). According to the Federal Bureau of Investigation (FBI; 2011), membership in criminal organizations (gangs) in the US increased 40% between 2009 and 2011, to a total of about 1.4 million active members of about 33,000 active street, organized motorcycle, and prison gangs.

There are similar concerns with terrorist violence. According to the Centre of Excellence – Defense Against Terrorism (COE-DAT), an accredited organization by the North Atlantic Treaty Organization (NATO), although the number of Terrorist attacks worldwide in 2012 at their lowest level in 6 years (down 12% from 2009 and down 40% from 2011
(possibly as a result of intense counter-terrorist activity), they still resulted in 7,294 incidents of terrorist violence resulting in 11,450 deaths and an additional 21,218 injuries, and 813 abductions globally (COE-DAT, 2013). COE-DAT reports that the majority of terrorist incidents (59%) took place in Iraq \(n = 1900\), Pakistan \(n = 1428\), and Afghanistan \(n = 972\).

1.2. Risk Assessment

According to Hart (2008), risk assessment actually involves two related processes, *risk assessment* and *risk management*. Risk assessment is the process of identifying both risk and protective factors that influence violence. Risk management is the prevention of or a reduction in violence by influencing risk and protective factors (Hart, 2008).

Over the past 15 years, an increasing number of risk assessment tools have been developed (Dahlberg & Krug, 2002; Hart, 2008; Singh, Serper, Reinhart, & Fazel, 2010). Singh and colleagues (2010) identified more than 120 such tools. They may be divided into two broad categories: discretionary and non-discretionary approaches. (Hart, 2008). Discretionary approaches, which rely heavily on the judgment of evaluators, include unstructured clinical judgment (UCJ) and structured professional judgment (SPJ). UCJ relies on intuition and the evaluator’s own experience and discretion of what they deem to be important risk factors. This method has been criticized for lacking an empirical basis for reliable and valid judgments (Hart, 2008; Pedersen, Rasmussen, & Elsass, 2010). SPJ, in contrast, is based on the current state of knowledge in the field through empirical evidence and professional practice. The SPJ method provides evaluators with a set of guidelines that outlines systematic steps for evaluating the presence and relevance of risk factors to inform case formulation and scenario planning, management strategies, and conclusory opinions about a case. This type of assessment aims for consistency and transparency. Pedersen and colleagues (2010) reported that the SPJ method has been deemed reliable and valid and that the use of SPJ tools is considered valuable and good practice. The SPJ method requires expertise and training (Hart, 2008).
The non-discretionary approach to violence risk assessment includes psychological testing and actuarial risk assessment instruments (Hart, 2008). Psychological testing uses an individual’s scores on psychological tests (e.g., the Hare Psychopathy Checklist-Revised or PCL-R; Hare, 2003) to determine an individual's level of risk. Actuarial risk assessment instruments are designed to predict outcomes for a specific population within a given time frame (e.g., the probability that a given person will engage in a new sexual offense in the coming 5 years). The items in actuarial risk assessment instruments are derived rationally or empirically, and then weighted and combined according to a fixed algorithm to make a prediction. The primary benefits of these methods are that they are consistent and transparent; they may also be empirically derived. Their limitations include a reliance on strong assumption of validity generalization (i.e., what worked in the construction or validation samples will work in new samples), an inability to incorporate important dispositional or contextual factors in individual cases, uncertain ability to make individual-level predictions based on group data, and a lack of attention to risk management (Hart, 2008; Hart & Cooke, 2013; Pederson et al., 2010).

1.2.1. Individual Risk Assessment of Group-Based Violence

The individual risk assessment of GBV brings up a unique challenge in risk assessment – that is, we are assessing an individual, within the context of a group. Individual risk assessment tools exist for various forms of violence, including general violence (e.g., HCR-20\textsuperscript{V3}; Douglas et al., 2013) – these include factors applicable to individuals, such as active symptoms of major mental illness and lack of insight. Group risk assessment tools exist for group violence (e.g., the SLEIPNIR version 2.0 Organized Crime Groups Capability Measurement Matrix; RCMP, 2010) – these include factors applicable to groups, such as group discipline and cohesion. However, evidence from the literature and practice suggests an increasing need to bridge these two levels of assessment. Specifically, there is a need to bring both individual and group assessment together to inform a comprehensive approach to GBV risk assessment and management for threat assessment specialists.
Researchers have argued for the importance and potential usefulness of a SPJ tool for GBV, specifically for terrorism (Monahan, 2012; Roberts & Horgan, 2008). This argument has been echoed in professional practice, but has expanded to other forms of GBV, including organized crime and freeman sovereign citizen groups (Burton & Amat, 2013; Van Allen, 2012). Currently, there are no sources (to my knowledge) that call for an actuarial measure for the assessment of GBV. Actuarial measures are dependent on large data sets of offenders from a specific target group. There is insufficient data to support an actuarial tool for the assessment of GBV at this time.

There are no formal threat assessment tools for the assessment and management of GBV for individuals (i.e., that spans the various forms of GBV). There are several threat assessment tools used for the assessment of groups (i.e., tools that identify the level or risk or threat of an entire group). One tool used in Canada by the Royal Canadian Mounted Police (RCMP) is the SLEIPNIR version 2 Organized Crime Groups Capability Measurement Matrix. The SLEIPNIR provides analysts with a framework to develop a matrix that informs assessments of groups. Attributes coded in the assessment include “Discipline” and “Cohesion” and are rated for the overall attributes of the group to determine the level of threat of the group in question (RCMP, 2010). For example, the SLEIPNIR is used to assess the level of threat posed by the Hell’s Angels in the Lower Mainland of British Columbia. In Canada, cases where group-level assessments have limited application or do not apply, such as when determining if an individual member of a group should serve as a confidential informant for law enforcement, or if an individual should receive witness protection for providing information on the illegal activities of their group, contemporary threat management is often used (Burton & Amat, 2013). In this context, contemporary threat management refers to applying an unstructured assessment process based on professional experience and peer-consultation.

Although there are no SPJ tools to assess individual risk for GBV, there are currently two SPJ tools that are intended to assess risk for terrorism specifically. Both measures—the Violent Extremist Risk Assessment (VERA, Pressman, 2009; VERA-2, Pressman & Flockton, 2010) and the Structured Professional Guidelines for Assessing Risk of Extremist Offending (ERG 22+, National Offender Management Service, 2011)—focus on the assessment of risk for extremist violence. Although Pressman & Folkton (2014)
make reference to studies supporting the utility of the VERA in high-secure correctional samples and face validity of the indicators by correctional and security experts with experience of working with terrorists, none of this research has been published in the peer-reviewed literature to date. One published study by Beardsley & Beech (2013) examined the utility of the VERA in the assessment of extremist violence of the application of the VERA to five terrorism case studies. The results also indicated the level of agreement for ratings on each of the VERA items by two raters on each of the five cases was 85.7%, and Kappa values for each of the cases were all > 0.76 ($p < .0001$). No published literature has examined the validity of the tool. The ERC 22+ has not been validated in published literature. Although there is only a limited research base for the VERA and no published evidence for ERG 22+, both can potentially provide useful guidance in the area of extremist violence. The two existing assessment tools are limited to extremist groups and extremist members.

There is also one tool that has been applied to the assessment and management of gang violence, the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004). The LC/CMI is type of SPJ tool – as it relies on empirically guided judgment or anchored clinical judgment - designed to assess both static and dynamic risk factors related to future violence. Guay (2012) evaluated the LS/CMI in a sample of gang and non-gang involved offenders in Canada. The results indicated that gang involved offenders evidenced a greater number of criminogenic risks and needs on the LS/CMI compared to their non-gang counterparts, and the LS/CMI was predictive of new arrests and new convictions of gang involved offenders for any new convictions (AUC = .712, $p < .01$) and any new violent convictions (AUC = .703, $p < .01$) during follow-up (mean follow-up 1024 days).

1.2.2. The Application of Structure Professional Judgment

In an article recently published in Psychology, Public Policy, and Law, Monahan (2012) offered four principal conclusions on the status of individual risk assessment of terrorism. Monahan concluded that (1) a prerequisite of work in this area is to specify the nature and scope of risk for terrorism; (2) a SPJ approach to risk assessment may be useful in assessing risk for terrorist violence, albeit the content of a SPJ tool would differ
substantially from existing instruments and guidelines; (3) the highest priority for research is to identify robust individual risk factors for terrorism; and (4) researchers require access to retrospective information on known groups of terrorists and non-terrorists to achieve these goals.

Other scholars have reinforced Monahan’s (2012) conclusions. Roberts & Horgan (2008) stated that there is an “urgent” need to prioritize the use of counterterrorism resources. The authors suggested that psychology can play a critical role in the development of a risk assessment tool that would assist counterterrorism professionals in effectively managing terrorism risk by triaging, or prioritizing, the highest risk cases. Specifically, Roberts and Horgan suggested the development of SPJ guidelines for the assessment of terrorism risk.

Similarly, Horgan (2011) argued that it is essential to consider influences beyond the individual, such as collective identity, in understanding terrorism and terrorism violence by individuals (see also Stevens, 2005). Horgan reiterated the pressing need for research in this area and called for a transparent and systematic evaluation of terrorist violence that is interdisciplinary and policy-relevant. Further, Gudjonsson (2009) stated that traditional risk assessment tools (i.e., HCR-20; Webster et al., 1997) have limited applicability to individual members of terrorist groups. Gudjonsson suggested that terrorism risk assessment should utilize multi-modal functional analysis and case formulation that incorporates knowledge of social psychology as well as individual-level processes specific to risk for terrorism. The method he suggested was in line with the method of SPJ guidelines for assessment and management of risk (see also Pressman & Folkton, 2014).

The current project utilized Monahan’s (2012) four principal conclusions for individual risk assessment of terrorism as a framework for the development and evaluation of SPJ guidelines for GBV. This project directly addresses Monahan’s (2012) four principal conclusions with empirical evidence and expanded Monahan's conclusions from exclusively individual risk for terrorist violence to more generally individual risk for members of violent groups, or GBV.
Although Monahan’s conclusions and others (Gudjonsson, 2009; Horgan, 2011; Roberts & Horgan, 2008) highlight the need for systematic research and assessment on terrorist violence, there is a body of literature that suggests that this need is not limited to terrorism. Rather, research suggests that Monahan’s conclusions may also apply to individual’s who are members of other groups, outside of terrorist organizations, who pose a risk for violence.

Pynchon and Borum (1999) argued for the role of social psychology, and specifically group-level processes, in understanding and assessing risk for violence of individuals who identify with a larger group. Notably, the authors did not limit their argument for the utility of group-level factors in violence risk assessment to terrorist violence. Pynchon and Borum suggested that group-level processes would influence an individual’s violent behaviour for any form of violent group, such as terrorist groups and gangs. The authors stated that an individual’s engagement in GBV could be understood on a spectrum that ranged from non-violence, or helpful behaviour (e.g., charities), to violent groups (e.g., terrorist and gangs) through principles from social psychology. Pynchon and Borum suggested that principles from social psychology (i.e., group theory, group dynamics, and group interaction) provide a basis for understanding the impact one’s group will have on the individual’s engagement in behaviour, including violent behaviour.

There is overwhelming support that group-level processes influence an individual’s behaviour (e.g., Ellemers, Spears, & Doosje, 2002). Ellemers and colleagues stated, “there is a substantial body of research reporting on phenomena that illustrate the powerful impact of people’s social identities on their perceptions, emotions, and behaviour” (p. 163). These processes include multiple levels: individual (or self), intra-individual (or self interactions with groups), inter-individual (or group-level interactions), and societal (or group in relation to society; e.g., Decker, 1996,).

There is some empirical evidence that has supported the role that group-level factors play in an individual's violent behaviour, but this has been primarily limited to gang-violence (Barnes, Beaver, & Miller, 2010; Bouchard & Spindler, 2010). While empirical support has been limited to gang violence, researchers of other forms of violent groups have drawn similar conclusions in reviews and theoretical papers. Arena & Arrigo (2005)
argued that behaviours in terrorist groups are a result of group interaction. The authors
ably state: “psychological researchers have been guilty of committing the fundamental
attribution error…in short, they overestimate the internal causes for terrorist behaviours”
(p. 487). Arena & Arrigo add that the application of group interaction to understanding
terrorist violence can be extended to other forms of GBV, such as youth gang violence.
There has also been a “blurring of lines” among violent groups in the literature. In line
with Pynchon and Borum’s (1999) argument that groups, whether violent or non-violent,
exist on a spectrum, researchers have suggested that different types of violent groups,
such as cults, terrorist organizations, and gangs, have similar characteristics,
development, and operations (Dawson, 2010; Gorringe, 2006; Lalich, 2009; Lichtenwald,
Steinhour, Perri, 2012; RCMP, 2008; Wright & Wright, 1982). In sum, they argue that the
same processes apply across different contexts, across different types of groups.

There are two main arguments against the comparison of violent groups. Some authors
argue that not all violent groups pose the same level of risk. For example, a multi-
national terrorist organization with sizeable financial support and firearms likely poses a
greater risk for widespread life-threatening violence than a low level street gang with a
few members who have no weapons. There is considerable evidence and well-
established models that specify the range of risk that groups as a whole pose (e.g.
Chermak, Freilich, & Shemtob, 2009; RCMP, 2010; Zhivan, 2012). At a group-level,
some groups pose a greater threat than others. The aim of this project is to examine not
just the group-level, but to examine the individual-level – the individual within the group.
Group-level and group process variables are necessary, but not sufficient in assessing
individual risk for GBV (Albanese, 2008; Horgan, 2011). I argue that similar group-level
processes play a role in the individual risk for GBV, regardless of the type of violent
group being assessed.

Other authors who argue against the comparison of multiple types of violent groups state
that violent groups differ substantially in their motivations for violence. There is no doubt
that different groups can have different motivations for violence (Gupta, Horgan, &
Schmid, 2009; Short, 2004; Lichtenwald et al., 2012). One group may be motivated to
commit violence for ideological reasons, while others do so for revenge, reputation, or
even economic gain. The same is true of motivations for violence at the individual-level.
For example, one individual may be motivated to engage in intimate partner violence to control or change their partner, whereas another person may assault their partner for release or expression. This does not mean that intimate partner violence is different, just that there are different motivations for the violent act (i.e., equifinality finality). Motivations for violence will play a role in assessing and managing risk for GBV - particularly for case formulation - but motivations do not discount groups from being subjected to the same group-level processes as others. Social psychological phenomena will impact all groups, regardless of motivations.

Taken together, traditional methods of risk assessment have limited applicability to individuals who identify with larger groups; this limitation is a result of the systematic and unique influences all groups have on an individual’s behaviour (Gudjonsson, 2009; Horgan, 2011; Monahan, 2012; Roberts & Horgan, 2008). This is not to say that only group variables are important in understanding individual behaviour. Rather, group-level factors in addition to individual-level variables associated with violence risk need to be considered.

I hypothesize that risk factors for GBV will be similar across different types of groups. There is evidence to suggest that the usefulness of examining risk factors that are applicable to all groups, not just terrorist organizations, warrants careful examination (e.g., Arena & Arrigo, 2005; Barnes et al., 2010; Gorringe 2006; Monahan, 2012; Pynchon and Borum, 1999). Such a model would have general application to a variety of groups, such as youth gangs, religious fundamentalist terrorists, organized crime, environmental bandits, sports teams, etc. This hypothesis is tested on a limited basis in the current project. Evaluation of this hypothesis will be dependent on future systematic empirical examination of the individual risk of GBV across different types of groups.

It is my opinion that limiting our scope to the examination of one type of violent group at this time would be omitting important information from other areas that may help inform our understanding, assessment, and management of individuals who are at risk for individual violence among a variety of groups. Monahan (2012) made a similar suggestion, stating that work from gang literature on affiliation can help inform our understanding of the role of affiliation on the risk of terrorist offenders.
To date, there are no systematic and empirically based guidelines for making decisions about risk for future violence for individuals who engage (or may engage) in GBV. No authors have conducted a systematic review of the social science literature to translate our knowledge into a practical model for the evaluation of GBV. In the current project, I applied a systematic review of the literature and consultation with experts to determine what factors are associated with risk for GBV. The aim of this review was to inform systematic and empirically-based guidelines to aid in the assessment and management of GBV.
Chapter 2.

Results, Part 1: Development of the Multi-Level Guidelines

2.1. Campbell Collaboration Methodology

I conducted a systematic and standardized review of the literature on GBV and related social sciences literature (e.g., group cohesion) using Campbell Collaboration methodology to examine what factors are associated with GBV. A Campbell Collaboration is the equivalent of a Cochrane Review for the criminal justice system (Campbell Collaboration, 2011; Cooke & Johnstone, 2010). The methodology explicated a comprehensive examination of the existing literature, in this case, GBV, that is transparent, can be replicated, and minimizes bias.

According to Campbell Collaboration (2011), a Campbell Collaboration systematic review must include clear inclusion/exclusion criteria, an explicit search strategy, systematic coding and analysis of included studies, and, where possible, the inclusion of meta-analyses. A committee of social scientists approved the method of the current review.

I used broad inclusion criteria for both violent and non-violent group-based behaviour given the limited explicit examination of GBV to date (see Monahan, 2012). Specifically, there was no date restriction on the review, both published and unpublished (e.g., PhD theses) sources were considered, the literature included both experiments and non-experiments, and research in languages other than English. For rationale regarding the broad inclusion criteria see Table 2.1.

1 R. Corrado, Ph.D; S. D. Hart; P. R. Kropp, Ph.D; R. Roesch, Ph.D
Table 2.1. Inclusion criteria for the Campbell Collaboration Review

- There was no date restriction on the review.  
  Rationale: All relevant literature was examined given the limited examination of the influence of group membership on behavior.
- Both published and unpublished (e.g., PhD thesis) were considered.  
  Rationale: In line with Campbell Collaboration guidelines, this helps minimize publication bias in the review.
- Literature included both experiments and non-experiments (e.g., book chapters, theoretical papers).  
  Rationale: All types of research methodology on this area will be informative for the current review, especially theoretical papers on group-based violence given the limited literature in this area.
- Research in languages other than English were included  
  Rationale: This added to the international scope of included literature. Translations were obtained as necessary.
- Types of group included in the review will be those specified by the WHO (2002): terrorist groups (including religious, political, and environmental terrorists), gangs, organized crime, bandits, and hooligans. Group violence not specified above, such as spontaneous group violence (e.g., riots) and racial and political genocide were excluded.  
  Rationale: The WHO list of subtypes of group-based violence is assumed to have international relevance for the development of the measure. The excluded types of violence are not relevant to the scope of the current measure.

The Campbell Collaboration included two sets of explicit search terms used to search 17 relevant social science databases to ensure broad coverage of social sciences and the law.

The following search terms were entered into the appropriate databases:

1. ((group-based OR group dynamics, group action OR collective action OR

2. group membership OR social identity) AND (violence or assault or aggressive))

3. ((violence OR assault OR aggressive) AND (terrorist OR gang OR organized crime OR bandits OR hooligans OR cult)) OR ((terrorist OR gang OR organized crime OR bandits OR hooligans OR cult) AND (risk assessment OR risk management OR risk))

The search included research databases available through the Simon Fraser University Library. Databases were selected to ensure broad coverage of social sciences and the law of both published and unpublished literature. All databases were selected for relevance to the review questions through a search of the library databases and their respective descriptions. When database searches produced over 500 sources with the
above search terms the review was limited to the top 100 sources returned by the search. This limit was set due to time and researcher constraints— additionally this decision was informed by the overall low rate of inclusion of the top 100 sources returned by the search (less than 10% in each case). The review was conducted between August and December 2011.

The included databases were:

Best Case: A Canada Law Book, Canadian Research Index, Cochrane Database of Systematic Reviews, Curia: Court of Justice of the European Union, Database of Abstracts of Reviews of Effectiveness, Database of Abstract of Reviews of Effects, Digital National Security Archive, Humanities & Social Sciences Indexes, International Encyclopaedia of the Social & Behavioral Sciences: IEBS, LegisInfo, Mental Measurements Yearbook, PsycARTICLES, PsycBOOKS, PsycINFO, PubMed Central, Social Theory, Sociological Abstracts. Additionally, government websites were searched for relevant reports from the following countries: Australia, Belgium, Canada, Denmark, France, New Zealand, Sweden, United Kingdom, and United States.

Experts in the field were also solicited for relevant sources related to risk for GBV. Sources provided by December 2011 were included in the analysis.

2.1.1. Data screening

The data were screened for inclusion in the review based on the three-step process implemented by Gadon, Cooke, & Johnstone (2005): (1) the titles of citations were scanned for relevance. Relevant titles were retained in the evaluation process; (2) abstracts of the retained titles were reviewed. Relevant titles were retained in the evaluation process; and (3) full-text articles were evaluated. Relevant sources were retained for the synthesis for findings.

Typically, two independent researchers conduct Campbell Collaboration Reviews. In the current review, I made initial judgments regarding the inclusion of sources due to researcher constraints. To ensure the initial judgments of inclusion were not conducted in a biased manner, a random sample of 10 initially judged ‘included’ sources and 10 initially judged ‘excluded’ sources were reviewed by Drs. Stephen D. Hart and P. Randall
Kropp. Intraclass correlations for single ratings were excellent ($\text{ICC}_1 = 0.88$).\(^2\) Discrepancies in ratings were discussed until agreement was reached. Other Campbell Collaborations (see Gadon et al., 2005) have employed similar methodology. This process was used in place of alternatives, such as a research assistant, to ensure expert opinion was considered in assessing for bias.

2.1.2. Data management

Brief citations from each database search were saved into a Microsoft Excel document (i.e., author(s), titles, database used, and search terms used to locate the source). Duplicate citations were recorded for analysis of overlap of the databases used. The Excel document indicated if the sources met inclusion criteria for the current review.

2.2. Results

The search resulted in 1,973 potentially eligible sources. Of the 1,973 sources 136 (7\%) were duplicate sources and were excluded. The remaining 1,837 sources’ titles were reviewed and 1,074 sources were retained (763 were excluded as irrelevant). The remaining 1,074 sources’ abstracts were reviewed and 753 sources were retained (321 were excluded as irrelevant). The remaining 753 sources’ full-text were reviewed and 201 sources were retained for data coding (552 were excluded as irrelevant). Of the initially eligible sources, a total of 201 sources (10\%) were retained for data coding. The full-text of 1 of the 201 sources was unable to be located (0.5\%) and 3 (1.5\%) were unable to be translated into English ($N = 197$).

2.2.1. Data coding

Each of the retained sources was coded on the basis of four key features: (1) identification of key factors related to GBV, (2) violent, non-violent, or both, (3) types of

\(^2\) See Methods section for further information on intraclass correlation statistical analysis.
group(s) examined or referenced (terrorist, gang, organized crime, hooligan, bandit, other), (4) level(s) of variable (self, intra-group, inter-group, societal, other).

2.2.2. **Scope of the existing evidence**

The retained articles represented the inclusion criteria set out prior to the Campbell Collaboration search (see Table 1). The majority of the 197 sources identified in the current systematic and comprehensive review of the literature were articles in peer-reviewed journals \((n = 122; 62\%)\). Of the peer-reviewed articles on GBV, the majority were review articles \((n = 69; 58\%)\), 13 utilized qualitative methodology \((11\%)\), 13 utilized quantitative methodology \((11\%)\), 13 were theoretical articles \((11\%)\), 10 were case studies \((8\%)\), and 2 were experiments \((2\%)\).

The remaining sources that were not articles in peer-reviewed journals \((n = 75; 38\%)\) were comprised of 22 books/book chapters \((29\%)\), 22 were government reports \((29\%)\), 15 were encyclopaedia entries \((20\%)\), 8 were case law in Canada \((10\%)\), 6 were dissertations \((8\%)\), 2 were newspaper articles \((3\%)\), and 2 were existing SPJ tools \((2\%; i.e., ERG 20+, National Offender Management Service, 2011; VERA, Pressman, 2009)\).

The publication date of the sources ranged from 1963 to 2011. Of the 197 sources, 167 \((83\%)\) referenced violent group behaviour, 24 \((12\%)\) covered both violent and non-violent groups, and 6 \((3\%)\) covered exclusively non-violent group behaviour. Gender was examined in 41 \((21\%)\) of the sources: of the 41 sources, 26 examined only males \((63\%)\), 6 examined only females \((15\%)\), and 9 \((22\%)\) examined both males and females. Youth were represented in 45 \((23\%)\) of the included sources.

The data were international in scope. The sources reviewed represented samples from 29 countries (see Table 2.2). In addition, one source had a sample of several European Countries and another source had a sample of several Central American countries. An additional 20 sources were comprised of data that was international in scope \(i.e., \) not limited to one country or continent).
Table 2.2. Countries represented in Campbell Collaboration review

<table>
<thead>
<tr>
<th>Country</th>
<th>Country</th>
<th>Country</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>Germany</td>
<td>Lebanon Netherlands</td>
<td>Sri Lanka</td>
</tr>
<tr>
<td>Australia</td>
<td>Greece</td>
<td>New Zealand</td>
<td>South Africa</td>
</tr>
<tr>
<td>Canada</td>
<td>India</td>
<td>Nicaragua</td>
<td>Spain</td>
</tr>
<tr>
<td>China</td>
<td>Ireland</td>
<td>Palestine</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Columbia</td>
<td>Israel</td>
<td>Peru</td>
<td>United States of America</td>
</tr>
<tr>
<td>El Salvador</td>
<td>Italy</td>
<td>Philippines</td>
<td>Venezuela</td>
</tr>
<tr>
<td>France</td>
<td>Japan</td>
<td>Scotland</td>
<td>Yugoslavia</td>
</tr>
</tbody>
</table>

The sources represented multiple levels of assessment: individual (or self), intra-individual (or self interactions with groups), inter-individual (or group-level interactions), societal (or group in relation to society). The sources also represented all groups identified by the WHO (2002) as violent groups and several additional violent and non-violent groups that fit within the definition of GBV: gangs, terrorist organizations, religious groups, ethnic and cultural groups, political groups, peer groups, cults, bandits, sports teams, hooligans, organized crime, protesters, genocide groups, national groups, fraternities, specific groups of employees or prisoners, guerrilla groups, extremist groups, military, freeman sovereign citizens, and ideological groups.

2.3. Discussion

2.3.1. Group-Based Violence Risk Factors

The results of the Campbell Collaboration review indicated a number of unique factors in the assessment and management of risk for GBV. I synthesized these findings within a nested (or ecological) model, similar to that of Brofremmbrener’s biopsychosocial model of development (Bronfenbrenner & Morris, 1998) along with my senior supervisor, Dr. Stephen D. Hart, and the third author of the guidelines, Dr. P. Randall Kropp. Others (i.e., Gorringe, 2006) have used Brofremmbrener’s model to understand gang violence. The model we developed (see Figure 2.1) is not intended to reflect a statistical model or to be considered the only model to conceptualize the factors associated with GBV.
There were four distinct domains represented in the GBV literature. The first domain, *Individual*, refers to risk factors that are relatively independent from that of the individual’s group membership. The second domain, *Individual-Group*, refers to risk factors related to the individual in relation to the group and captures an individual’s identity, attitudes, and role in relation to the group. The third domain, *Group*, refers to risk factors related to group process and structure. The final domain, *Group-Societal*, items refers to risk factors related to external or peripheral contributions to violence risk. This domain captures the societal and intergroup-level influence on groups and individual members of groups to engage in violence.

The results indicated that the scope of the literature for GBV is diverse and primarily comprised of articles published in peer-reviewed journals, with the majority of articles a review of the current findings, and about equal presentation of articles that utilized quantitative, qualitative, theoretical, and case study methods. The sources spanned almost 50 years of research on violent and non-violent groups internationally. Both females and youth were represented in the current review. The sources also represented multiple levels of assessment (individual, intra-individual, inter-individual, societal) and covered a variety of groups that fit within the definition of GBV.

The current review did not systematically assess for quality of sources included in the development of the GBV Model (see Figure 1). Quality of sources was considered during synthesis of the findings into the model and factors. However, this is a limitation of the
current review. Authors argue that the existing literature is limited, and importantly limited by the lack of quality research on risk factors associated with GBV (see Monahan, 2012).

I would like to underscore that this project examines risk for violence by a member or affiliate of a group, not risk for becoming a member of a violent group. Not all members of violent groups are at risk for violence (Baloch, 2010; Horgan & Taylor, 2011; RCMP, 2009). For example, Angie and colleagues (2011) examined 29 ideological groups online and found that not all members of violent ideological groups included in their study expressed violence or violent intent. The authors reported that only a subgroup of members of the violent groups reported engaging in violent acts. Horgan and Taylor (2011) add that terrorism and radicalism are not synonymous. The authors specify that engagement in terrorist ideologies and groups are not the same as engaging in terrorist violence (see also Baloch, 2010). There are similar findings in the gang literature. Taylor (2004) summarized the literature on gang violence stating that researchers have found that the majority of gang members are only peripherally involved in violence, the majority of violent acts by a gang are committed by a small percentage of members. This is an important distinction to make, as many scholars include risk for being a member of a violent group as risk for terrorism, or gang violence. The proposed factors are intended to help distinguish members or affiliates at low risk to engage in violence from those members or affiliates at moderate or high risk to engage in violence.

2.3.2. **The Multi-Level Guidelines (MLG)**

Based on the above review, Drs. Stephen Hart and P. Randall Kropp and myself developed the draft version of the *Multi-Level Guidelines*, or MLG (Cook et al., 2013). The MLG was developed as a comprehensive, management-oriented set of guidelines of GBV for threat assessment specialists. The MLG is comprised of 20 risk factors in four domains that are assessed as part of a structured and systematic process of evaluating violence risk. See Table 2.3 for a list of the MLG factors and domains. The MLG is both evidence- and consensus-based. International experts in the field of GBV
were consulted while the draft was under development. Expert feedback was solicited from a diverse group of experts with various training backgrounds in social sciences and from different countries. Experts had specialization in a variety of types of groups (i.e., terrorism, gangs, cults, environmental terrorism, organized crime) and each expert reviewed the guidelines for utility and comprehension, this included the application of the guidelines to a case they had been involved with. Experts included members of my doctoral thesis committee: Drs. Raymond Corrado and Ronald Roesch. Additional experts have also provided feedback at conferences the draft was presented at (i.e., American Psychology-Law Society, Association of European Threat Assessment Professionals), which was not included in the draft guidelines, but will be incorporated into the published manual.

3 Detective Gwyneth Amat, Calgary Police Service, Calgary, AB, Canada; Randy Borum, Ph.D., University of South Florida, Tampa, FL, USA; Martin Bouchard, Ph.D., Simon Fraser University, Burnaby, BC, Canada; Christopher Dean, Ph.D., National Offender Management Service, HM Ministry of Justice, London, United Kingdom; Kevin S. Douglas, LL.B., Ph.D., Simon Fraser University & Mental Health Law & Policy Institute, Burnaby, BC, Canada; Detective/Constable Michael Fealing, Vancouver Police Department, Vancouver, BC, Canada; John Horgan, Ph.D., International Center for the Study of Terrorism & Pennsylvania State University, University Park, PA, USA; Jack Martin, Ph.D., Simon Fraser University, Burnaby, BC, Canada; J. Reid Meloy, Ph.D., ABPP, San Diego, CA, USA; John Monahan, Ph.D., University of Virginia, Charlottesville, VA, USA; Russell Palarea, Ph.D., Operational Psychology Services, Washington, DC, USA; Robert Rennebohm, M.S.W., Boston, MA, USA; Stephen Wright, Ph.D., Simon Fraser University, Burnaby, BC, Canada

4 For example, Mario Scalora, Ph.D., University of Nebraska-Lincoln, Lincoln, NB.
Table 2.3.  Group-Based Violence Risk Factors by Domain

<table>
<thead>
<tr>
<th>Domain</th>
<th>Categories</th>
</tr>
</thead>
</table>

Note: For item descriptions see the Multi-level Guidelines (MLG) User Manual (Cook et al., 2013). These are the factors of the MLG, if there are additional factors in any of the domains the other factors should also be considered in the assessment and management of risk.

The MLG is applied in six systematic and standardized steps. The six steps of the MLG are similar to those utilized in other widely used SPJ tools (e.g., HCR V3, Douglas et al., 2013; SAM, Kropp, Hart, & Lyon, 2008). First, evaluators collect the information to draw opinions about a case. Second, evaluators determine the presence of the 20 risk factors. Each risk factor is coded on a three-point scale (“absent”, “possibly or partially present,” “present”). Each factor is coded for presence for the recent status of the case and for the previous status of the case. Third, evaluators determine the relevance for the 20 risk factors.
factors for future violence risk. Each risk factor is coded on a three-point scale ("not relevant", “possibly or partially present”, “relevant”). The relevance of factors are conceptualized as motivators, disinhibitors, or destabilisers in formulating risk for future violence. Forth, evaluators develop risk scenarios to inform management plans needed to prevent future violence in each case. Each management strategy is developed to include the nature of the potential future violence, the severity of the violence, the imminence of the violence, the frequency or duration of the violence, the potential victims of violence, and the likelihood that violence may occur. Fifth, evaluators recommend management strategies to prevent the plausible scenarios developed in step four. Finally, evaluators make conclusory opinions about the case: case prioritization, risk for future violence, risk for life-threatening violence, imminent violence, and the likely victims. See Cook and colleagues (2013) for a complete description of the use of the MLG.

Although SPJ guidelines are inherently empirically based (i.e., SPJ’s are developed through an evaluation and synthesis of the relevant literature and consensus in clinical practice), it is not sufficient to rely on the authors’ evaluation and synthesis of the literature and practice as evidence of the effectiveness of the guidelines.
Chapter 3.

Results, Part 2: Evaluation of the Multi-Level Guidelines

3.1. Evaluation of Risk Assessment Guidelines

It is imperative that all risk assessment guidelines are evaluated for utility, reliability, and validity in the assessment and management of the specific form of violence risk for which the guidelines were developed. We need to determine if the guidelines are useful for threat assessment professionals in assessing and managing risk of violence. We need to determine if users can reliably use the guidelines, in terms of the ratings of the factors, conclusory opinions, case formulations, and management strategies. We also need to determine if the judgments made on the basis of these guidelines are related to future violence, predictive validity, are related to similar assessments of violence risk, concurrent validity, and are dissimilar to assessments of violence or other forms of risk unrelated to the specific form of violence risk for which the guidelines were developed, discriminant validity.

An important first step in the evaluation of a new set of guidelines is to assess for utility and reliability of the tool (Douglas & Kropp, 2002). Initial evaluations of other SPJ tools have used similar methodology, such as the Historical, Clinical, Risk Guide version 3 (HCR V3, Douglas et al., 2013), the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER, Kropp, Hart, & Belfrage, 2010), the Short-Term Assessment of Risk and Treatability (START; see Doyle, Lewis, & Brisbane, 2008) and the Child Abuse Risk Evaluation (CARE; see Agar, 2003). Without evidence of clinical and practical utility and reliability sophisticated evaluations of the tools are valueless. As put by Douglas and Kropp (2002), “reliability is a requisite to validity” (p. 649).
Utility may be assessed through surveys about the usefulness and practical implementation of the guidelines in practice – a customer satisfaction survey, of sorts. A part of assessing utility is determining the ability for threat assessment professionals to be trained on these tools. There are few published studies that have examined utility of and training on risk assessment guidelines. Studies have demonstrated that training for SPJ guidelines has improved risk assessment skills for psychologists, psychiatrists, social workers, occupational therapists, police, and other threat assessment professionals (Doyle et al., 2008; McNiel et al., 2008; Reynolds & Miles, 2009; Storey, Gibas, Reeves, & Hart, 2011).

One study, conducted by Storey and colleagues (2011), directly evaluated outcomes of a threat assessment training course. Storey and colleagues evaluated a violence risk assessment training program for police and other criminal justice professionals on participant’s knowledge about violence risk assessment, competence in conducting comprehensive violence risk assessment, and self-reported competence in violence risk assessment. Storey and colleagues evaluated knowledge, skills, and competence through a pre- and post-training examination. The results indicated that participant’s knowledge significantly increased, their skills for identifying risk factors significantly increased, and they were significantly more likely to provide a correct rating, per expert ratings, on case vignettes post-training. Further, the quality and quantity of risk management recommendations increased. Participants’ self-reported confidence, competence in risk assessment, and familiarity with risk factors also all increased post-training. However, participants’ self-report of confidence, competence, and familiarity did not correlate with improvement in their scores on the multiple-choice test given. Storey and colleagues suggested that self-reported confidence might have exceeded objective skill level.

While training programs have been effective for various forms of violence risk to a wide variety of professionals, Carroll (2007) expressed concerns about barriers that contribute to gaps between research and practice, the most significant of which was that users felt that SPJ tools, or a specific SPJ tool, added nothing of utility to their practice. The evidence presented by Carroll primarily comes from experiences in Australia and he aptly points out that the research in this area is only beginning to develop; he
nonetheless raised an important point. Without the belief that tools or guidelines have utility, there will be little use by front-line evaluators for whom the tools were developed. Carroll states: “The challenge is to meet the needs and address the concerns of front-line clinicians, while still ensuring a scientifically valid basis for the decision-making process that flow from such techniques” (p. 302). As such, investigations of utility should be the foundation for evaluations of SPJ tools.

Reliability of guidelines may be assessed through examining the similarity, or agreement, of ratings made by two or more trained evaluators who have applied the guidelines to a series of cases. Evaluations of widely used SPJ tools have found comparable findings on reliability estimates for domain, total scores, and conclusory opinions. Researchers have demonstrated good interrater reliability coefficients for the Historical, Clinical, Risk Guide, or HCR-20 (Webster et al., 1997), the Spousal Assault Risk Assessment Guide, or SARA (Kropp, Hart, Webster, & Eaves, 1999), the Guidelines for Stalking Risk Assessment & Management, or SAM (Kropp et al., 2008), and the Short Term Assessment for Risk and Treatability, or START (Webster, Martin, Brink, Nicholls, & Desmarais, 2009), among others.

As there are hundreds of papers and conference presentations evaluating the reliability of commonly used SPJ tools. I will focus on examples for the HCR-20, SARA, SAM, and START. All studies reviewed report interrater reliability using Intraclass Coefficients (ICC). Douglas and Ogloff (2003) found good reliability for HCR-20 total scores (ICC = .85), the H scale (ICC = .90), C Scale (ICC = .79), and moderate ICCs for the R scale (ICC = .47). They also found good ICCs for structured clinical risk judgments (ICC = .61). Kropp & Hart (2000) found good reliability the SARA total scores (ICC = .84). For the SAM, Kropp and colleagues (2011) found good reliability for the SAM Nature of stalking factors (ICC2 = .77), Perpetrators risk factors (ICC1 = 0.76), Victim vulnerability factors (ICC1 = .44), and total scores (ICC1 = 0.82). Kropp and colleagues (2011) findings

---

5 Note. The HCR-20 is the most widely researched SPJ tool (Guy, Packer, & Warken, 2012) and there is considerable variability in the quantity and quality of research of SPJ tools generally (see Guy, 2008).

6 ICC1, intraclass correlation coefficient for single ratings; ICC2, intraclass correlation coefficient for average ratings.
for concursory opinions ranged from fair to good: Case prioritization ($\text{ICC}_1 = .39$), Continued Stalking ($\text{ICC}_1 = .71$), and Serious Physical Harm ($\text{ICC}_1 = .44$). For the START, Desmarais, Nicholls, Wilson, and Brink (2011) found good to excellent START total scores for Strength ($\text{ICC}_2 = .93$) and Vulnerability ($\text{ICC}_2 = .95$) facets and violence risk estimates ($\text{ICC}_2 = .85$).\(^7\)

The above studies all used two or three trained raters coding participants’ files and/or interview to obtain ratings on a large number of cases in various settings, primarily forensic psychiatric inpatient and outpatient services. Two other studies have investigated ICCs for the RSVP (Sutherland et al., 2012; Wilson, 2013) with smaller sets of practice cases (6-10 cases) immediately following training of mental health professionals, all of whom had varying degrees of experience in risk assessment and the use of SPJ tools (Sutherland et al., 2012, $N = 28$; Wilson, 2013, $N = 17$). Sutherland and colleagues (2012) found fair reliability for the RSVP Sexual Violence History Domain ($\text{ICC}_2 = .45$), Psychological Adjustment Domain ($\text{ICC}_2 = .52$), and Mental Disorder Domain ($\text{ICC}_2 = .51$), Social Adjustment Domain ($\text{ICC}_2 = .52$), and Manageability Scores ($\text{ICC}_2 = .48$).\(^8\) ICCs for structured clinical risk judgment were fair to good (Case Prioritization, $\text{ICC}_2 = .62$; Risk of Serious Physical Harm, $\text{ICC}_2 = .69$; Immediate Action Required, $\text{ICC}_2 = .43$). Similarly, Wilson (2013) found poor to good reliability for the RSVP Sexual Violence History Domain ($\text{ICC}_1 = .07$), Psychological Adjustment Domain ($\text{ICC}_1 = .45$), Mental Disorder Domain ($\text{ICC}_1 = .45$), Social Adjustment Domain ($\text{ICC}_1 = .48$), $\text{ICM}_1 = .69$), Manageability Scores ($\text{ICC}_1 = .78$) and Total Score ($\text{ICC}_1 = .56$). ICCs for structured clinical risk judgment were fair to poor (Case Prioritization, $\text{ICC}_1 = .29$; Risk of Serious Physical Harm, $\text{ICC}_1 = .44$; Immediate Action Required, $\text{ICC}_1 = .21$).\(^9\)

Both Sutherland and colleagues (2012) and Wilson (2013) stated that the lower levels of agreement in their studies, compared to others examining the RSVP and other SPJ tools, may be due to the methodology used (i.e., a smaller set of practice cases with a greater number of raters just trained in the use of the tool whom have varying degrees of

\(^7\) Note: Authors only reported $\text{ICC}_2$ scores.
\(^8\) Note: Authors only reported $\text{ICC}_2$ scores. Domain ICCs represent domain average scores.
\(^9\) Both Sutherland and colleagues (2012) and Wilson (2013) used more conservative interpretation guidelines (Fleiss, 1981) to remain consistent with past studies of the RSVP.
experience). This is notable as the current study employed a similar methodology to Sutherland and colleagues (2012) and Wilson (2013).

While there is often good support for the reliability of the total and domain scores on SPJ tools, there is frequently variability evidenced at the item level. At the item level of the HCR-20, ICC\(_1\) ranged from .01 to 1.00 (Douglas & Ogloff, 2003). At the item level of the SARA, ICC\(_1\) ranged from .45 to .86, with a median ICC\(_1\) of .65 (Kropp & Hart, 2000). At the item level of the SAM, ICC\(_1\) ranged from .33 to 1.00 (Kropp et al., 2011).\(^{10}\) In each of these papers, the authors suggest while the variability in item-level ratings may reflect actual differences in rating, it may also be an artifact, in that, the low ICCs are a function of a restricted range of the items or due to limitations in the information available to raters in the research design (e.g., rating relational or interpersonal items on the basis of file information only). This was also the case in studies using fewer cases and a larger number of raters immediately after SPJ training (i.e., Sutherland et al., 2012; Wilson, 2013). In Wilson’s (2013) study, ICC\(_1\) at the item level for the RSVP\(^{11}\) ranged from .11 to .91 (see also Sutherland et al., 2013, ICC\(_2\) ranged from .20 to .87).

I evaluated the utility and reliability of the MLG through two MLG training courses for criminal justice and mental health professionals. The methodology of the current study was developed in line with other studies evaluating the utility of risk assessment guidelines and of training programs for other forms of violence risk (Agar, 2003; Doyle et al., 2008; McNiel et al., 2008; Reynolds & Miles, 2009; Storey et al., 2011; Wilson, 2013). I had three hypotheses for the evaluation of the MLG: (1) criminal justice and mental health professionals confidence, competency, and knowledge for assessment and management of GBV would significantly increase after learning how to administer the MLG (2) criminal justice and mental health professionals would report that the MLG is a useful set of guidelines for practice and will provided integral feedback for the development of the tool and (3) the reliability of the MLG would be comparable to those found with other SPJ tools.

\(^{10}\)Kropp and colleagues (2011) were unable to calculate two of the SAM Perpetrator risk factors and five of the SAM Victim Vulnerability risk factors due to restricted range.

\(^{11}\)Wilson (2013) was unable to calculate one item (Problems with self-awareness) due to restricted range.
3.2. **Method**

Two MLG training courses for criminal justice and mental health professionals were utilized to implement the study design. Institutional ethics approval was granted prior to recruitment of participants or data collection.

3.2.1. **Participants**

A total of 46 criminal justice and mental health professionals attended the MLG training workshops. Participants were recruited through threat assessment and mental health professional organizations, psychology-law graduate departments, and the professional contacts of the members of my doctoral thesis committee. Potential participants were offered full reimbursement of their training fees (up to $550.00 CAD) as incentive to participate in the research. Participants were also eligible to receive 24 hours of Continuing Education credits, however no one elected to receive CE Credits for their participation.
Table 3.1. Demographics of Course Participants, N = 46

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>27</td>
<td>59%</td>
</tr>
<tr>
<td>Female</td>
<td>19</td>
<td>41%</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(mean)</td>
<td>40.93</td>
<td>9.12</td>
</tr>
<tr>
<td><strong>Country of Residence</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>44</td>
<td>96%</td>
</tr>
<tr>
<td>United States</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>Singapore</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Primary profession</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Law enforcement/corrections</td>
<td>35</td>
<td>76%</td>
</tr>
<tr>
<td>Corporate security</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>Psychologist</td>
<td>3</td>
<td>7%</td>
</tr>
<tr>
<td>Psychology graduate student</td>
<td>6</td>
<td>13%</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Highest degree obtained</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college or university</td>
<td>11</td>
<td>24%</td>
</tr>
<tr>
<td>Associates degree</td>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td>Bachelor degree</td>
<td>20</td>
<td>44%</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>8</td>
<td>18%</td>
</tr>
<tr>
<td>Professional degree</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>Doctorate degree</td>
<td>3</td>
<td>7%</td>
</tr>
<tr>
<td><strong>Threat assessment qualifications</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualified in court as a threat assessment expert</td>
<td>9</td>
<td>20%</td>
</tr>
<tr>
<td>Published research on threat assessment in peer-reviewed journals</td>
<td>4</td>
<td>9%</td>
</tr>
<tr>
<td>Published research on group-based violence in peer-reviewed journals</td>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td>Certified in use of SPJs for threat assessment</td>
<td>21</td>
<td>46%</td>
</tr>
<tr>
<td>Certified in use of other approaches of threat assessment</td>
<td>12</td>
<td>26%</td>
</tr>
</tbody>
</table>

Participants were primarily law enforcement or corrections officers residing in Canada (59% male). The entire sample had some college or university education, with the majority holding a Bachelors degree (44%). The participants had varying degrees of
experience in threat assessment. On average, participants had been conducting threat assessment for 6 years ($SD = 4.6$), with a range of 0 to 18 years. One fifth of the sample had been qualified in court as a threat assessment expert ($n = 9; 20\%$). Few participants had published on threat assessment in peer-reviewed journal ($n = 4; 9\%$) and on group-based violence in peer-reviewed journals ($n = 2; 4\%$), though this is expected given the professional backgrounds of the participants. Twenty-one (46\%) of participants were certified in the use of SPJ for threat assessment prior to attending the course and 12 (26\%) were certified in the use of other approaches of threat assessment (e.g., actuarial tools). Participant demographics are presented in Table 3.1. Participant’s competency and use of threat assessment tools was also varied (see Table 3.2). There was a wide range of experience – the range of total threat assessment reports written was 0 to 1,020 ($M = 74.43$, $SD = 190.41$).
Table 3.2. Experience of course participants with widely used threat assessment instruments, N = 46

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Competent in Use</th>
<th>Number of Reports written</th>
<th>Use in Practice</th>
<th>Frequency of Use Per Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>HCR-20</td>
<td>26</td>
<td>57%</td>
<td>68</td>
<td>87</td>
</tr>
<tr>
<td>SVR-20</td>
<td>15</td>
<td>33%</td>
<td>34</td>
<td>57</td>
</tr>
<tr>
<td>SARA</td>
<td>27</td>
<td>59%</td>
<td>53</td>
<td>110</td>
</tr>
<tr>
<td>SAM</td>
<td>20</td>
<td>44%</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>WAVR-21</td>
<td>12</td>
<td>26%</td>
<td>21</td>
<td>44</td>
</tr>
<tr>
<td>CARE</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>START</td>
<td>4</td>
<td>9%</td>
<td>68</td>
<td>43</td>
</tr>
<tr>
<td>RSVP</td>
<td>15</td>
<td>33%</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>B-SAFER</td>
<td>12</td>
<td>26%</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>SRP</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WRA &amp; ERA</td>
<td>1</td>
<td>2%</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>EARU</td>
<td>1</td>
<td>2%</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>SAVRY</td>
<td>15</td>
<td>33%</td>
<td>17</td>
<td>26</td>
</tr>
<tr>
<td>ERASOR</td>
<td>2</td>
<td>4%</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>PATRIARCH</td>
<td>6</td>
<td>13%</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>VRAG</td>
<td>8</td>
<td>17%</td>
<td>48</td>
<td>66</td>
</tr>
<tr>
<td>STATIC-99-2000</td>
<td>12</td>
<td>26%</td>
<td>50</td>
<td>65</td>
</tr>
<tr>
<td>SAPROF</td>
<td>3</td>
<td>7%</td>
<td>38</td>
<td>44</td>
</tr>
</tbody>
</table>

Note: Participants also listed other threat assessment for which they are competent and use in practice (verbatim responses provided): SLEIPNIR, Path to Intended Violence, JACA, LS/CMI, PCL-R, EARL, STATIC-99, ODARA, NCRTA, TPS model (Greyfell variation – scoring/RCMP approach), Ontario Threat Scale for Organized Crime, Pathway/Contemporary Threat Management, Informal RCMP.

3.2.2. **Description of the MLG Training**

The MLG training course was developed by the authors of the MLG and was modeled on training courses for other similar risk assessment guidelines (e.g., HCR-20) that have been offered for over ten years to hundreds of threat assessment professionals. The modality of training used has been shown to be effective for training threat assessment professionals in the use of SPJ guidelines for various forms of violence risk (e.g., Storey et al., 2011). Public Safety Canada, the Simon Fraser University Mental Health, Law, and Policy Institute, and ProActive ReSolutions, Inc sponsored the training course. One
training course took place in Western Canada and the other in Eastern Canada in Spring 2013.

The course was offered over three consecutive days for a total of approximately 24 contact hours. The first day was didactic. The course program focused on general principles of violence risk assessment and management and the assessment of risk for GBV. On the first day, all participants applied the MLG to a case of GBV and the case was reviewed as a group. The latter two days of the course were practical. The course program focused on the administration of the MLG through the completion of four cases. Participants completed up to five cases over the three-day course.\(^{12}\)

**Materials**

**Utility Questionnaires**

Participants were asked to complete a general evaluation of the MLG and MLG course. Evaluation materials consisted of a pre-training questionnaire and a post-training questionnaire. The pre-training questionnaire consisted of two parts. Part one assessed general demographic questions (i.e., gender, age, country of residence, primary procession, highest degree obtained, and threat assessment qualifications) and questions regarding experience as a threat assessor (i.e., years conducting threat assessment, threat assessment competencies, number of threat assessment reports written, and types and frequency of use of other threat assessment tools). Part two assessed self-ratings of confidence, knowledge of risk assessment, knowledge of risk factors, and competency for assessing an individual’s risk in relation to both general violence and group-based violence. Each of these self-ratings was scored on a 10-point likert scale (see Table 3.3 for list of questions). The post-training questionnaire consisted of the same self-ratings of confidence, knowledge of risk assessment, knowledge of risk factors, and competency assessed in the pre-training questionnaire. The post-training questionnaire asked participants to indicate what forms of GBV they plan on using the MLG with. The post-training questionnaire also included two blank pages for participants

\(^{12}\) A number of participants needed to miss an afternoon or a day of the course due to employment or travel restrictions. For these participants less than five practice cases were completed.
to provide any comments or recommendations on the MLG, the MLG training, or the general approach to the assessment and management of group-based violence, and for the MLG manual.

**Table 3.3. Pre- and post- training self-ratings questions for general and GBV**

<table>
<thead>
<tr>
<th>General Violence</th>
<th>10-point likert rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>How confident are you in your ability to assess an individual’s risk for violence?</td>
<td>1= not at all confident</td>
</tr>
<tr>
<td>How much knowledge do you feel you have about risk assessment for violence?</td>
<td>1= almost none or no knowledge</td>
</tr>
<tr>
<td>How much knowledge do you feel you have about risk factors for violence?</td>
<td>1= almost none or no knowledge</td>
</tr>
<tr>
<td>How competent do you feel you are at assessing an individual’s risk for violence given you current practical and technical skills?</td>
<td>1= not at all competent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group-Based Violence</th>
</tr>
</thead>
<tbody>
<tr>
<td>How confident are you in your ability to assess an individual’s risk for group-based violence?</td>
</tr>
<tr>
<td>How much knowledge do you feel you have about risk assessment for group-based violence?</td>
</tr>
<tr>
<td>How much knowledge do you feel you have about risk factors for group-based violence?</td>
</tr>
<tr>
<td>How competent do you feel you are at assessing an individual’s risk for group-based violence given you current practical and technical skills?</td>
</tr>
</tbody>
</table>

*Note: Questions were asked both pre- and post- training to assess for change in self-assessed confidence, knowledge, and competency in risk assessment of general violence and GBV.*

**MLG Worksheets**

Participants were also provided an MLG Worksheet for each case they completed. The MLG Worksheet was developed in conjunction with the MLG to efficiently and effectively document the assessment process and is intended to be used in conjunction with the MLG manual (Cook et al., 2013). It contains indicators and space to document each of the six steps applied to implement of the MLG.

**Cases**

A set of eleven standardized and sanitized cases of GBV were used in this study. Cases were provided to me in standardized and sanitized form from Dr. Stephen D. Hart's
private practice. That is, historical cases had been standardized to occur in 1997 and modern cases were standardized to have occurred in 2007 (i.e., if an index offence of a case originally occurred in 2005, all dates had been altered as if the case occurred in 2007). Cases from Canada were standardized to occur in Ontario in cities of comparable population size (i.e., if a case originally took place in Vancouver, BC it was altered to take place in Toronto, ON). Cases from international locations had been altered to occur in cities, states/provinces, or countries of comparable population size and demographics. Cases had been sanitized to exclude any names, locations or addresses, date, dates of birth, any identifying numbers (including personal security numbers), employer information, and information of those related to the individual.

The cases used in this study were actual cases of GBV to ensure legitimacy and applicability of the training for law enforcement and mental health professionals. The use of actual cases, over fictional cases, allowed participants to gain real-world practice with cases that were of sufficient quality and complexity that they would encounter in practice. As such, there were cases with limited information, conflicting information, and misleading information. However, each case was selected to contain sufficient information to code the MLG. The eleven cases reflected the diversity of GBV in type, location, perpetrator gender, purpose of the risk assessment, and conclusory opinions of level of risk (as determined by the authors of the MLG). See Table 3.4 for a summary of cases utilized.
<table>
<thead>
<tr>
<th>Case</th>
<th>Type of GBV</th>
<th>Location</th>
<th>Date</th>
<th>Perpetrator Gender</th>
<th>Purpose of Risk Assessment</th>
<th>Risk Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Small youth gang</td>
<td>Canada</td>
<td>2007</td>
<td>Male</td>
<td>Report to Crown Counsel</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>Small organized crime group</td>
<td>Canada</td>
<td>2007</td>
<td>Male</td>
<td>Sentencing</td>
<td>Low</td>
</tr>
<tr>
<td>3</td>
<td>Large organized crime group</td>
<td>Canada</td>
<td>2007</td>
<td>Male</td>
<td>Sentencing</td>
<td>High</td>
</tr>
<tr>
<td>4</td>
<td>Large gang</td>
<td>Canada</td>
<td>2007</td>
<td>Male</td>
<td>Report to Crown Counsel</td>
<td>Low</td>
</tr>
<tr>
<td>5</td>
<td>Freeman Sovereign Citizen</td>
<td>Canada</td>
<td>2007</td>
<td>Male</td>
<td>Bail Hearing</td>
<td>Moderate</td>
</tr>
<tr>
<td>6</td>
<td>Environmental extremism</td>
<td>Canada</td>
<td>2007</td>
<td>Female</td>
<td>Report to Crown Counsel</td>
<td>Moderate</td>
</tr>
<tr>
<td>7</td>
<td>Violent religious Extremism</td>
<td>Canada</td>
<td>2007</td>
<td>Male</td>
<td>Bail Hearing</td>
<td>Moderate</td>
</tr>
<tr>
<td>8</td>
<td>New religious Movement/Cult</td>
<td>USA</td>
<td>1997</td>
<td>Female</td>
<td>Court Report</td>
<td>Moderate</td>
</tr>
<tr>
<td>9</td>
<td>Violent religious extremist</td>
<td>USA</td>
<td>1997</td>
<td>Male</td>
<td>Court Report</td>
<td>High</td>
</tr>
<tr>
<td>10</td>
<td>Urban guerilla/Insurgent</td>
<td>Canada</td>
<td>1997</td>
<td>Female</td>
<td>Parole Hearing</td>
<td>Low</td>
</tr>
<tr>
<td>11</td>
<td>Nationalist Extremism</td>
<td>United Kingdom</td>
<td>1997</td>
<td>Male</td>
<td>Parole Hearing</td>
<td>Low</td>
</tr>
</tbody>
</table>

Note: Case 11 was completed by all participants as a learning case on the first day of the course and was excluded from analyses. Crown Counsel is the official name for criminal prosecutors in Canada. Risk level as judged by the authors of the MLG.

### 3.2.3. Procedure

To maintain anonymity, participants in both courses were provided a unique participant identification code used on all of their training materials. The pre-training questionnaire was administered to each participant when they arrived at the MLG training and the post-training questionnaire was administered to each participant during the final hour of
the MLG training. Participants all completed “Case 11” on the first day of the course. “Case 11” was completed individually and group feedback was provided on the application of the MLG to the case. This case was used as a learning, or practice case, for the subsequent cases on days two and three of the training and, therefore, was excluded from analysis.

Participants completed the remaining 10 cases during latter two days of the course. Participants were given 1.5 hours to review the case materials and provide ratings on the risk factors, recommended management strategies, and conclusory opinions. Two to six participants completed the same case during the morning or afternoon of the latter two days. Participants were then given 1 hour to discuss the case in their small group to develop group (or consensus) ratings and discuss the application of the MLG to the case. Group ratings were marked clearly in a different color marker than original individual item ratings. All small groups were provided expert feedback by Dr. Stephen D. Hart or myself in the small group format. Participants provided verbal feedback on the utility of the MLG throughout the training course. All participants consented to allow the verbal feedback provided during the course to be used anonymously in the evaluation of the guidelines.

Participants were randomly assigned the practice cases they completed individually and reviewed in small groups. Each case was completed both on the second day of the training and the third day of the training by multiple evaluators to control for practice effects. For example, “Case 2” was completed by half of attendees on day two and half of attendees on day three in each of the training courses.

Importantly, this course needed to balance training and practical applications of the MLG with the goals of the research. In situations that training/practical applications would conflict with an ideal research design, the training participants were provided instructions to inform use and practice. This occurred on two occasions that affected the study methodology. One, participants in the first training expressed that they would have benefited from receiving a briefing on the political climate related to “Case 11”, which

13 If a participant needed to leave prior to the end of the MLG course they were provided the post-training questionnaire prior to leaving.
involved the Palestine-Israeli conflict from the 1970s-1990s. We provided a case summary and brief political history relevant to the case to the second training group. This case was not used in the quantitative analyses and did not impact results, but did improve the training course.

Two, participants in the first training expressed that they would have benefited from reviewing a second case as a large group on the last day of the workshop. To accommodate this, we revised our workshop model for the second course to review a second case of GBV as a large group on the last day. This improved the quality of training to participants and did not affect quantitative analyses because the second case review was done after all cases were complete. As in the first workshop, participants completed up to five cases over the three-day course.

### 3.2.4. Statistical Analysis

The utility of the MLG and MLG training was assessed using both quantitative and qualitative techniques. Quantitative techniques included paired t tests to compare the participant’s pre- and post-training questionnaires on confidence, knowledge, and competence on various aspects of the MLG and course training. A-priori power analyses were calculated using G Power. The sample was sufficient for detecting both large (.80, one-tail, \( \alpha \) err prob = 0.5) and medium effect sizes (.50, one-tail, \( \alpha \) err prob = 0.5).

Qualitative content analysis was used to assess user feedback provided verbally during the MLG training and in writing through the anonymous training questionnaires. Content analysis is widely used in health research and can provide a useful and flexible method for analyzing text data (Hsieh & Shannon, 2005). Content analysis is a research method by which text data (including text collected in focus groups, through narrative responses, and open ended survey questions) is systematically classified through coding to identify common themes or patterns (Berg, 2004; Hsieh & Shannon, 2005). According to Hsieh and Shannon (2005), there are three approaches to qualitative content analysis: (1) conventional, (2) directed, and (3) summative. The three approaches differ in the coding schemes utilized (codes/keywords identified before and during analysis or only during analysis), how codes originate (text data, theory or relevant research, counting and
comparisons of key words and the interpretation of the context underlying the keywords, respectively), and the “threats to trustworthiness”, or credibility and internal validity of the themes developed.

The current study utilized conventional content analysis. Conventional content analysis is applied to studies whose aim is to describe text data when existing theory or research is either limited or non-existent. Conventional content analysis was chosen for the current study because of the flexibility of the approach to be applied to any text data (including text collected in this study through focus groups, through narrative responses, and open ended survey questions) and because no categories were defined prior to collection of participant’s feedback (Hsieh & Shannon, 2005).

The reliability of the MLG was assessed using distributions, intraclass coefficients (ICC), and structural reliability of the guidelines. The procedure allowed me to uniquely assess the interrater reliability of ratings by independent groups of assessors. Four independent groups developed consensus ratings on each case after completing their individual ratings. We can reasonably assume the groups were independent (all completing the cases on different days or in different trainings) and randomly comprised of course participants (each group was randomly selected and group members varied group to group, that is, no two groups were alike). This type of analysis is integral to the validation of the MLG, as the guidelines were developed for completion by groups of individuals with varying levels of specialization (i.e., individual-level expert, group-level expert), not independently.

Intraclass Coefficients (ICCs) were all calculated using a one-way random effects model, absolute agreement method. ICCs apply a procedure for estimating the reliability of sets of ratings. The application of ICCs has been well established in social sciences for over 40 years (Bartko, 1966) and has been used for the evaluation of item, numerical, and conclusory opinions of various SPJ tools (e.g., Douglas, Yeomans, & Boer, 2005). ICC models measure chance corrected agreements of raters through the comparison of rater variability on items to the total variability of all raters on all items (i.e., the ratio of between-groups variance to total variance; Barko, 1996; Shrout & Fleiss, 1979). A one-way random effects model is used to remove between-rater variance in the ICC
analyses when each case is completed by a different set of raters and the raters of cases are randomly selected from a larger pool of raters (Nichols, 1998; Shrout & Fleiss, 1979). It is appropriate for use over a two-way model when all raters do not complete every case and there is an assumption of no interaction or systematic variability among raters (Bartko, 1996; Nichols, 1998).

### 3.3. Results

#### 3.3.1. Utility

*Confidence, Knowledge, Competence*

There was a significant difference between participants’ pre- and post-self rating of confidence, knowledge, and competence for both risk assessment of violence generally and risk assessment of GBV (see Table 3.5). On average, participant’s self-ratings of confidence, knowledge, and competence of violence risk assessment generally significantly increased ($p < .001$), with medium effect sizes ($d = -0.42$ to $-0.52$). On average, participant’s self-ratings of confidence, knowledge, and competence of GBV risk assessment also significantly increased ($p < .001$), with large effect sizes ($d = -1.54$ to $-1.88$).
Table 3.5. Mean pre- and post-training self-ratings for general and GBV, N = 46

<table>
<thead>
<tr>
<th></th>
<th>Pre-training M</th>
<th>SD</th>
<th>Post-training M</th>
<th>SD</th>
<th>MD</th>
<th>t</th>
<th>Cohen’s D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Violence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How confident are you in your ability to assess an individual’s risk for violence?</td>
<td>7.07</td>
<td>2.43</td>
<td>8.09</td>
<td>1.59</td>
<td>1.02**</td>
<td>4.77</td>
<td>0.50</td>
</tr>
<tr>
<td>How much knowledge do you feel you have about risk assessment for violence?</td>
<td>7.09</td>
<td>2.57</td>
<td>8.15</td>
<td>1.40</td>
<td>1.07**</td>
<td>3.92</td>
<td>0.51</td>
</tr>
<tr>
<td>How much knowledge do you feel you have about risk factors for violence?</td>
<td>7.52</td>
<td>2.16</td>
<td>8.28</td>
<td>1.39</td>
<td>.76**</td>
<td>3.73</td>
<td>0.42</td>
</tr>
<tr>
<td>How competent do you feel you are at assessing an individual’s risk for violence given you current practical and technical skills?</td>
<td>7.22</td>
<td>2.27</td>
<td>8.17</td>
<td>1.24</td>
<td>.96**</td>
<td>4.31</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>Group-Based Violence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How confident are you in your ability to assess an individual’s risk for group-based violence?</td>
<td>5.09</td>
<td>1.24</td>
<td>7.57</td>
<td>1.39</td>
<td>2.48**</td>
<td>9.69</td>
<td>1.88</td>
</tr>
<tr>
<td>How much knowledge do you feel you have about risk assessment for group-based violence?</td>
<td>4.65</td>
<td>2.08</td>
<td>7.65</td>
<td>1.29</td>
<td>3.00**</td>
<td>10.35</td>
<td>1.73</td>
</tr>
<tr>
<td>How much knowledge do you feel you have about risk factors for group-based violence?</td>
<td>4.93</td>
<td>1.97</td>
<td>7.76</td>
<td>1.12</td>
<td>2.83**</td>
<td>11.00</td>
<td>1.77</td>
</tr>
<tr>
<td>How competent do you feel you are at assessing an individual’s risk for group-based violence given you current practical and technical skills?</td>
<td>5.07</td>
<td>2.08</td>
<td>7.67</td>
<td>1.18</td>
<td>2.70**</td>
<td>10.30</td>
<td>1.54</td>
</tr>
</tbody>
</table>

Note: MD = mean difference; ** p < .01; Cohen’s D effect size interpreted 0.2 is small, 0.5 is medium, 0.8 is large.

At the end of the course, participants were asked how practical the training course was for their purposes and the ease of use and understanding of the MLG. Self-ratings were scored on a 10-point likert scale. Participants indicated that the training course was practical for their purposes (M = 8.07, SD = 1.95 with 1 being “not practical” and 10 being “extremely practical”). The results also indicated that on average the MLG was easy to use (M = 7.78, SD = 1.35 with 1 being “difficult” and 10 being “easy”) and easy to understand (M = 8.00, SD = 1.13 with 1 being “difficult” and 10 being “easy”).
When asked to rank the type of GBV participants anticipate using the MLG with, a large proportion of participants anticipated using the MLG with Organized Crime (72%), Gang (70%), and youth gang (46%) cases, with 85%, 97%, 91%, respectively, of prospective users indicating these types of violence a top three rank for use in their work. A large number of participants (52%) also indicated anticipated use of the MLG on cases of freeman sovereign citizens, with 79% of anticipated users ranking this type of violence in their top three for use in their work. Another large proportion of participants (48%) indicated they would use the MLG on religious extremism cases, with 73% of prospective users indicating this as a top three rank for use in their work. A smaller, but sizable portion of users (39%) also indicated they would use the MLG on environmental extremism cases, with 61% indicating this type of violence as a top three rank for use in their work. See Table 3.6 for proportions of participants indicating intended use of the MLG with various forms of GBV.

Table 3.6. Participants indicating intended use of the MLG with various forms of GBV, N = 46

<table>
<thead>
<tr>
<th>Form of GBV</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organized crime</td>
<td>33</td>
<td>72%</td>
</tr>
<tr>
<td>Gangs</td>
<td>32</td>
<td>70%</td>
</tr>
<tr>
<td>Freeman sovereign citizens</td>
<td>24</td>
<td>52%</td>
</tr>
<tr>
<td>Terrorism (religious extremists)</td>
<td>22</td>
<td>48%</td>
</tr>
<tr>
<td>Youth gangs</td>
<td>21</td>
<td>46%</td>
</tr>
<tr>
<td>Terrorism (environmental extremists)</td>
<td>18</td>
<td>39%</td>
</tr>
<tr>
<td>Honor based violence</td>
<td>16</td>
<td>35%</td>
</tr>
<tr>
<td>Ethnic groups</td>
<td>15</td>
<td>33%</td>
</tr>
<tr>
<td>Cults/new religious movements</td>
<td>12</td>
<td>26%</td>
</tr>
<tr>
<td>Hooligans</td>
<td>8</td>
<td>17%</td>
</tr>
<tr>
<td>Guerilla/insurgency groups</td>
<td>7</td>
<td>15%</td>
</tr>
<tr>
<td>Bandits</td>
<td>5</td>
<td>11%</td>
</tr>
<tr>
<td>Sports teams</td>
<td>5</td>
<td>11%</td>
</tr>
<tr>
<td>Fraternities</td>
<td>4</td>
<td>9%</td>
</tr>
</tbody>
</table>

Note: Not all course participants assess GBV in their professional practice, which may restrict the intended use of the guidelines among participants who regularly assess GBV in practice.

Participants were also asked how often they are required to assess GBV in their professional activities and the likelihood of use of the MLG in their professional activities.
Self-ratings were scored on a 10-point Likert scale. The results indicated that participants were required to assess GBV in the professional activities ($M = 6.13$, $SD = 2.67$ with 1 being “not at all often” and 10 being “very often”) and were likely to use the MLG in their professional activities ($M = 6.95$, $SD = 3.19$ with 1 being “not at all likely” and 10 being “very likely”). There was significant variation in the frequency participants are required to assess GBV in their practice (range 1-10) and the likelihood of use in their professional practice (range 1-10).

**Content Analysis of Feedback**

Both verbal and written feedback provided during the course was analyzed and synthesized using content analysis. Content analysis was used to examine themes provided in the feedback. A total of 129 general comments ($n = 22$), verbatim quotations from the course-end feedback session ($n = 23$), and written comments by participants ($n = 84$) were reviewed. Although general comments and verbatim quotes are less frequent in the analysis, participants often agreed with comments made verbally to the entire course. This feedback will be considered with equal weight to the written feedback provided by participants for revisions of the MLG. Frequencies of the themes and subthemes of the content analysis will not be reported because agreed upon comments made verbally to the entire course would not be accurately represented and thus frequencies would be misleading for the strength of the themes.

Prior to coding, a superordinate set of categories was created that captured the range of data. When the initial set of themes was too broad, a second set of themes was created to capture the data in more detail. The set of themes increased from 4 to 15. After the themes were finalized, the text data was coded for the relevant themes.

Content analysis revealed two levels of themes for the utility feedback (see Figure 3.1). There were four primary themes: (1) process of applying the MLG, (2) manual content, (3) training, and (4) global appraisals. Feedback related to the process of applying the MLG included three subthemes: need for explicit instruction (e.g., “Extra emphasis/guidance on selecting group to assess”), communication findings (e.g., “How do you recommend presenting the assessment (report/court)?”), and practical
applications (“As a psychologist, where might I get the intelligence to rate group characteristics?”).

**Figure 3.1. Content Analysis Results**

Feedback related to the manual content included three subthemes: additional instructions (e.g., “would like to see a definitive definition of “violence”, i.e.: is property damage such as arson that has the potential to harm people considered violence even if no harm resulted”), items (e.g., “Differentiate more clearly between IG1 and IG3”), and typographical (e.g., “p. 32- last para, end of 2nd line – ‘rate’ should be ‘rated’”).

Feedback related to training included three subthemes: format (“It would have been helpful for the last case study to be done as an entire group and talked about by Dr. Hart or Alana.”), enhanced didactics (e.g., “One suggestion is to maybe consider spending more time on the management area to give an understanding of what’s acceptable.”), and cases (e.g., “Many of the scenario examples include individuals who are in custody; however from an intelligence perspective we are asked to assess risk prior to the violent act. Perhaps a couple based on intelligence information prior to the anticipated violence.”).

Feedback related to the global appraisals include two subthemes: expressions of benefit (e.g., “This has opened a new dimension in risk assessment and established a way to organize and understand and articulate factors relevant to group based violence.”) and statements for future applications (e.g., “[we are] going to be more reliant on experts in the field; we will have to work with a team.”).
3.3.2. Distributions, Interrater Reliability, and Structural Reliability

I examined the distribution, interrater reliability, and structural reliability of the MLG group ratings \( N = 40 \). The individual rating results are presented in the Appendix. Consistent with Douglas and colleagues (2005), ICC\(_1\), or reliability of a single rater, was the primary index of reliability; ICC\(_2\), for averaged ratings, are also presented. ICCs were interpreted using guidelines for strength of agreement developed by Chichetti and Sparrow (1981) for specific items\(^{14,15}\) and Landis and Koch (1977) for categorical data.\(^{16}\) Values of greater than .50 or .60 are generally considered to be good for practical purposes (Chichetti & Sparrow, 1981; see also Douglas & Ogloff, 2003; Kropp & Hart, 2000).

The distributions and interrater reliabilities of group ratings of the individual MLG risk factors are presented in Table 3.7. The results indicate that each item spanned the potential range of scores on the MLG, with the exception of I1. No items were omitted due to missing information in the consensus ratings. The interrater reliabilities of individual risk items ranged from, poor to excellent (Chichetti & Sparrow, 1981) or slight to substantial (Landis & Koch, 1977; ICC\(_1\) = .15 to 1.00). ICC values for three items (IG1, IG3, and G1) were not calculated due to insufficient variability in the ratings.

\(^{14}\) Individual items and conclusory opinions are considered single items (Chichetti & Sparrow, 1981; see also Douglas et al., 2005).

\(^{15}\) Chichetti and Sparrow (1981): \( < .40 \) is poor, \( .40 \) to \( .59 \) is fair, \( .60 \) to \( .74 \) is good, and \( \geq .75 \) is excellent.

\(^{16}\) Landis and Koch (1977): 0 is poor, \( .01 \) to \( .20 \) is slight, \( .21 \) to \( .40 \) is fair, \( .41 \) to \( .60 \) is moderate, \( .61 \) to \( .80 \) is substantial, and \( .81 \) to 1 is almost perfect.
### Table 3.7. Presence of the Multi-level Guidelines (MLG) factors: Distribution and Interrater Reliability of Ratings

<table>
<thead>
<tr>
<th>Individual Domain factors</th>
<th>Distribution</th>
<th>Interrater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>P</td>
</tr>
<tr>
<td>I1</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>I2</td>
<td>20%</td>
<td>12.5%</td>
</tr>
<tr>
<td>I3</td>
<td>15%</td>
<td>5.0%</td>
</tr>
<tr>
<td>I4</td>
<td>43%</td>
<td>25.0%</td>
</tr>
<tr>
<td>I5</td>
<td>20%</td>
<td>20.0%</td>
</tr>
<tr>
<td>I6</td>
<td>18%</td>
<td>30.0%</td>
</tr>
<tr>
<td>Individual-Group domain factors</td>
<td>Distribution</td>
<td>Interrater reliability</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>P</td>
</tr>
<tr>
<td>IG1</td>
<td>8%</td>
<td>13%</td>
</tr>
<tr>
<td>IG2</td>
<td>15%</td>
<td>15%</td>
</tr>
<tr>
<td>IG3</td>
<td>3%</td>
<td>5%</td>
</tr>
<tr>
<td>IG4</td>
<td>8%</td>
<td>26%</td>
</tr>
<tr>
<td>Group domain factors</td>
<td>Distribution</td>
<td>Interrater reliability</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>P</td>
</tr>
<tr>
<td>G1</td>
<td>3%</td>
<td>8%</td>
</tr>
<tr>
<td>G2</td>
<td>20%</td>
<td>15%</td>
</tr>
<tr>
<td>G3</td>
<td>13%</td>
<td>8%</td>
</tr>
<tr>
<td>G4</td>
<td>8%</td>
<td>8%</td>
</tr>
<tr>
<td>G5</td>
<td>15%</td>
<td>25%</td>
</tr>
<tr>
<td>G6</td>
<td>20%</td>
<td>30%</td>
</tr>
<tr>
<td>Group-Society domain factors</td>
<td>Distribution</td>
<td>Interrater reliability</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>P</td>
</tr>
<tr>
<td>GS1</td>
<td>18%</td>
<td>18%</td>
</tr>
<tr>
<td>GS2</td>
<td>15%</td>
<td>5%</td>
</tr>
<tr>
<td>GS3</td>
<td>15%</td>
<td>15%</td>
</tr>
<tr>
<td>GS4</td>
<td>65%</td>
<td>23%</td>
</tr>
</tbody>
</table>

**Note:** For distribution (N = 40): N = no; P = partially or possibly present; Y = present. For interrater reliability (N = 40): ICC₁ = intraclass correlation coefficient for single ratings and ICC₂ = intraclass correlation coefficient for averaged ratings, both calculated using a mixed effects model, absolute agreement method; - = insufficient variability to calculate interrater reliability.

Distributions, median-corrected item-total correlations (Mdn CITCs), and ICCs of the numerical scores for the Individual, Individual-Group, Group, and Group-Society factors for the group ratings are presented in Table 3.8. Distributions of the ratings were similar across the domains and CITCs were generally poor to fair (Chichetti & Sparrow, 1981)
or fair to moderate (Landis & Koch, 1977). The ICCs for the numerical scores ranged from fair to excellent (Chichetti & Sparrow, 1981) or moderate to almost perfect (Landis & Koch, 1977; .47 to .83).

Table 3.8. Presence of the Multi-level Guidelines (MLG) factors: Distribution, Structural Reliability, and Interrater reliability of Numerical scores

<table>
<thead>
<tr>
<th>Factors</th>
<th>Distribution</th>
<th>Structural reliability</th>
<th>Interrater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>Mdn CITC</td>
<td>ICC₁</td>
</tr>
<tr>
<td>I factors</td>
<td>8.58 (2.55)</td>
<td>0.38</td>
<td>.83</td>
</tr>
<tr>
<td>I-G factors</td>
<td>6.78 (1.83)</td>
<td>0.59</td>
<td>.58</td>
</tr>
<tr>
<td>G factors</td>
<td>9.53 (2.87)</td>
<td>0.54</td>
<td>.47</td>
</tr>
<tr>
<td>G-S factors</td>
<td>5.15 (2.18)</td>
<td>0.53</td>
<td>.60</td>
</tr>
<tr>
<td>Total</td>
<td>30.03 (8.01)</td>
<td>0.59</td>
<td>.58</td>
</tr>
</tbody>
</table>

Note: N = 40; I = Individual; I-G = Individual-Group; G = Group; G-S = Group-Societal. For distributions: maximum = 12 for I, and G; 8 for I-G and G-S; and 40 for Total. For structural reliability: Mdn CITC, median-corrected item-total correlation. For interrater reliability: ICC₁, intraclass correlation coefficient for single ratings and ICC₂, intraclass correlation coefficient for averaged ratings, both calculated using a mixed effects model, absolute agreement method.

The distributions and ICCs for the conclusory opinion ratings are presented in Table 3.9. The ratings are well distributed with a relatively greater amount of cases in the sample rated as high for case priority, risk of violence, and risk of life-threatening violence. The ICCs for the conclusory opinions were all good or excellent (Chichetti & Sparrow, 1981) or substantial (Landis & Koch, 1997).

Table 3.9. Multi-level Guidelines (MLG) Conclusory Opinions: Distribution and Interrater Reliability of Ratings

<table>
<thead>
<tr>
<th>Presence</th>
<th>L</th>
<th>M</th>
<th>H</th>
<th>Omit</th>
<th>ICC₁</th>
<th>ICC₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case Prioritization</td>
<td>22%</td>
<td>23%</td>
<td>56%</td>
<td>0%</td>
<td>.76</td>
<td>.93</td>
</tr>
<tr>
<td>Risk of Violence</td>
<td>23%</td>
<td>33%</td>
<td>45%</td>
<td>0%</td>
<td>.69</td>
<td>.90</td>
</tr>
<tr>
<td>Risk of life-threatening violence</td>
<td>23%</td>
<td>38%</td>
<td>40%</td>
<td>0%</td>
<td>.73</td>
<td>.92</td>
</tr>
<tr>
<td>Risk of Imminent Violence</td>
<td>38%</td>
<td>33%</td>
<td>30%</td>
<td>0%</td>
<td>.67</td>
<td>.89</td>
</tr>
</tbody>
</table>

Note: For distribution (N = 40): L = low; M = moderate; H = high. For interrater reliability (N = 40): ICC₁ = intraclass correlation coefficient for single ratings and ICC₂ = intraclass correlation coefficient for averaged ratings, both calculated using a mixed effects model, absolute agreement method.
3.4. Discussion

The utility and reliability of the MLG was evaluated through two MLG training courses for criminal justice and mental health professionals. This was the first evaluation of the MLG. The sample was a relatively large and diverse sample of primarily front-line law enforcement officers. The results supported my three hypotheses: (1) criminal justice and mental health professionals confidence, competency, and knowledge for assessing and managing GBV significantly increased after learning how to administer the MLG, (2) criminal justice and mental health professionals found the MLG to be a useful set of guidelines for practice and provided integral feedback for the development of the tool, and (3) the reliability of the MLG is comparable to those found with other SPJ tools. Limitations and implications of the findings are discussed throughout this section.

3.4.1. Utility

Quantitative

The results indicated that participants perceived a significant improvement in their confidence, knowledge, and competence in assessing and managing risk for GBV after learning to administer the MLG. The results also indicated that participants perceived a significant improvement in their confidence, knowledge, and competence in assessing and managing risk for violence more generally, though this was to a lesser degree than improvement in the area of GBV.

Although self-reported confidence, knowledge, and competence increased, there is no way to determine if this reflects an actual improvement in skill. Storey and colleagues (2011) found similar findings of improved confidence, knowledge, and competence via self-report. However, in their study, self-report ratings did not correlate with improvement of participant's scores on a multiple-choice competency test. Storey and colleagues concluded that self-reported confidence might exceed objective skill level learned in violence risk assessment courses.

Although this may be the case, it is also possible that the multiple choice competency test administered in Storey and colleagues’ (2011) study was an arbitrary measure of
actual skill. The test was developed for the purpose of evaluating the training course and did not undergo any psychometric evaluation (e.g., no norms were developed, the authors did not evaluate the sensitivity and specificity of the items). Without this information we cannot conclude that the test developed was an objective and accurate measure of actual skill. That being said, it is important to recognize that self-reported improvement may be an over estimate of participant’s confidence, knowledge, and competence in this study. This could have important implications for practice. For example, if trained users of the MLG falsely believed they were competent and knowledgeable in this area there may be errors in their work. Given the nature of the decisions made based on threat assessments, these errors may literally mean life or death. It is important for users to engage in ongoing training and collaboration and consultation with their peers and supervisors to maintain competency. Further, supervisors and managers should consistently provide adequate feedback on the competency and knowledge of users under their supervision and secure additional training or support as needed.

The results also indicated that the course was practical for the purposes of the participants, and that on average the MLG was easy to use and easy to understand. The participants also indicated a variety of cases with which they anticipate using the MLG, and presumably types of violence they see the MLG as useful to assess in their own practice. The most common forms of GBV listed were organized crime, gangs, freeman sovereign citizens, terrorism (both religious and environmental extremists), and youth gangs. Participants also indicated use of the MLG on cases of honour-based violence, with ethnic groups, cults/new religious movements, hooligans, insurgency/insurgency groups, bandits, and with traditionally non-violent groups (i.e., fraternities and sports teams). This list covered the range of forms of GBV in the literature and supports the hypothesis that risk factors for GBV are useful across various forms of groups. The results provided evidence that the model captured by the MLG has a general application to a variety of groups.

Participant responses also indicated that they are required to assess GBV in their professional activities and were likely to use the MLG, but there was variation in these responses. This may reflect the diversity of the sample, with some participants working
as members of GBV threat assessment units (e.g., organized crime and gang or terrorist units) and others working primarily in areas where assessment of GBV are less frequent, such as in forensic psychiatric hospitals or in research. This may also reflect a lack of intention to use the MLG in practice. This explanation, however, is less likely given the high ratings of the practicality of the course and the ease of use of the MLG. Notably, the diversity of the course participants may have blunted the rates of intended use of the guidelines of those who regularly assess GBV in practice. Therefore, the proportions of participants indicating use with various forms of GBV should be considered a conservative estimate.

Qualitative Content Analysis. A significant aspect of the development and evaluation of the MLG was the active dialogue Dr. Hart and I strived for with participants to modify and improve the guidelines to reflect both the empirical literature and current practice. The results indicated several themes for the feedback on the MLG and MLG Course: (1) process of applying the MLG (i.e., need for explicit instruction, communication of findings, and practical applications), (2) manual content (i.e., additional instructions, items, and typographical), (3) training (format, enhanced didactics, and cases), and (4) global appraisals (i.e., expressions of benefit and statements for future applications). Throughout each theme the request for further instruction or clarification of the MLG was noted. Overall, the participants provided valuable feedback for the development of the tool across a diverse set of themes.

While this feedback is useful, one potential limitation of the methodology used (i.e., conventional content analysis) is the possibility that the data might not be accurately represented because codes originate solely from the data and discretion of the coder (Hsieh & Shannon, 2005). This is primarily a concern when the coder does not adequately understand the context of the data and is therefore unable to identify the key categories. In this study, I was intimately familiar with the context of the data: I conducted the literature review and reviewed all expert feedback to develop the MLG, I am the first author of the MLG and therefore contributed significantly to the content of the manual, I developed and delivered the MLG training course with Dr. Hart and I was present for the entirety of the two training courses. As a result, concerns of credibility and internal validity of the themes developed are minimal or non-existent. Further,
content analysis is limited to providing a description of the feedback. The purpose of content analysis in the current project is to describe the qualitative findings. No specific revisions of the MLG on the basis of these results will be outlined here. However, these findings, along with the other findings of this project will inform the revisions of the MLG.

Another limitation of the methodology used for the qualitative data was the lack of a method for collecting feedback in the group format. There are methodologies specifically for collecting such data (i.e., focus group methodologies; Berg, 2004); however, a focus group method was not used in the current study. Further research obtaining qualitative feedback from violence risk training courses should be developed with focus group methodology prior to data collection.

3.4.2. Reliability

The results provide support for the reliability of the MLG. Consistent with my hypothesis, the reliability estimates for the MLG are comparable to those found with other SPJ tools.

*Distributions.* The distributions of the individual MLG risk factors (with the exception if I1, Individual Violent Behaviour), total scores of the four domains (i.e., Individual, Individual-Group, Group, and Group-Society), and the conclusory opinions for case priority, risk for violence, risk for life-threatening violence, and risk for imminent violence were all well distributed across the possible ratings. This adequate variability indicates that users can communicate low, moderate, and high risk within each of the domains and for conclusory judgments. Revisions of the MLG should consider the specificity and importance of the items with ratings favouring absence and presence (e.g., I1, IG3, GS4). However, it is important to keep in mind that the cases used in this study were selected for the purposes of training from a restricted sample. While the results do demonstrate variability in the potential range of ratings in the small set of cases included; the results do not represent the distribution of items across GBV cases generally. Future research is needed to determine the distribution of items in a large sample of GBV cases.

*Interrater Reliability.* The interrater reliability results were generally consistent with those found in evaluations of other SPJ tools. At the individual item-level the reliability
coefficients ranged from poor to excellent, or slight to substantial. Generally, SPJ tools evidence variability at the item-level (e.g., HCR-20 reliability coefficients ranged from poor to excellent; Douglas & Ogloff, 2003). The reliability coefficients for the domain and total numerical scores were fair to excellent, or moderate to almost perfect. These results are generally consistent with past research of other SPJ tools (e.g., Douglas & Ogloff, 2003; Kropp & Hart, 2000), particularly for the Individual factors. The reliability coefficient of the Group factors was lower than the domain and total numerical scores of other SPJ tools, but was consistent with studies that used a similar methodology to the current study (Sutherland et al., 2012; Wilson, 2013). The reliability coefficients of the conclusory opinions were all good to excellent or substantial. This result was comparable to those found in evaluations of other SPJ tools, and slightly better than reliability coefficients of the HCR-20, RSVP, and SAM in the limited studies reviewed in this paper (i.e., Douglas & Ogloff, 2003; Kropp et al., 2011; Sutherland et al., 2012; Wilson, 2013). The results demonstrate that conclusory opinions made using the MLG can be made reliably by groups of raters on a diverse set of cases.

The results also demonstrate some support for the structural reliability of the MLG. The correlations between each factor and domain total scores should be modest, that is, correlations should demonstrate that factors are contributing unique information to the total scores of the domains. The reliability of the Individual, Individual-Group, Group, and Group-Societal factors was fair to moderate at best; demonstrating some support for the unique contribution of the factors in each of the domains.

The study used professionals with varying degrees of experience. Like Sutherland and colleagues (2012) and Wilson (2013) suggest: The strength of this approach is increased ecological validity of the study and a clearer picture of reliability in the field; however, the methodology used may explain some of the lower reliability estimates evidenced in diverse, multi-rater, training studies. Further, although the methodology of the current study attempted to control for practice effects that may have confounded the results (i.e., by randomly assigning cases to be completed across the later two days in varying order), the results are still a lower bound estimate because I am testing reliability of raters in training. It will be important for future research to utilize a methodology.
consistent with other studies of violence risk that have two or three trained raters coding a larger number of cases.

### 3.4.3. Implications

The findings of the utility and reliability analyses demonstrate support for the application of the MLG to the assessment and management of GBV. The cases utilized in the evaluation were varied in type of GBV (i.e., youth gang, organized crime, freeman sovereign citizen, environmental extremism, violent religious extremism, new religious movement/cult, lone actor violent religious extremism, urban insurgent/insurgent, and nationalist extremism), varied in location (i.e., in Canada and International), perpetrator gender (i.e., male and female), purpose of the risk assessment (i.e., report to crown counsel, sentencing, bail hearing, court report, and parole hearing), and objective level of risk (i.e., low, moderate, and high). Taken together, the results support the use of the MLG across these varying types of groups, both within Canada and internationally, and with females as well as males, for a variety of purposes, across all levels of risk for violence.

In addition to the suggested research noted above, there are three applications of the MLG that warrant attention in the coming years: lone actor risk for GBV, female perpetrators risk for GBV, and youth perpetrators risk for GBV. First, lone actors risk for violence has been a contentious issue for many years (e.g., Corrado, 1982). Lone actors within the context of GBV will be defined as lone actors of violence whose affiliation or connection with a group is non-existent or unimportant to their risk for violence. This can occur in a variety of different ways. One, their affiliation or connection may primarily be a result of psychotic processes. That is, the individual holds a delusional belief that he or she is a member of a group; he or she may use their delusional affiliation with group to motivate or justify their violent acts. Two, their affiliation or connection may also be unimportant to their risk for violence, regardless of the etiology of the group connection. That is, although an individual may identify with larger groups that on the surface seem to be important in understanding and managing their risk, upon evaluation the group plays little to no role in their decision to engage in violence. There was one case of a lone actor in the current evaluation. Based on individual and small group discussion it
became apparent that there was evidence of the factors in all four domains of the MLG, but the relevance of the Individual-Group, Group, and Group-Societal factors were rated as “No”, or not relevant, to the individual’s risk for violence. However, factors within the Individual domain were identified as relevant to the individual’s risk for future violence, which, in turn, informed case formulation and management strategies (e.g., I5 Mental Health Problems). It would be interesting to apply the MLG to prominent cases such as perpetrator Anders Behring Breivik and suspect Nidal Malik Hasan17 to flesh out the application to lone actors further. In both Breivik’s and Hasan’s cases the uncertainty of group affiliation has raised questions regarding law and policy (i.e., criminal responsibility, terrorism laws, policies of the American Military) as well as risk for future violence.

The application of the MLG to female perpetrators of GBV is another area of special interest for future evaluation. The current project provides evidence for the application of the MLG to females. The methodology included females in the Campbell Collaboration review and three of the 10 cases used in the evaluation of the tool were female offenders. There was no evidence from the current study that suggested the MLG would not be applicable to female offenders. However, there may be unique characteristics of female terrorism offenders that should be considered when applying the MLG in such cases. Jacques and Taylor (2013) conducted a bibliographic study of male (n = 269) and female (n = 222) terrorist offenders across a range of terrorist activities (i.e., offences related to nationalist-separatist, social revolutionary, and religious fundamentalist issues). The results indicated that, in general, males and females were similar on several characteristics (i.e., similar age of recruitment or activity in terrorist group, immigrant status, and role played in terrorism) and different on other (i.e., females had a higher level of education at time of recruitment/involvement in terrorist group, higher rates of unemployment at time of recruitment/involvement in terrorist group, and lower rates of previous activist connections). The results also indicated that female offenders were a heterogeneous group, indicating the need for a flexible approach to

17 Anders Behring Breivik was convicted for the mass murder of 70 in Oslo and Utøya, Norway in 2011. Nidal Malik Hasan is the suspect for the mass murder in of 13 individuals at the Fort Hood Military Base in Texas, U.S in 2009.
assessing and managing these cases. Jacques and Taylor only analyzed static (not dynamic) characteristics of offenders.

A third area of special interest is the application of the MLG to youth. Although one case included in the current study was a youth gang case, it will be important to continue investigating the MLG in the assessment and management of youth GBV. The MLG may be applied individuals age 14 and older (Cook et al., 2013). However, youth can differ substantial from adults in criminal offending. Although social processes apply across age groups, there may be unique considerations when applying the MLG to youth. One reviewer suggested that evaluators be cognizant of the high rates of gang involvement and peer-influenced crime in youth that may not translate to the youth’s adult behaviour. The reviewer was referring to the distinction between adolescent limited and life-course persistent offending in youth (Moffit, 1993) and the distinctive roles groups play in adolescent offending (Sherif & Sherif, 1964; Short & Strodbeck, 1965).

Adolescent limited offenders are youth who begin to engage in criminal activity in adulstescences, with only a small portion continuing to offend in adulthood (Moffit, Caspi, Harrington, & Milne, 2002). This course of violence is more frequent than the alternate course of violence, the life-time persistent group (Moffit, 1993). Monahan, Steinberg, and Cauffman (2010) investigated the role of peers on criminal behaviour in youth transitioning to adulthood in a longitudinal study of 1,354 youth. Their results indicated that after age 20 the influence of peers (i.e., the degree adolescents’ acted autonomously during interactions with their peers) on criminality significantly decreased. They suggested that the change seen in youth transitioning to adulthood may be related to increased maturity of youth as they age.

Corrado and Freedman (2011) identify antisocial peers as a key variable for models of persistent criminal behaviour into adulthood. Specifically, having antisocial peers is associated with both increased violent behaviours and the likelihood of escalation of violent behaviours for antisocial youth who are attracted to violent groups (i.e., gangs). The authors point out that there are risk factors for the smaller proportion of adolescent onset offenders who continue to engage in crime into later adulthood – including gang involvement. These risk factors include, for example, concentrated social or economic
disadvantage and family criminality. Bouchard & Spindler (2010) added that findings from their study indicate that it is not simply association with delinquent peers or membership in groups that contributes to criminal offending in youth; rather, a formal structure and organization (e.g., a group name, leaders, hierarchy, rules, specific territory, defense of honor/reputation) of youth delinquent groups acted as a key mechanism for delinquency and violence, among other outcomes in their study.

Evaluators will need to recognize the role of adolescence in communicating risk for GBV. While some youth may be at risk for GBV into adulthood, others may not. Group membership alone is not necessarily indicative of risk for GBV, even among youth. Evaluators need to assess and communicate when risk for GBV is limited to adolescence and use risk assessment as a means to inform management and intervention for youth. For examples of intervention strategies for adolescent onset offenders see Corrado and Freedman (2011, p. 215).
Chapter 4.

Conclusion

This project details the development and evaluation of a new set of guidelines for the assessment and management of GBV: the Multi-level Guidelines, or MLG (Cook et al., 2013). The draft guidelines were developed through a synthesis of a Campbell Collaboration review and expert feedback. The MLG reflects an ecological, or nested, model of GBV that consists of 20 risk factors across four domains: Individual, Individual-Group, Group, and Group-Societal. The draft guidelines were evaluated for utility and reliability through two training courses of criminal justice and mental health professionals. The results of the evaluation provided evidence that the MLG is a useful and reliable tool for the assessment and management of various forms of GBV.

This project addressed each of Monahan’s (2012) four principal conclusions about individual risk for terrorism and expanded his conclusions beyond terrorist violence to all forms of GBV.

Conclusion (1) a prerequisite of work in this area is to specify the nature and scope of risk for terrorism. I argue that the nature and scope of risk assessment for terrorism should be extended to include all GBV. The current Campbell Collaboration review and expert consensus supports this conclusion. The evidence- and consensus-based guidelines spanned heterogeneous sources of data that included social science principles applicable to all forms of GBV and reviewers that spoke to the utility of an inclusive model that is applicable to various groups. Empirically, the evaluation results presented in this paper support the nature and scope of risk assessment to include all forms of GBV. The results of the utility and reliability evaluation suggest that a comprehensive model of all forms of GBV is applicable. This is evident both in the user feedback of applications of the MLG to multiple forms of GBV and the reliability
evidenced among a diverse set of cases of various forms of GBV. The proposed conclusion is an application of group processes in different contexts, among different types of groups.

Conclusion (2) a structured professional judgment (SPJ) approach to risk assessment may be useful in assessing risk for terrorist violence, albeit the content of a SPJ tool would differ substantially from existing instruments and guidelines. Based on the review, Drs. Stephen D. Hart and P. Randall Kropp and I developed a SPJ tool for the assessment and management of group-based violence, the MLG (Cook, et al., 2013). The MLG differs substantially from existing instruments and guidelines. The tool is both evidenced- and consensus-based.

Although the MLG reflects current best practice, typically, SPJ guidelines are developed based on extensive bodies of literature that have directly examined variables related to the behaviour under examination. There is debate in the field whether there is enough information to determine what factors influence risk for GBV (Dernevik, Beck, Grann, Hougue, & McGuire, 2009; Gudjonsson, 2008). Scholars (e.g., Borum, 2011; Dernevik et al., 2009; Monahan, 2012) have stated that there is little existing research on the individual risk factors for terrorism. For example, Monahan (2012) stated, "existing research has largely failed to find valid nontrivial risk factors for terrorism" (p. 184).

Based on the current review, the scope of the existing evidence is limited; however, peer-reviewed papers that are primarily qualitative, quantitative, and reviews characterize the evidence. The evidence also includes a large number of exploratory analyses, qualitative findings, and case studies. The limited existing evidence is, in part, due to our current lack of ability to systematically evaluate GBV risk factors. The research is limited, but I disagree that the field has “failed” to date to find any risk factors of terrorism, or more broadly risk factors for GBV. Exploratory analysis, qualitative findings, and case studies are often the basis for theory and development of systematic research. This field is at an early state of development. We are in a position to translate the preliminary identification of risk factors for GBV into a systematic model, or heuristic, to stimulate research and allow for rigorous examination of this phenomenon.
Not only do we have sufficient information to begin rigorous examination of GBV, there is a pressing need for practitioners to have empirically-based guidelines that reflect the current state of the field for informing the assessment and management of GBV, despite current limitations. Gudjonsson (2008), Monahan (2012), and Roberts and Horgan (2008) all agree that there is a need for SPJ guidelines based on the current evidence for terrorist violence. I argue that the same is true for all GBV. The MLG was developed to meet this need as the new guidelines are intended for use by criminal justice, security, and mental health professionals working in a variety of contexts where there is risk or contact with those involved with GBV. These guidelines are implemented through the use of interdisciplinary teams that can assess the complexity of the multiple levels of analysis required.

Evaluation of this conclusion is dependent on future systematic empirical examination of the individual risk of GBV across different types of groups. There is sufficient evidence to suggest that group processes will apply similarly across different types of violent groups and would be important to consider in the assessment and management of risk across these groups. It may be the case that upon empirical evaluation some risk factors for GBV may be important, or more important, for some types of groups than others. Similarly, some risk factors for GBV may be unimportant, or less important, for some types of groups than others. This is true of risk factors at the individual-level and group-level; however, this is an empirical question.

Conclusion (3) the highest priority for research is to identity robust individual risk factors for terrorism. There are two ongoing goals of the current project: (1) to develop a systematic and empirically-based model of GBV with clear definitions to facilitate high quality research on risk assessment and management of GBV and (2) to develop a new comprehensive heuristic for thinking about, researching, assessing, and managing GBV risk. In line with Monahan’s (2012) third principal conclusion, my primary objective is to stimulate systematic and transparent research on GBV. These goals are also in line with Roberts and Horgan’s (2008) conclusions that there is an urgent need for the development of a risk assessment tool to assist counterterrorism professionals in the management of terrorism to prioritize the highest risk cases. There are similar conclusions from social science literature. In a recent special issue published in
Aggression and Violent Behaviour, Hollin (2010) concluded that there is a “pressing need” for examining social psychological principles that lead to violence. The MLG (Cook et al., 2013) fills this need and extends the tool to all GBV as the new guidelines provide a set of structured and systematic steps in evaluating GBV. Using the guidelines professionals are better able to make systematic, transparent, and empirically informed decisions about the level of individual risk for GBV. The nested model approach addresses the need for incorporation of an interdisciplinary perspective, particularly social psychology, criminology, and criminal analysts, in the assessment of GBV risk (Gordon, 2010; Gudjonsson, 2009; Horgan, 2011; Stevens, 2005). Although this study provides support for the utility and reliability of the MLG, further evaluation of the guidelines is needed. This research needs to be systematic and transparent (Hollin, 2010; Horgan, 2011).

Douglas & Kropp (2002) put forth a four step prevention-focused paradigm informed validation procedure for violence risk assessment and management: (1) validation of the procedure used to identify risk factors, (2) evaluate the relationship between identified risk factors and violence, (3), evaluate the relationship between the clinical risk and management decisions made on the basis of the identified risk factors and violence, and (4) evaluate impact of using the identified risk factors to inform clinical decisions and management the reduction or prevention of violence. The current study informed Step 1 in this process. The identified risk factors in the MLG are face-valid, content-valid, legally-valid, and evidence clinical and practical utility. Further investigations should include the examination of the MLG for concurrent, incremental (above general risk assessment measures), and predicative validity. Further, there is a large body of high-quality research on other SPJ guidelines that can inform the research for validation of the current guidelines. For examples of high quality studies evaluating the validity of SPJ guidelines see Nicholls, Ogloff, & Douglas (2004), Viljoen, Nicholls, Greaves, de Ruiter, and Brink (2011), and Vincent, Guy, Fusco, Gerhenson (2012). See also Douglas, Skeem, and Nicholson (2011) for information about research methods in violence risk assessment. Several important lines of discussion have followed regarding the evaluation of predicative validity of violence risk assessment tools that those validating the MLG may want to consider (see Singh, 2013).
Conclusion (4) researchers require access to retrospective information on known groups of terrorist and non-terrorist to achieve these goals. Research on GBV requires access to highly protected populations. I want to underscore the call for access to specialized groups (Monahan, 2012; Horgan, 2011). Psychologists need to work together with counterterrorism, gang, and organized crime officials to effectively evaluate these populations for informing reduction in violence and victimization. By working together, this research can have direct implications for policy in jurisdictions internationally. Policy implications can range from criminal justice, security personnel, and national security. Research can inform policy on decisions of allocation of investigative resources, custodial release, workplace safety, and threat to national security.

Lack of systematic and policy-relevant research has had implications. For example, Fasihuddin (2012) reported that front-line law enforcement officers in Pakistan have encountered significant barriers in combating terrorism because of a lack of effective national policies. The results of Fasihuddin’s study indicated that law enforcement officers faced these barriers due to a lack of policy-relevant investigation of terrorism risk, such as effective case building. The MLG may serve as a framework for systematic and transparent evaluation for GBV risk. That being said, institutional, regional, and national policy should not explicate the MLG as the only option of assisting the evaluation of GBV risk. It is important to recognize the MLG may not be applicable to all evaluators in all contexts. For example, the United Kingdom has developed the ERG 20+ (National Offender Management Service, 2011) that appears to be meeting their need in the assessment and management of extremist risk. Along the same lines, some law enforcement officers may be required by their departments to refer threat assessment cases to specialty units, as several units do in Canada (e.g., RCMP and Ontario Provincial Police) and the use of the MLG regularly may fall outside of their duties.

In sum, this project is the first to develop and evaluate a comprehensive, management-oriented set of guidelines of GBV for threat assessment specialists: the Multi-level Guidelines for the Assessment and Management of Group-Based Violence, or MLG (Cook et al., 2013). This project addressed a significant gap in current threat assessment practice with the development of the MLG. The evaluation of the tool as a
useful and reliable instrument for the assessment and management of GBV is also encouraging.

The ongoing goal of this project is to facilitate empirical research on the assessment and management for GBV risk. I would like to again emphasize the importance of sound empirical investigation into the risk factors for GBV. Although the existing evidence was sufficient to develop guidelines for the purposes of introducing a heuristic of GBV and providing a framework for research and practice in the area, these guidelines will need to reviewed and updated as new research advances the field. The findings from this study will inform the revisions of the MLG.
References


Borum, R. (2011, April) Assessing individual risk for terrorism involvement. Keynote presentation to the annual meeting of the International Association of Forensic Mental Health Services, Miami, FL.


R. v. Gaya, [2008] ONSC 24539


Appendix.

Distributions, Interrater Reliability, and Structural Reliability: Individual ratings

I also examined the distributions and interrater reliabilities of ratings of the individual MLG risk factors for group-based violence for individual ratings. ICC₁, or reliability of a single rater, was the primary index of reliability; ICC₂, for averaged ratings, are also presented. As with the group consensus data, ICCs were interpreted using guidelines for strength of agreement developed by Chichetti and Sparrow (1981) for specific items and Landis and Koch (1977) for categorical data. Values of greater than .50 or .60 are generally considered to be good for practical purposes (Chichetti & Sparrow, 1981).

Importantly, there were two additional occasions to those explicated in the body of the paper where training/practical applications conflicted with an ideal research design that should be considered when interpreting the individual rater results. One, participants were instructed to leave items blank if they were unsure or did not know how to rate the item. Those ratings were then discussed in small group formats with expert feedback. Raters did make a rating after group discussion and were considered when determining consensus rating for each group. This occurred for the individual risk factors ratings, management strategies, and conclusory opinions. While this resulted in a large number of missing items or ratings at the individual rating level (see Table A1) it did improve the quality of training provided to the participants and did not effect the group consensus ratings used in analysis for this project.

Two, participants were asked to focus on risk factors item ratings in the assessment to learn the item descriptions and ratings. As risk item ratings are the root or core of the assessment process that have substantial impacts on the scenario planning and management stages we felt it was important to instruct and weight time to risk item assessment. Most of the participants were experienced in scenario development and management strategies and were learning the unique assessment piece of the MLG specifically. This improved the quality of training to participants, but did result in a large number of cases with no scenario planning or management strategies indicated (see Table A3). The results should be viewed with this in mind. Ideally for research participants would provide ratings on all items and all aspects of administration of the MLG.

The distributions and interrater reliabilities of individual ratings of the individual MLG risk factors are presented in Table A1. The results indicate that each item spanned the

---

18 Individual items and conclusory opinions are considered single items (Chichetti & Sparrow, 1981; see also Douglas et al., 2005).
19 Chichetti and Sparrow (1981): < .40 is poor, .40 to .59 is fair, .60 to .74 is good, and > .75 is excellent.
20 Landis and Koch (1977): 0 is poor, .01 to .20 is slight, .21 to .40 is fair, .41 to .60 is moderate, .61 to .80 is substantial, and .81 to 1 is almost perfect.
potential range of scores on the MLG. A larger number of items were omitted than expected, particularly in the Group and Group-Society domains. The interrater reliabilities of individual risk items ranged from poor to excellent (Chichetti & Sparrow, 1981) or slight to substantial (Landis & Koch, 1977; ICC\(_1\) = .12 to .76).

Table A1. Presence of the Multi-level Guidelines (MLG) Factors: Distribution for Individual Raters

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Interrater Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Individual Domain factors</td>
<td></td>
</tr>
<tr>
<td>I1</td>
<td>8%</td>
</tr>
<tr>
<td>I2</td>
<td>25%</td>
</tr>
<tr>
<td>I3</td>
<td>17%</td>
</tr>
<tr>
<td>I4</td>
<td>41%</td>
</tr>
<tr>
<td>I5</td>
<td>23%</td>
</tr>
<tr>
<td>I6</td>
<td>18%</td>
</tr>
<tr>
<td>Individual-Group domain factors</td>
<td></td>
</tr>
<tr>
<td>IG1</td>
<td>6%</td>
</tr>
<tr>
<td>IG2</td>
<td>17%</td>
</tr>
<tr>
<td>IG3</td>
<td>4%</td>
</tr>
<tr>
<td>IG4</td>
<td>16%</td>
</tr>
<tr>
<td>Group domain factors</td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>1%</td>
</tr>
<tr>
<td>G2</td>
<td>23%</td>
</tr>
<tr>
<td>G3</td>
<td>11%</td>
</tr>
<tr>
<td>G4</td>
<td>5%</td>
</tr>
<tr>
<td>G5</td>
<td>18%</td>
</tr>
<tr>
<td>G6</td>
<td>23%</td>
</tr>
<tr>
<td>Group-Society domain factors</td>
<td></td>
</tr>
<tr>
<td>GS1</td>
<td>18%</td>
</tr>
<tr>
<td>GS2</td>
<td>17%</td>
</tr>
<tr>
<td>GS3</td>
<td>15%</td>
</tr>
<tr>
<td>GS4</td>
<td>55%</td>
</tr>
</tbody>
</table>

Note: For distribution (N = 172): N = no; P = partially or possibly present; Y = present. For interrater reliability (N = 172): ICC\(_1\) = intraclass correlation coefficient for single ratings and ICC\(_2\) = intraclass correlation coefficient for averaged ratings; both calculated using a mixed effects model, absolute agreement method.

Distributions, median-corrected item-total correlations (\(Mdn\) CITCs), and ICCs of the numerical scores for the Individual, Individual-Group, Group, and Group-Society factors
for the individual ratings are presented in Table A2. Distributions of the ratings were similar across the domains and CITCs were generally good. The ICCs for the numerical scores ranged from fair to good (Chichetti & Sparrow, 1981) or moderate to substantial (Landis & Koch, 1977; .47 to .65). Only the I factors and Total were considered good for practical purposes (Chichetti & Sparrow, 1981).

### Table A2. Presence of the Multi-level Guidelines (MLG) Factors: Distributions and Structural Reliability of Numerical Scores for Individual Raters

<table>
<thead>
<tr>
<th></th>
<th>Distribution</th>
<th>Structural reliability</th>
<th>Interrater Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M (SD)</td>
<td>N</td>
</tr>
<tr>
<td>I factors</td>
<td>164</td>
<td>8.66 (2.80)</td>
<td>164</td>
</tr>
<tr>
<td>I-G factors</td>
<td>169</td>
<td>6.75 (2.01)</td>
<td>169</td>
</tr>
<tr>
<td>G factors</td>
<td>157</td>
<td>10.08 (3.29)</td>
<td>157</td>
</tr>
<tr>
<td>G-S factors</td>
<td>145</td>
<td>5.64 (2.70)</td>
<td>145</td>
</tr>
<tr>
<td>Total</td>
<td>145</td>
<td>30.78 (8.10)</td>
<td>145</td>
</tr>
</tbody>
</table>

*Note: I = Individual; I-G = Individual-Group; G = Group; G-S = Group-Societal. For distributions: maximum = 12 for I and G; 8 for I-G and G-S; and 40 for Total. For structural reliability: Mdn CITC = median corrected item-total correlation. For interrater reliability (N = 172): ICC₁ = intraclass correlation coefficient for single ratings and ICC₂ = intraclass correlation coefficient for averaged ratings, both calculated using a mixed effects model, absolute agreement method.*

Proportions of Other considerations coded as present, Scenario planning present, and management strategies recommended are presented in Table A3. A larger proportion of I domain other considerations were coded as present in the practice cases (23% versus 3% for all other domains). Scenario planning was completed in 80% of the cases completed by participants (N = 172). Similarly, management strategies of surveillance/monitoring, control/supervision, assessment/treatment were indicated in over 70% of cases completed, with victim safety planning indicated as a management strategy relatively less frequently.
Table A3. Multi-level Guidelines (MLG) Other Considerations, Scenario Planning, and Management Strategies: Distribution for Individual Raters

<table>
<thead>
<tr>
<th>Other considerations</th>
<th>Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>I domain</td>
<td>23%</td>
</tr>
<tr>
<td>I-G domain</td>
<td>3%</td>
</tr>
<tr>
<td>G domain</td>
<td>3%</td>
</tr>
<tr>
<td>G-S domain</td>
<td>3%</td>
</tr>
<tr>
<td>Scenario planning</td>
<td>80%</td>
</tr>
<tr>
<td>Management strategies recommended</td>
<td></td>
</tr>
<tr>
<td>Surveillance/Monitoring</td>
<td>72%</td>
</tr>
<tr>
<td>Control/Supervision</td>
<td>85%</td>
</tr>
<tr>
<td>Assessment/Treatment</td>
<td>70%</td>
</tr>
<tr>
<td>Victim Safety Planning</td>
<td>35%</td>
</tr>
</tbody>
</table>

Note: N = 172; I = Individual; I-G = Individual-Group; G = Group; G-S = Group-Societal.

The distributions and ICCs for the conclusory opinion ratings for the individual ratings are presented in Table A4. The ratings are well distributed with a relatively greater amount of cases in the sample rated as high for case priority, risk of violence, and risk of life-threatening violence. The ICCs for the conclusory opinions were all poor (Chichetti & Sparrow, 1981) or fair (Landis & Koch, 1977; .33 to .38). These ratings significantly increased to acceptable levels when two cases (Case 1 & 3) were removed from analysis. None of the conclusory opinions were considered good for practical purposes (Chichetti & Sparrow, 1981).

Table A4. Multi-level Guidelines (MLG) Conclusory Opinions: Distribution and Interrater Reliability of Ratings for Individual Raters

<table>
<thead>
<tr>
<th>Presence</th>
<th>Interrater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
</tr>
<tr>
<td>Case Prioritization</td>
<td>20%</td>
</tr>
<tr>
<td>Risk of Violence</td>
<td>20%</td>
</tr>
<tr>
<td>Risk of life-threatening violence</td>
<td>24%</td>
</tr>
<tr>
<td>Risk of Imminent Violence</td>
<td>33%</td>
</tr>
</tbody>
</table>

Note: For distribution (N = 172): L = low; M = moderate; H = high. For interrater reliability (N = 172): ICC₁ = intraclass correlation coefficient for single ratings and ICC₂ = intraclass correlation coefficient for averaged ratings; both calculated using a mixed effects model, absolute agreement method.

When compared to the group consensus ratings, the individual-level results support the use of the MLG in groups (as the MLG was intended to be used), rather than completed
The individual rater results also reflect the impact of the decisions for the quality of training over research design. Specifically, a larger number of items than expected were omitted when cases were completed individually. This was due to the course instructions to “Omit” items if participants were unsure of their responses. Interrater reliabilities for conclusory opinions were poor to fair (ICC₁ = .33 to .38). This was lower than expected, but consistent with studies that have investigated ICCs in a smaller set of practice cases with a greater number of raters whom have varying degrees of experience. For example, Sutherland et al. (2012) reported ICC₁ = .43 to .69 for conclusory opinions on the RSVP, whereas Wilson (2013) reported ICC₁ = .21 to .44 for conclusory opinions on the RSVP. The results also demonstrate that each item was score across the range of possible scores, and for numerical scores (Table A2) and conclusory opinions (Table A4) when assessed individually. Among individual raters the MLG numerical scores and conclusory opinions are distributed and you can communicate low, moderate, and high risk within each of the domains and conclusory judgments. In sum, although these results are not as strong as the consensus rater results, there is also support to a limited degree of the MLG for the individual rater results.