

**THE RELIABILITY AND VALIDITY OF  
THE VIOLENCE RISK SCALE - EXPERIMENTAL VERSION 1**

**A Thesis Submitted to the College of Graduate Studies and Research  
in Partial Fulfillment of the Requirements for the Degree of Master's in Education  
in the Department of Educational Psychology  
University of Saskatchewan  
Saskatoon**

**By  
Audrey Gordon  
Fall, 1998**

**© Copyright Audrey Gordon, 1998. All rights reserved.**

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of the University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Educational Psychology

University of Saskatchewan

Saskatoon, Saskatchewan S7N 0W0

## Abstract

Increasing concern for the safety of society has resulted in a heightened demand for instruments that can accurately predict the risk of violence. As a result, efforts to improve risk prediction techniques have intensified. However, the accurate prediction of violence has proved to be a difficult task, often plagued by inaccuracies and a general lack of validity (Douglas, MacFarlane, & Webster, 1996). The present study represents the first phase in a large scale evaluation of a newly developed risk assessment instrument, the *Violence Risk Scale-Experimental Version 1 (VRS-E1)* (Wong & Gordon, 1996). The VRS-E1 was specifically designed to assess the risk of violence in offender populations. Using a sample of provincial and federal offenders (n=60), subjects were assessed with the VRS-E1 by two independent raters based on a file review and semi-structured interview. The interrater reliability of the VRS-E1 was determined based on the degree of concordance between the scores of two raters for each item and for the total scores using the Kappa (k) Coefficient of Agreement Statistic and Pearson's (r) Product Moment Coefficient. To assess the extent to which the VRS-E1 items are measuring the same or a similar construct, an analysis of the internal consistencies of the scale items was conducted using the Cronbach's alpha. Item analyses were also used to evaluate the independent contribution of each VRS-E1 item to the scale total score. The convergent validity of the VRS-E1 was evaluated by investigating the relationship of the VRS-E1 scores with the scores obtained on validated measures of criminal violence (i.e., the Criminal Career Profile analysis and the Psychopathy Checklist-Revised), criminal risk/need (i.e., General Statistical Information

on Recidivism scale and Level of Service Inventory-Revised), and aggression/hostility (i.e., the Aggression Questionnaire and Interpersonal Behaviour Survey). Discriminant validity of the VRS-E1 was assessed by correlating the VRS-E1 scores with the aggressiveness and assertiveness domains of the IBS, and the correlation of VRS-E1 scores with violent and non-violent offending. Stepwise multiple regression analyses were used to evaluate the relative contribution of the VRS-E1 and comparison measures to the post-diction of violent and nonviolent criminal convictions based on official records. The VRS-E1 demonstrated good interrater reliability. The item by item interrater agreement was strong. The obtained Kappa value using all pairs of ratings (n=1651) was in the 'good' concordance range (Fleiss, 1981). Over 58% of the ratings were in complete agreement and only 7% differed by more than one scale point. A very high alpha coefficient of .92 was obtained indicating that the VRS-E1 is internally consistent and that the items comprising the VRS-E1 are most likely measuring one unidimensional construct. The convergent validity of the VRS-E1 was supported by its significant relationship to validated measures of criminal/violent risk. The discriminant validity of the VRS-E1 was supported by significant correlations with violent but not with non-violent convictions, and with the IBS aggressiveness but not with the assertiveness domains. As well, the results indicate that the VRS-E1 outperforms the other risk assessment instruments in the postdiction of violent convictions. Among the comparison measures, the VRS-E1 contributed most significantly to the postdiction of violent convictions. Overall, the results suggest that the VRS-E1 can be rated reliably by individuals trained to administer it; the scale is internally consistent, and that the VRS-E1 is a valid instrument for assessing the risk of violence in offender populations.

## ACKNOWLEDGMENTS

I want to acknowledge the many individuals who assisted me with this research. First, I want to thank my supervisor, Dr. Donald Saklofske, for his enthusiasm and insightful suggestions. He has provided me with encouragement throughout the project.

I also wish to express my sincere appreciation and gratitude to Dr. Stephen Wong, my co-supervisor (and the co-author of the Violence Risk Scale-Experimental Version 1). Dr. Wong has been a true mentor in every sense of the word. I will always be thankful for his unending encouragement, support, and intellectual guidance. His good humor, kindness, and respect for all people are characteristics that have made our long working relationship especially enjoyable and meaningful.

My gratitude is also extended to the other members of my thesis committee, Dr. Ivan Kelly, for his helpful comments, and Dr. David Jackson, for agreeing to act as my external examiner.

There are many individuals at the Regional Psychiatric Centre (Prairies) who assisted me with this project. I want to express my sincere appreciation to the Clearwater and MacKenzie unit staff for their help in facilitating this research. I also wish to express my sincere gratitude to Dr. Deqiang Gu for his expertise and guidance with the statistical analyses. His many insightful comments and suggestions proved invaluable. My gratitude is extended to Tyson Sawchuk for his assistance with the provincial sample data collection and for allowing me access to his data set. I want to thank Ms. Wendy Pollock for her enthusiastic and conscientious assistance with the federal sample data collection, her devotion to this project exceeded the call of duty. Sarah VanderVeen and Chantal Marin's editorial assistance with the final draft is also gratefully acknowledged.

The Saskatoon Correctional Centre staff were gracious and helpful. I am especially grateful to Rob Arthur (Phoenix Program Director) for his enthusiasm and for facilitating the data collection.

I also wish to express my gratitude to the inmates at the Saskatoon Correctional Centre and Regional Psychiatric Centre (Prairies) who volunteered to participate in the study. Their time and willingness to share their life stories is gratefully acknowledged.

My deepest gratitude is extended to my family for their immeasurable support. I want to thank my mother, for her unfailing love, and my father for all his teachings and his ever-present love.

Finally, I want to thank my husband, John, for his unwavering love, support, and encouragement throughout my many years of university education, and also my children, Sara and Jesse, for their understanding and unconditional love. We have all walked this long road together.

## **DEDICATION**

This thesis is dedicated to my children, Sara and Jesse.

You are my shining stars - thank you for brightening my life.

May the kindness in your hearts and joy in your souls forever guide your way.

## TABLE OF CONTENTS

	PAGE
ABSTRACT .....	iii
ACKNOWLEDGMENTS .....	v
TABLE OF CONTENTS .....	vii
LIST OF TABLES .....	xi
LIST OF FIGURES .....	xii
LIST OF APPENDICES.....	xiii
 1 INTRODUCTION .....	 1
Thesis Foci.....	3
 2 LITERATURE REVIEW .....	 4
Constituent Constructs in Violent Risk Assessment.....	4
The Prediction of Violence - Overview.....	7
THE EVOLUTION OF RISK ASSESSMENT.....	9
First Generation Risk Assessment .....	9
Second Generation Risk Assessment.....	10
The General Statistical Information on Recidivism Scale (GSIR) - Overview .....	10

	Third Generation Risk Assessment.....	15
	Third Generation Risk Assessment-Type A.....	16
	Level of Service Inventory-Revised - Overview .....	16
	Self-Report Version of the Level of Service Inventory (SRI) - Overview .....	20
	Third Generation Risk Assessment-Type B .....	21
	The Psychopathy Checklist-Revised (PCL-R) - Overview.....	21
	New Developments in Violence Risk Assessment.....	26
	The Violence Risk Scale (VRS) - Overview.....	30
	Factor Selection and Operationalization .....	31
	Research Hypotheses .....	33
3	METHODS AND MEASURES.....	34
	Subjects .....	34
	Procedures.....	34
	Measures .....	38
	Aggression Questionnaire.....	38
	Criminal Career Profile .....	38
	Interpersonal Behavior Survey .....	39
	The General Statistical Information on Recidivism Scoring System.....	40
	Level of Service Inventory-Revised .....	40
	Level of Service Inventory - Self-Report Inventory .....	41
	Psychopathy Checklist-Revised .....	41
	Data Analyses.....	42



4	RESULTS .....	43
	Sample Characteristics .....	43
	Provincial Sample Characteristics.....	43
	Federal Sample Characteristics .....	44
	Descriptive Statistics of the VRS and Comparison Measures.....	45
	Descriptive Statistics of the Violence Risk Scale (VRS) .....	45
	Descriptive Statistics for the AQ, GSIR, and PCL-R.....	47
	Descriptive Statistics for the IBS.....	48
	Interrater Reliability of the VRS .....	52
	Interrater Reliability of the VRS (Item by Item Analyses ).....	52
	Interrater Reliability of the VRS Total Score and Comparison Measures....	59
	Internal Consistency of the VRS.....	64
	Validity of the VRS.....	67
5	DISCUSSION .....	69
	Overview .....	69
	Interrater Reliability of the VRS.....	70
	Item by Item Interrater Concordance.....	70
	Interrater Reliability of Total Score Correlations.....	73
	Internal Consistency of the VRS.....	74
	Validity of the VRS .....	75
	Overview of the Validity Investigations.....	75
	Validity Correlations .....	76
	The Relationship Between VRS Scores and the Postdiction of Violence.....	79

Conclusions.....	80
Limitations of the Study .....	81
Recommendations for Future Research.....	82
REFERENCES.....	83
APPENDICES.....	91

## LIST OF TABLES

TABLE	PAGE
1.1	Correlations Between PCL-R Scores and Recidivism.....24
4.1	Means and Standard Deviations of the VRS for the Provincial, Federal, and Combined Samples.....46
4.2	Shapiro-Wilks' Test of Normality for Distribution of VRS Scores.....47
4.3	Mean Scores and Standard Deviations for the AQ, GSIR, and PCL-R for the Provincial, Federal, and Combined Samples.....48
4.4	Means and Standard Deviations for the IBS-Combined Sample.....50
4.5	Means and Standard Deviations for the Level of Service Inventory-Revised Provincial and Federal Samples .....51
4.6	VRS Item by Item Kappa and Pearson Correlation Coefficients Provincial and Federal Samples .....53
4.7	Interrater Reliability of the VRS and Degree of Difference Between Raters in the Combined Sample.....56
4.8	Interrater Reliabilities of the VRS, GSIR, LSI-R, and PCL-R Provincial, Federal, and Combined Samples.....59
4.9	VRS Full Scale and Split-Half Alpha Coefficients - Combined Sample .....66
4.10	Correlations of the Violence Risk Scale with other Criminal Risk/Violence Measures.....68

## LIST OF FIGURES

FIGURES	PAGE
4.1 Percentage of interrater agreement across the total number of ratings (n=1651) expressed as a function of the difference between the scores of Rater 1 (R1) and Rater 2 (R2) .....	58
4.2 Frequency Distributions of VRS Scores Provincial Sample .....	61
4.3 Frequency Distributions of VRS Scores Federal Sample.....	62
4.4 Frequency Distributions of VRS Scores Combined Sample .....	63

## **LIST OF APPENDICES**

Appendix A.	Violence Risk Scale - Experimental Version 1 (VRS-E1) Scale Items.....	91
Appendix B.	Informed Consent - Provincial Sample.....	92
Appendix C.	Informed Consent - Federal Sample.....	93
Appendix D.	Ethics Approval.....	94
Appendix E.	Description and Psychometric Review of VRS Comparison Measures .....	95

## CHAPTER 1

### INTRODUCTION

Predicting who among us will commit a violent act is, increasingly, a paramount consideration in the criminal justice community and among members of psychiatric and psychological professions. As a result, efforts to develop risk assessment techniques which can accurately predict violent recidivism have intensified (e.g., Andrews, 1995; Andrews & Bonta, 1991; Hare, 1991; Gendreau, Little, & Goggin, 1996; Harris, Rice, & Cormier, 1991; Lidz, 1993; Mossman, 1994; Webster & Eaves, 1993; Webster, Harris, Rice, Cormier, & Quinsey, 1994). Although the methodology for risk prediction has become more sophisticated, there has not been a corresponding increase in the precision of risk prediction instruments. This is most notable in the prediction of *violent* re-offending. A number of risk-assessment instruments have been used to predict risk of violence but they have demonstrated weak predictive accuracy (i.e.,  $r \leq .30$ , Serin, 1996). To date, no risk assessment instrument that can measure change in risk level has proved adequately reliable and valid in the prediction of violent recidivism.

The forensic literature suggests that in order to increase accuracy rates in the prediction of *violent* re-offending, the predictor variables considered must be

theoretically or empirically linked to the criterion variable of *violent* recidivism (e.g., Farrington and Tarling, 1985; Lidz, Mulvey, & Gardner, 1993). Furthermore, to ensure optimal utilization of predictor variables, they must be objectively and systematically employed. Therefore it is imperative that the variables are adequately and clearly operationalized.

Wong and Gordon (1996) embraced these recommendations in the design and development of the Violence Risk Scale - Experimental Version 1 (VRS-E1). Based on a comprehensive review of the violence-risk literature and an evaluation of current risk assessment technologies, the VRS-E1 employs 23 Dynamic (changeable) factors and 6 Static (historical) factors that are empirically or theoretically related to violent re-offending. (Hereafter, for the sake of brevity, VRS will be used to indicate VRS-E1). The operationalizations of the selected factors have been subjected to peer review and edited accordingly. Preliminary data on an earlier version of the scale demonstrated good interrater reliability (Gordon & Wong, 1995). Incorporating both research and clinical perspectives, the VRS is designed to provide an objective, systematic and comprehensive evaluation of the risk of violence for incarcerated offenders who are to be released to the community. As well, the VRS is designed to identify high-risk areas that can then be used as treatment targets. Although not investigated at the present time, a section of the VRS (Part B of the scale) is designed to evaluate changes in risk levels as a result of treatment. The VRS is thus designed to evaluate the risk of violence both pre- and post-treatment.

### **Thesis Foci**

The present study represents the first stage in a large-scale evaluation of the interrater reliability, internal consistency and validity of the VRS (Part A). Using a file review and semi-structured interview, two raters independently rated a sample (N=60) of provincially and federally incarcerated male offenders on each of the 29 Factors.

The interrater reliability of the VRS was assessed based on the degree of concordance between the scores of two raters for each item and the total score correlations between Rater 1 and Rater 2. The internal consistency of the VRS was evaluated (Cronbach's alpha) to assess the extent to which the factors comprising the VRS are measuring the same construct. The extent to which the VRS is tapping the construct of 'risk of violent recidivism' was evaluated by investigating the relationship of the VRS (Part A) with valid measures of criminal violence that collectively capture the construct of risk of violent recidivism (i.e., the Psychopathy Checklist-Revised and Criminal Career Profile analysis), criminal risk/need (i.e., the Level of Service Inventory-Revised), and aggression/hostility (i.e., the Aggression Questionnaire and Interpersonal Behavior Survey).



## CHAPTER 2

### LITERATURE REVIEW

#### Constituent Constructs in Violence Risk Assessment

Violence is a multi-faceted construct. Although a certain degree of consensus has been reached in defining violence, no one definition seems to be all encompassing. Violence is defined as an "exertion of physical force so as to injure" (Webster Dictionary, 1989). Megargee (1976) defines violent behavior as "acts characterized by the application or overt threat of force which is likely to result in (physical) injury to people". Both definitions indicate the use of force resulting, or likely to result, in injury. "Threat " is intended to encompass situations such as armed robbery in which injury may be threatened but not accomplished. Although not explicitly stated, "threat of injury" also implies *psychological* injury that may occur under *threat* of physical injury (e.g., as may be experienced by an individual held hostage). The notion of "likely" is included to also capture violent *attempts* (e.g., intentional shooting at someone should be considered as violent even if the bullets miss). A distinction also exists between the *description* and the *evaluation* of violence. For example, an individual who kills another in an act of self-defense has, by definition, committed a violent act. However, this same person is unlikely to be *evaluated* as a violent individual.

Criminal violence is sometimes used to denote various categories of violent acts which have specific legal meaning, for example, assault, armed robbery, sexual assault, etc. (Blackburn, 1993). The term "dangerousness" is often used interchangeably with the term "violent." However, as Monahan (1980) suggests, 'dangerousness' confuses issues regarding what one is predicting with the probability one is assigning to its prediction. In practice, the residual effect is that the word has a tendency to degenerate from a characteristic of behavior to an inherent personality trait (Monahan, 1980). Monahan (1980) states that the term "dangerous behavior" becomes problematic because it is implicitly predictive and instead advocates the conceptually crisper term "violent behavior" (or "violence") that simplifies the separation of definitional issues from probabilistic ones.

Aggression is also sometimes used interchangeably with violence. Aggression is defined as "a forceful action especially when intended to dominate or master" (Webster's Dictionary, 1989). Threatening displays of behavior which stop short of physical injury are considered aggressive. All violence is necessarily aggressive but not all aggressive behaviors result in violence. Violence is more serious, but tends to occur at a lower rate than aggression. Aggression is often the antecedent to many forms of violence. Therefore, in assessing the risk of violence, it is imperative that aggression also be considered.

Aggression, especially verbal aggression (yelling, swearing, shouting, etc.), is a common occurrence among offenders who are prone to violence. Farrington and Gunn (1985) suggest that at its most simplistic level, all criminal activity involves some form of aggression. Research definitions of aggression have focused on the intent, the means

of expression, and the affective factors that underlie aggression (Williams, Boyd, Cascardi, & Poythress, 1996). Aggressive behavior thus consists of behavioral, affective, and cognitive components. The behavioral component of aggression is represented by physical and/or verbal aggression. Anger, represents the emotional or affective component of aggression and the cognitive component of aggression is exemplified by hostility or hostile attitudes (Buss & Perry, 1992). These components of aggression should also be addressed in assessments of risk of violence.

A distinction between assertive and aggressive behaviors has also been made. Mauger and Adkinson (1980) state that assertive and aggressive behaviors are independent response classes. Assertiveness is described as behavior directed at reaching desired goals. Assertive individuals tend to continue in the direction of goals in spite of obstacles in the environment or the opposition of others. In contrast to aggressive individuals, if desired goals are blocked by others, the assertive person's actions are aimed solely at eliminating the interference and not at attacking the offending individual. Although assertive people may be described as competitive in their behaviors, they are more likely than aggressive individuals to abide by societal conventions of fairness (Mauger & Adkinson, 1980).

Mauger and Adkinson (1980) posit that aggressive behavior originates from attitudes and feelings of hostility toward others. Often the purpose of aggressive behavior is to attack or exert power over others. Aggressive behavior may be only incidentally directed toward some instrumental goal. Attainment of the goal may be used as a rationalization for the aggressive actions. Aggressive people may deliberately violate or simply disregard the rights of others in pursuing their goals. Assertive people

may use violent behaviors, however unlike aggressive individuals, they only resort to violence in clearly threatening situations in which other assertive behaviors have proved insufficient as a means of defense (Mauger & Adkinson, 1980). In assessing the risk of violence, it is therefore imperative to distinguish between violent/aggressive behaviors and assertive behaviors.

Due to the complexities in defining violence, throughout this manuscript, violence, criminal violence, and violent behavior, will be used interchangeably to denote any act against a person(s) that results in physical and/or significant psychological injury.

### **The Prediction of Violent Recidivism - Overview**

Violent recidivism refers to an individual being convicted of a violent criminal offense after commission of a prior violent criminal offense. For the purposes of evaluating the predictive validity of various violence prediction instruments, violent conviction(s) is often used as the dependent or criterion variable because it can be objectively measured using criminal records. However, non-adjudicated acts of violence or aggression should also be considered in violent-risk assessment. These acts may include characteristics such as an aggressive interpersonal style, threats of weapon-use, and the association with gangs or peers who are prone toward violence (Gendreau et al., 1996).

In addition to overt behavioral manifestations of violence, certain personality characteristics are highly correlated with violent behaviors. For example, the incidences of Psychopathy and/or anti-social personality disorder are much higher among violent offender populations than in non-violent or non-offender samples (Cornell et al., 1996).

The risk assessment literature strongly suggests that in order to improve the predictive power of risk assessment instruments, it is essential to measure more relevant variables more reliably, more validly and more sensitively (Farrington & Tarling, 1985). A recent meta-analytic study (Gendreau et al., 1996) demonstrated that both static (historical) and dynamic (changeable) factors are important in the prediction of recidivism. Accordingly, the personality, behavioral and contextual characteristics (both static and dynamic) that are theoretically and/or empirically relevant to the criterion variable (i.e., violent re-offending) must be considered in comprehensive assessments of violence risk (e.g., Borum, 1996; Lidz, Mulvey, & Gardner, 1993).

Violence rarely occurs in a vacuum. Therefore, attention to the contextual characteristics of violence is an important consideration in violence-risk assessments. Lidz et al. (1993) for example, advocate a "situational-factor" approach to violence-risk assessment by evaluating the extent to which the occurrence of violence is dependent upon the presence of specific environmental characteristics. Such an approach allows one to consider situations that are *potentially* related to violence (e.g., an offender being released to a high-risk situation).

Given the multi-dimensional nature of the violence construct, the accurate prediction of violent recidivism requires a multi-faceted approach that considers the individual's overt behaviors (both past and present), interpersonal style, personality characteristics, and idiosyncratic contextual triggers that increase the potential for violence. This penchant toward multi-dimensional evaluations of risk is the outcome of an evolving risk-assessment technology.

## **The Evolution of Risk Assessment**

### **First Generation Risk Assessment**

Risk assessment techniques have evolved over time. Bonta (1996) has classified the advancement of risk assessment techniques within a developmental framework. First-generation risk assessments or clinical risk assessments are amalgamations of intuition and judgment. The clinical assessment of risk has been praised for its rich and comprehensive approach. Clinicians usually attempt to capture all the variables that might potentiate violence, independently and/or interactively (Pollock, McBain, & Webster, 1989). In general, clinicians focus on at least three general types of predictor variables, namely personality factors, situational factors and interactions between the two (Monahan, 1980). These predictor variables may be evaluated in terms of either their facilitating or inhibiting influence on the commission of violent acts. The subjectivity of clinical risk assessment however has resulted in relatively less precision and reproducibility compared to statistical prediction techniques (Meehl, 1954). Monahan (1981) stated that among mentally disordered offenders, two out of every three clinical predictions of violence would likely be erroneous. A recent six-year longitudinal study reported that "*professional* clinicians were no more accurate than raters in the assessment of risk" (Menzies & Webster, 1995). This finding supports previous "first-generation" research (see Borum, 1996; Monahan, 1980) which concluded that, compared to statistical techniques, clinical assessments are less valid and less reliable.

### Second Generation Risk Assessment

The lack of precision of clinical assessments spurred the development of second-generation risk assessment instruments. Second-generation risk assessment employs standardized, objective risk prediction tools that are comprised solely of *static* (criminal historical) variables. A recent comprehensive review of the second-generation risk assessment indicated that it has contributed to improved precision in short-term predictions of violence (Otto, 1992). According to Otto (1992) this improvement is likely a result of more contemporary conceptualizations of violence and advances in prediction techniques.

The General Statistical Information on Recidivism (GSIR) scale typifies second-generation risk assessment. At present, the GSIR scale is one of the most widely used assessment tools in the prediction of recidivism. The GSIR was used as a comparison measure in the present study and is reviewed below.

#### The General Statistical Information on Recidivism Scale (GSIR) - Overview

Burgess (1928) developed one of the earliest forms of actuarial risk assessment. Factors that are empirically related to recidivism are rated 0 or 1 indicating the absence or presence of the factor. These scores are summated to produce a total score with an associated 'probability of recidivism'. The GSIR scale and many other risk assessment instruments in use today are based on this methodology (Bonta, Harman, Hann, & Cormier, 1996).

The GSIR scale (Nuffield, 1982), developed in the early 1970's, was designed to provide a structured framework for objective assessments of potential parolees. The

GSIR was constructed and its predictive validity was evaluated using a random sample of roughly 2500 male releasees from federal penitentiaries from 1970 to 1972 inclusive. The follow-up period was three years with re-arrest for any indictable offense as the major outcome variable. Fifteen variables were found to be related to recidivism in the sample (e.g., age at admission, current offense, inmate demographic characteristics, interval at risk since last offense, etc.). Nuffield (1982) used a weighted Burgess method such that the risk factors were assigned weights depending upon their deviation from the base rate of success. For example, an inmate with no prior record of incarceration would be assigned a score of -4 for this factor. On the other hand, inmates with five or more imprisonments would receive a score of +2 for this same factor. Nuffield (1982) reported scores on the scale ranging from -24 to +19 which were used to define one of five risk categories: Very Good Risk, Good Risk, Fair to Good Risk, Fair Risk, Poor Risk. The results indicated that GSIR scores predicted re-arrests over a three-year follow-up period (Bonta et al., 1996) . A linear relationship between Nuffield scores and general success rates was reported. Those with a Nuffield risk score of -14 had a general success rate of 97%, while those with a risk score of +16 had a success rate of only 17% (Hann & Harman, 1992) .

Nuffield (1982) reported that there was a close correspondence between the statistical risk of reoffending (for offenders with similar risk prediction scores) and the likelihood of actually being paroled by the National Parole Board (NPB). However, inmates with the best risk scores were actually paroled at a somewhat lower rate than those with slightly less favorable scores (Nuffield, 1982). This finding suggests that



clinical judgment retained an overpowering influence in the Parole-decision-making process.

Hann and Harman (1989) reported that among categories of inmates having the same GSIR release risk score, those granted parole by the NPB were more successful after release than those who had been denied parole. He compared the release success rates of inmates with worse release risk scores who were paroled and those with better release risk scores who had been denied a parole. In many cases, inmates in the former groups had better release success rates than did inmates in the latter groups. For instance, the success rates for paroled inmates with each of the scores of 0 to 4 (i.e., "Fair Risk" category) were all above the success rates for those who were denied parole (but were eventually released) and had scores of -5 to -1 (i.e., "Good Risk" category).

The predictive accuracy of the GSIR is more restricted when comparing neighboring risk categories rather than the more extreme categories. For example, a decision to release an offender who scored in the "Very Good Risk" category would be relatively less ambiguous and less precarious than making a decision to release an offender categorized as "Fair Risk". Thus, in practical terms, the five GSIR risk categories may have more statistical than functional significance.

Bonta et al. (1996) re-examined the GSIR on a new sample of federal inmates with the same post-release follow-up period as used with Nuffield's construction sample. The usefulness of the GSIR with respect to the prediction of violent recidivism was investigated. The researchers examined the suggestion that the federal inmate population has become more violent in the past decade (relative to Nuffield's early 1970's sample which had a very low base rate for violent re-offending - 12.6%, with

robbery included in the definition). Bonta et al. (1996) proposed that finding a higher base-rate for violent re-offending would provide a better test of the GSIR's ability to predict violent re-offending.

Two measures of violent recidivism were employed. The narrow violence category included homicide, sexual assault, and aggravated assault. The broad category of violence encompassed the narrow violence offenses as well as weapons offenses, robbery, and less serious sexual offenses. Narrowly defined violent crimes comprised 9.8% of the offenses, and broadly defined violent crimes, 18.6% of the offenses. GSIR scores ranged from -20 to +24. All items on the modified GSIR scale, with the exception of *Previous Conviction for Violent Sex Offence*, were predictive of recidivism. Although this item did not predict either general recidivism or broadly defined *violent* recidivism, it was important for predicting narrowly defined violent recidivism. Bonta et al., (1996) suggests that this finding is consistent with the contention that previous behavior is generally one of the better predictors of future behavior, and that predictive power is improved when these behaviors are specifically defined (Andrews & Bonta, 1994). Also, these findings support the contention that predictive precision will be enhanced by focusing on predictors that are conceptually related to the criterion variable and are precisely defined (Bonta et al., 1996) .

Scores on the GSIR scale showed only modest associations with future violent behavior. Pearson product-moment correlations between scores on the GSIR scale and both narrowly-defined and broadly defined violent recidivism were low to moderate ( $r=.15$ ,  $p \leq .001$  and  $r=.20$ ,  $p \leq .001$ , respectively). Multiple regression analyses were

used to assess the validity of the GSIR in the prediction of violence resulting in correlations of only  $r=.22$  for narrowly-defined and  $r=.27$  for broadly-defined violence.

Bonta et al. (1996) concluded that the GSIR performed adequately as a stable predictor of general recidivism for a federal offender population. With the exception of violent sex offence history, all the items demonstrated predictive validity for general recidivism. The risk categories comprising the scale showed systematic associations with general recidivism outcomes, although the middle categories were less informative. With respect to violent recidivism, the GSIR's predictive validity is limited.

As discussed above, more contemporary research suggests that the predictive validity of violent-risk instruments will be enhanced by considering particular contextual or environmental conditions that may enhance or diminish violent tendencies. From this perspective, the GSIR's limited predictive accuracy concerning violent recidivism may be related to the exclusion of dynamic predictor variables. Indeed, the major criticism against the GSIR and other static risk scales is that they fail to take into account present functioning. As well, changes in offender functioning are not reflected in subsequent re-assessments. Unlike dynamic risk assessment that considers the changes in predictor variables over time and treatment, variables in static risk scales are not malleable. Undoubtedly historical criminal factors are an important consideration in risk-assessment. However, more contemporary research indicates that violent-risk instruments that also include dynamic predictor variables (that are empirically or theoretically related to violence) outperform the predictive validity of

actuarial risk-instruments which rely solely on historical variables (e.g., Gendreau et al., 1996; Loza & Simourd, 1994).

Monahan and Steadman (1994) suggest a number of means by which violence prediction could be enhanced. They contend that (a) to maximize the external validity of risk assessments, widely representative cohorts of participants must be selected; (b) more sophisticated dimensional measures should replace the more traditional forced-choice "yes-no" dualisms of risk predictor and criterion variables; (c) to appraise their relative and collective usefulness, actuarial and clinical risk assessments should be directly compared; (d) multiple and detailed socio-demographic, medical, and psycho-legal factors need to be incorporated into risk evaluations, which should occur in various contexts and recurrently over time; (e) assessments must consider both situational as well as individual-specific factors associated with violent behavior; and (e) the criterion of violent recidivism must be evaluated using convergent sources and within a variety of contexts (e.g., psychiatric, penal, and community-situated). A number of third-generation risk-assessment instruments described below have incorporated many of these criteria.

### Third Generation Risk Assessment

Third-generation risk assessment instruments are theoretically based rather than empirically derived. Two main categories of third-generation instruments have been identified (Bonta, 1996). Although both types are similar, they vary with respect to the emphasis placed on assessing static and dynamic factors and the rationale for developing the measures. For example, the Level of Service Inventory (Type A) is

based upon a social learning theory of criminal behavior and Hare's Psychopathy Checklist (Type B) was derived from theoretical conceptions of psychopathy.

### Third Generation Risk Assessment - Type A

Type A-third generation risk assessment evaluate risk based on a wide range of static risk and dynamic (need) factors. The variables are broadly grouped according to criminogenic characteristics (e.g., abilities, attitudes, skills), environment (e.g., criminal associates, living arrangements), and circumstances (e.g., finances, employment, substance use/abuse) (Loza & Simourd, 1994) . One of the most widely researched *risk/need* instruments is the Level of Service Inventory- Revised (LSI-R)<sup>1</sup> (Andrews & Bonta, 1991) . The Level of Service Inventory is available in both a scored (LSI-R) and self-report (SRI) format. The LSI-R is designed to assess attitudes, behaviors and environmental characteristics and was developed to assist case managers in supervision decisions concerning adult probationers and parolees. The instrument is designed to assess the offender's risk of recidivism as well as identifying target areas for treatment interventions.

#### Level of Service Inventory-Revised - Overview

The current version of the LSI-R comprises 54 dynamic and static items denoting specific risk variables grouped into 10 subcategories (i.e., criminal history, education/employment, financial, family/marital, accommodations, leisure/recreation, companions, alcohol/drug problems, emotional/personal, attitudes/orientation).

Individual items are scored (using both interview and file information) in a binary

---

<sup>1</sup>Research studies refer to the scored instrument by different names Level of Supervision Inventory, Level of Supervision Inventory-Revised (i.e., LSI, LSI-R or LSI-IV). The final published version is referred to as the Level of Service Inventory-Revised (LSI-R). All of the research studies used essentially the same instrument and therefore all of the designations can be considered synonymously.

format indicating the absence or presence of the risk/need variable (with "1" indicating the presence of the attribute described by the item). When summed, higher scores reflect a greater risk of recidivism and need for clinical intervention. Psychometric evaluations of the LSI-R indicate that the underlying construct measured by the LSI-R is a "propensity for rule violation". Among probationers and inmates released to halfway houses, the LSI has demonstrated internal consistency ( $\alpha = .72$  and  $.71$  respectively), interrater reliability ( $r = .94$ ), and temporal stability ( $r = .80$ ) over several months (Andrews, 1982; Andrews, Kiessling, Mickus, & Robinson, 1986; Bonta & Motiuk, 1985). Compared with two second-generation static instruments (i.e., the Statistical Information on Recidivism scale (GSIR) and the Salient Factor Score (SFS), the LSI yielded the highest correlation with parole release outcome ( $r = .53$ ) followed by the GSIR ( $r = .44$ ) and SFS ( $r = .38$ ) respectively.

Gendreau et al. (1996) compared the LSI-R with another third-generation *need*-based classification system, the Wisconsin Classification System (Baird, 1981), as well as a second-generation actuarial instrument, the Salient Factor Score (Hoffman, 1983). Compared to these measures, the LSI-R produced a non-significant but higher correlation ( $r = .35$ ) with recidivism (Gendreau et al, 1996).

Other research has examined the validity of the LSI-R on a variety of criterion measures. For example, in a sample of provincially-sentenced inmates, the LSI-R produced moderate to strong relationships with staff evaluations of case progress ( $r = .40$ ), completion of sentences in half-way houses ( $r = .52$ ), recidivism while on probation ( $r = .47$ ), post-program recidivism ( $r = .47$ ), severity of reoffense ( $r = .39$ ), and

likelihood of reincarceration ( $r=.40$ ) (Loza & Simourd, 1994). Loza and Simourd (1994) examined the psychometric properties and convergent validity of the LSI with federal offender samples in Canada and concluded that the LSI is a reliable risk/need instrument for use with federal offenders. In agreement with previous research, satisfactory levels of internal consistency were found for the LSI total score and subtotals. Loza and Simourd (1994) predicted that LSI-R scores would be higher for violent offenders than for nonviolent offenders. Also tested and supported in the study was the hypothesis that the overall pattern of results for the federal sample would be consistent with previous LSI-R research on provincial incarcerates that has indicated that LSI scores among federal offenders is similar to that of provincial offenders (e.g., Motiuk, Motiuk, & Bonta, 1992; Loza & Simourd, 1994).

LSI total scores ranged from 1 to 41 ( $M=26.2$ ,  $SD=9.9$ ). Violent offenders comprised 64% of the sample and were classified according to a criterion of having committed at least one major (i.e., murder, manslaughter, assault, kidnapping, forcible confinement) or serious (robbery, sexual offenses) past or present offense. Violent offenders had statistically significant greater mean scores on the LSI total score ( $M=27.4$ ,  $SD=9.1$  versus  $M=23.9$ ,  $SD=10.9$ ,  $F[1,159]=4.95$ ,  $p \leq .05$ ).

The convergent validity of the LSI-R was examined using the General Statistical Information on Recidivism scale (GSIR) and the Psychopathy Checklist-Revised (PCL-R). Overall, the strongest correlations were found between the PCL-R and LSI-R (Loza & Simourd, 1994). Correlations between Factor 2 (behavioral component) of the PCL-R and LSI total score and subtotals were particularly strong (e.g., LSI-R criminal history:  $r=.75$ , LSI-R total score:  $r=.84$ ). These findings are consistent with other

research documenting the strength of the relationship ( $r=.78$  range) between the LSI-R and PCL-R (Gendreau et al., 1996) . Except for the LSI Alcohol/Drug subtotal ( $r=.17$ ), PCL-R Factor 1 (personality component) and LSI correlations reached acceptable levels ( $r=.30$  to  $.53$  range). The highest correlations were found for the Attitudes/Orientation, Education/Employment, Family/Marital, and Companions subtotals. Correlations between the GSIR and the LSI total score, and the Criminal History, Education/Employment, Family/Marital, Companions, Attitudes/Orientation, and Accommodations subtotals, were moderate to strong. Weak correlations were reported for the Financial and Alcohol/Drug LSI subtotals (Loza & Simourd, 1994). Loza and Simourd (1994) contend that these results are consistent with a social psychological approach to criminal conduct that suggests that a broad range of risk/need variables are related to an individual's propensity toward rule violation.

A recent meta-analysis (Serin, 1996) evaluated the most prominent risk instruments and predictor variables. The maximum predictive accuracy of violent risk prediction was in the  $r \leq .28$  range achieved via correlational analysis between violent recidivism and scores offenders obtained on the PCL-R (described below). Although a strong relationship between the LSI-R and PCL-R ( $r=.78$ ) has been reported, the validity of the LSI-R as an independent violent-risk assessment instrument has not yet been established. As a result, it is recommended that the PCL-R be used in conjunction with the LSI-R in assessing the risk of violence (Gendreau et al., 1996) . Used collaboratively, the LSI-R and PCL-R should provide a more accurate estimate of violence-risk than either measure used independently. Although the two instruments



overlap in some regards, they offer distinct yet compatible approaches to predicting the risk of violent recidivism.

#### Self-Report Version of the Level of Service Inventory (SRI) - Overview

Offender self-reports have been used extensively in forensic research (Motiuk, Motiuk, & Bonta, 1992). A self-report version of the interview-based LSI was constructed using the items in the inventory to generate self-referent statements regarding personal attributes and situations (Motiuk, 1988, cited in Motiuk et al., 1992). The LSI-R Self-Report Inventory (SRI) is a paper-and-pencil scale comprised of 78 items scored in a 0 to 1 format. The SRI can be administered to offenders who possess a Grade 6 reading level. The SRI is comprised of the same sub-components as the LSI-R (described above). The majority of SRI items are in a true-false or yes-no response format with the remainder requiring a specific answer (e.g., How many times have you been convicted of an offense since your 16th birthday?). One item has a 4-point Likert-type response format.

Using a sample of 100 provincially incarcerated inmates, Motiuk et al., (1992) evaluated the relationship between the SRI and LSI. SRI total scores (4 to 41) and LSI total scores (4 to 42) produced a correlation of .78 ( $p \leq .001$ ). SRI total scores were examined for predictive validity in relation to institutional performance as measured by the number of misconducts and the number of assaults. Controlling for number of days in custody, significant partial correlations were found for prison misconduct ( $r = .17$ ,  $p \leq .05$ ) and assaults ( $r = .19$ ,  $p \leq .05$ ).

One-year post-release data was reported for 97 of 100 inmates. The SRI total scores, although significantly related to both parole violation ( $r = .29$ ,  $p \leq .05$ ) and

reincarceration ( $r=.26$ ,  $p\leq.01$ ) were modest in accounting for the variance (Motiuk, 1992). The incremental predictive criterion validities of the SRI and LSI were evaluated through multiple regression analyses, using prison misconduct, assault, halfway house outcome, parole violations, and reincarceration as separate criterion variables. Only for halfway house outcome (success/failure) did the LSI demonstrate any incremental validity relative to the SRI (Motiuk et al., 1992).

The SRI predicted both institutional performance and post-release recidivism as accurately as the LSI. As yet, no data is available concerning the validity of the SRI with federal inmate populations although given the reported similarities between the LSI and SRI, it is expected that the SRI will also be reliable and valid in federal inmate populations.

### Third-Generation Risk Assessment - Type B

The second type of third-generation risk instrument is more narrowly focused and is aimed at evaluating specific populations such as those with anti-social personality, sociopathy, or psychopathy.

#### The Psychopathy Checklist-Revised (PCL-R) - Overview

As noted above, the recidivism literature reports that the PCL-R to date, is one of the most widely accepted risk prediction instruments. The PCL-R was developed to assess psychopathy in male forensic populations. It purports to measure both behavior and personality traits that are characteristic of the clinical conception of psychopathy (e.g., superficial charm, lack of remorse or shame, failure to follow life plan, etc.). Psychopathy has an early onset, is characteristic of the individual's long-term functioning, and results in social dysfunction or criminality (Fulero, 1995) .

Although the term 'psychopathy' is sometimes used synonymously with the DSM-IV classification of Anti-Social Personality Disorder (APD), there are distinctions between them. The DSM-IV classification has been criticized for placing too much emphasis on the behavioral characteristics of APD. This focus on behavioral characteristics results from indications that personality characteristics are not reliably and validly measurable (Fulero, 1995). In fact, dissatisfaction with the DSM-III classification spurred the development of the PCL (Hare, 1980). Psychopathy is a more focused assessment with only 20%-30% of those diagnosed as APD meeting the PCL-R criteria for psychopathy (Hart & Hare, 1989). This state of affairs has led to the recommendation that individuals in forensic settings be categorized as psychopathic or non-psychopathic APD (Meloy, 1995; cited in Serin, 1996). Studies indicate that correlations between APD and recidivism are typically lower than are correlations between psychopathy and recidivism (Hart & Hare, 1989; Harris, Rice, & Quinsey, 1993; Harris et al., 1991). Thus, distinguishing between APD and psychopathy is important in evaluations of risk of recidivism.

The original Psychopathy Checklist (PCL, 1980) contained 22 items. In the revised version of the PCL, the PCL-Revised, the criteria for scoring the items were revised, several item descriptions were modified, and two items were deleted.

The PCL-R is rated on a three-point ordinal scale, based on the degree of match between the personality/behavior of the individual and the item description in the manual (2=Yes, 1=Maybe/in some respects, 0=No). Both file review and the administration of a semi-structured interview are required to obtain information for the ratings. The PCL and the PCL-R have virtually the same psychometric properties and

external correlates and are also highly correlated. They can be considered as measures of the same construct (Hare, 1991). Factor analysis of the PCL-R revealed a two-factor structure. Factor 1 comprises the interpersonal/affective traits (i.e., enduring personality characteristics such as selfishness, callousness, remorseless, shallow affect, etc.). Eight of the items load .40 or above on this factor. Factor 2 is characterized by impulsive and antisocial behavioral characteristics. Nine of the items load above .40 on this factor. Three of the twenty items do not load on either Factor 1 or Factor 2. The two factors correlate at approximately .5 with each other, indicating the instrument is indeed tapping dissimilar but overlapping constructs.

In a recent review of the literature, Hemphill and Wong (1997) reported that follow-up studies on the predictive validity of the PCL-R indicate that it is one of the strongest single predictors of violent recidivism and of sexual recidivism (Harris, Rice, & Cormier, 1991; Harris, Rice, & Quinsey, 1993; Quinsey, Rice, & Harris, 1995; Rice & Harris (1995). Serin (1996) reported that the correlations between PCL-R scores and violent recidivism were  $r=.28$  and  $r=.29$  on the full 20-item version of the PCL-R and a shortened 18-item PCL-R deleting Item 10 (Poor Behavioral Controls) and Item 20 (Criminal Versatility) respectively. The respective correlations between PCL-R scores and non-violent recidivism were  $r=.31$  for the full length PCL-R and  $r=.28$  for the shortened 18-item PCL-R (Hemphill & Wong, 1997). Recidivism was more strongly correlated with PCL Factor 2 than with PCL Factor 1 in three of five studies. In two of the ten studies examined in Hemphill and Wong's review (because reconvictions represent a conservative estimate of re-offending), reconviction, conditional release violations, and rearrests or rehospitalization were also included as criterion variables for

recidivism. In the studies reviewed, the PCL-R predicted any (non-violent) recidivism and violent recidivism equally well, but actuarial risk scales predicted general, non-violent recidivism better than they predicted violent recidivism (Hemphill & Wong, 1997).

Hemphill and Wong's (1997) findings are consistent with another recent study that evaluated the predictive efficacy of the PCL-R using official recidivism within a two-year follow-up as the dependent variable (Serin, 1996). For this sample (n=81, minimum and medium risk, predominantly White offenders), 75% were categorized as violent offenders - a surprisingly high base rate considering that the sample did not comprise maximum security offenders. Non-violent offenses included property crimes such as theft. Violent offenses comprised robbery, assault, manslaughter, sexual assault, and murder. Psychopaths were defined as those offenders with PCL-R total scores of 29 or greater, nonpsychopaths as those with PCL-R total scores less than or equal to 16, and a mixed group comprised the balance. The mean PCL-R score was 22.1 (SD=6.7). The results of the recidivism data for the PCL-R are presented in Table 1.1.

**Table 1.1**  
**Correlations Between PCL-R Scores and Recidivism**

Measure	F1	F2	Recidivism	Violent Recidivism
PCL-R	.61**	.91**	.31*	.28*
Factor 1		.31*	.14*	.26*
Factor 2			.36**	.22
Recidivism				.29*

\*  $p \leq .01$ , \*\* $p \leq .001$

The overall failure rate of the sample was 57%. General failure rates were 40% for nonpsychopaths (n=20), 51.2% for the mixed group (n=41) and 85% for the psychopaths (n=20). Categorizing PCL-R scores 0-10, 11-20, 21-30, and 31-40 resulted in failure rates of 50%, 44.4%, 62.2%, and 80%, respectively. Overall the violent recidivism rate was 10%. None of the nonpsychopaths recidivated violently. However, 7.3% of the mixed group, and 25% of the psychopaths recidivated violently (Serin, 1996). Also significant is that although 85% in the PCL-R psychopathy category (i.e., 31-40) recidivated, only 25% of those offenders recidivated violently. Clearly, many violent offenders are not psychopathic. Therefore, to predict violence, measures other than those used to capture psychopathy are needed.

In summary, the PCL-R is a psychometrically sound and effective instrument for measuring the construct of psychopathy. Although primarily designed to assess psychopathy, the use of the PCL-R has been extended to the prediction of recidivism and violence.

These findings, and the results across studies that found larger PCL/violent recidivism correlations than actuarial risk scales/violent recidivism correlations, suggest that actuarial risk scales may be more limited in the prediction of violence (relative to the PCL-R) than in the prediction of general recidivism (Hemphill & Wong, 1997). In contrast to purely empirically-based assessment, the PCL-R predicts violent recidivism and any recidivism equally well because the relationship between the PCL-R and recidivism rests more strongly on a theoretical foundation (Hemphill & Wong, 1997). Construct-related predictors such as the PCL-R that are theoretically derived should be

associated with a broader range of socially deviant and violent behaviors than are actuarial risk scales that employ only historical factors to predict recidivism (Hemphill & Wong, 1997). As such, instruments designed specifically to assess violence-risk, and are comprised of empirically or theoretically related predictor variables, should logically predict violence more accurately than purely empirically derived or less focused instruments. The current trend in the development of a number of violence-risk instruments (e.g., the HCR-20) is based on this assumption.

### **New Developments in Violence Risk Assessment**

The results of a recent meta-analysis using 131 recidivism studies (Gendreau et al., 1996) concurred with previous findings indicating that variables such as age, criminal history, companions, family factors, gender, social achievement, and substance abuse are significant and potent predictors of recidivism. The results of the study also provided strong support for the inclusion of dynamic predictor variables in risk assessment. Of the two major static and dynamic categories, Gendreau et al. (1996) reported that criminal history and criminogenic needs were equivalent in predicting recidivism, producing mean  $r$  values of .16 and .18 respectively. Overall, the static and dynamic domains were comparable in their ability to predict recidivism (mean  $r$  values=.14 and  $r$ =.13 respectively).

Contemporary third-generation violence risk assessment instruments focus on predictor variables that are either theoretically or empirically linked specifically to violence. This approach differs from less focused instruments such as the LSI-R

(which tap the broader construct 'propensity for rule violation') or the PCL-R (designed to measure psychopathy).

The most notable instruments that were designed specifically to address the risk of violence are the Dangerous Behavior Rating Scheme (DBRS, Webster & Menzies, 1993); the Historical, Clinical, Risk Scheme (HCR-20, Webster & Eaves, 1993) ; and the Violence Prediction Scheme (VPS, Webster et al., 1994). By focusing on predictor variables related to the construct of violence, these tools offer an advancement in violence-risk assessment. At the present time however these instruments are still under development and have not been validated.

The Dangerous Behavior Rating Scheme (DBRS) has been criticized for unclear operational definitions for the items and the inclusion of items that were not empirically linked to violent behavior. Borum (1996) suggests that these inadequacies may be responsible for constraining the predictive validity of the instrument. In a recent study, the DBRS demonstrated weak predictive validity estimates ( $r=.16$ ,  $r=.18$  and  $r=.15$ ) after one year, three years, and six years respectively (Menzies & Webster, 1995) .

The HCR-20 (Webster & Eaves, 1993), a systematic model for the assessment of risk for future violent behavior in criminal and psychiatric populations, holds more promise. The HCR-20 utilizes both static and dynamic factors. The scale is divided into three assessment categories comprising a total of twenty predictor variables. Ten static (historical) factor variables are used to assess past behaviors (e.g., age at first violent offense, extent of previous violence). Five clinical variables (e.g., insight, attitude) evaluate present functioning and five risk variables (e.g., support and supervision, compliance, stress) are used to predict future functioning. The HCR-20



has been praised for its operationally defined coding system that facilitates increased reliability, and also for its practical utility (Borum, 1996). Preliminary data compiled via retrospective analysis demonstrated significant correlations between the Historical (H) and Clinical (C) scales of the HCR-20 and scores on the Psychopathy Checklist-Revised and the number of previous charges for violent offenses (Douglas, Webster, Eaves, Wintrup, & Hart, 1996; cited in Borum, 1996). Much of the existing HCR-20 data was compiled using samples of mentally-disordered offenders. In addition, the reliability and validity estimates of the HCR-20 are preliminary. Accordingly, the HCR-20 is primarily used as a research tool and/or as a checklist to prompt clinicians to consider the major relevant areas of risk inquiry (Borum, 1996).

The Violence Prediction Scheme (VPS) is one of the most recent additions to the violence-prediction domain. The VPS was developed to assess dangerousness and risk in a maximum security psychiatric hospital population using both actuarial and clinical factors. The core of the VPS is the Violence Risk Assessment Guide (VRAG), which has the PCL-R as one of its major predictor variables. Twelve static variables comprise the VRAG all of which have demonstrated predictive power with respect to violent recidivism: 1) PCL-R,  $r=.34$ ; 2) elementary school maladjustment,  $r=.31$ ; 3) DSM-III diagnosis of personality disorder,  $r=.26$ ; 4) age at index offense,  $r=.26$ ; 5) separated from parents under age 16,  $r=.25$ ; 6) failure on prior conditional release,  $r=.24$ ; 7) non-violent offence history,  $r=.20$ ; 8) never married,  $r=.18$ ; 9) DSM-III diagnosis of schizophrenia,  $r=.17$ ; 10) victim injury,  $r=.16$ ; 11) alcohol abuse,  $r=.13$ ; and 12) female victim index offense,  $r=.11$ .

In a seven-year follow-up, Harris et al. (1993) reported a multiple correlation of .46 between the VRAG and subsequent violence in the community in a sample of psychiatric offenders. The VPS also contains an interview-based assessment, the ASSESS-LIST incorporates actuarial, clinical and situational factors. The ASSESS-LIST is an acronym for the following risk prediction domains: 1) Antecedent History; 2) Self Presentation; 3) Social and Psychosocial Adjustment; 4) Expectations and Plans; 5) Symptoms; 6) Service; 7) Life Factors; 8) Institutional Management; 9) Sexual Adjustment; 10) Treatment Progress. The VPS represents a step forward in comprehensive risk assessment using static and dynamic predictors. The items in ASSESS-LIST are used as a guide for the assessment of violence. The most significant criticism against the VPS is that the recidivism estimates were derived from a sample of mentally disorder offenders and for this reason, the generalizeability of the results to other populations is questionable (Webster, 1994). As well, the VRAG consists of static variables only and therefore, it is unable to measure change.

Although the DBRS, HCR-20, and VPS were developed specifically to assess the risk of violent-recidivism, the data is too limited at the present time to permit their inclusion as violent recidivism prediction measures that are sensitive to change. As discussed above, the utilization of LSI-R, the GSIR and even the PCL-R, as violent risk measures (although quite common), is in fact an extension of the original intent and design of these instruments. Furthermore, the GSIR, PCL-R and VRAG are comprised solely of static variables. Therefore, it is not surprising that these instruments lack precision with respect to the prediction of violent recidivism as they are unable to capture the variance within the very important domain defined by the dynamic

variables. At the present time, there is no reliable and valid instrument available with both static and dynamic variables designed specifically to assess the risk of violent recidivism in offender populations.

The accuracy of a violence-risk prediction instrument should be maximized when the predictor variables encompass the many sub-components of the construct of violent recidivism, for example, interpersonal aggression, psychopathic personality, criminal history, etc.). The Violence Risk Scale-Experimental Version 1 (VRS, Wong & Gordon, 1996) was developed based on these premises.

### The Violence Risk Scale (VRS) - Overview

The VRS is a comprehensive instrument designed to assess the risk of violent recidivism of offenders that are to be released from incarceration. The design of the VRS results from a thorough review of the risk prediction literature together with an examination of the strengths and limitations of the current risk assessment methodologies. The VRS is comprised of both static and dynamic factors and incorporates both clinical and research perspectives. Thus, the VRS addresses Webster, Eaves, Douglas, and Withrup (1995) assertion that "the greatest challenge in what remains of the 1990s is to integrate the almost separate worlds of research on the prediction of violence and the clinical practice of assessment - at present the two domains scarcely intersect".

The VRS includes 6 static and 23 dynamic factors (Appendix A). The selection of the static and dynamic factors was based on their empirical or theoretical relationships to violent recidivism. The scale has a pre- and a post-treatment section. Each pre-

treatment factor in Part A describes a construct which can be rated on a 4-point (0,1,2,3) scale based on a file review and a semi-structured interview. The higher the rating, the more the construct is associated with violence for the offender in question. The sum of the scores for Part A for the 29 factors is a quantitative measure of the risk of violence pre-treatment; the higher the score, the higher the risk of violence. High individual factor scores are useful in pinpointing idiosyncratic risk areas where treatment interventions can be targeted. Although not investigated at the present time, the VRS-Part B is designed to evaluate the change in violence on each of the 23 dynamic factors.

#### Factor Selection and Operationalization

The VRS is different from other risk assessment instruments in that each predictor variable is operationalized in terms of its idiosyncratic relationship with violence.

For each Static Factor, a detailed rating description is provided for each of the 4-scale points. The Static Factors (S1 to S6) are: Current Age, Age at First Conviction, Number of Juvenile Convictions, Violence throughout Lifespan, Prior Release Failures/Escapes, Stability of Family Upbringing. With the exception of Current Age, the Static Factors do not change over time in such a way that would lead to a change in risk.

For the Dynamic Factors (D1 to D23), the prototypical characteristics for only the two extreme anchor points (i.e., 0 and 3) of the scale are provided. The rater is required to assess if the offender's characteristics match the rating descriptions. A rating of 1 is considered as more serious or intense than 0; and a rating of 2 is considered as less serious or intense than 3. The Dynamic Factors (D1 to D23) are: Violent Lifestyle,

Criminal Personality, Criminal Attitudes, Work Ethic, Respect for Authority and Social Conventions, Criminal Peers, Interpersonal Aggression, Emotional Disinhibition, Violence During Incarceration, Weapon Use, Insight into the Cause of Violence, Mental Disorder, Substance Abuse, Stability of Relationships with Significant Others, Community Support, Released to High Risk Situations, Violence Cycle, Impulsivity, Violent Sexual Behavior, Compliance with Community Service, Stages of Change, Security Level of Release Institution, Longest Violent Incident Free Period.

The VRS also contains a 'Clinical Override' item which allows the rater to make special judgments of the risk of violence independent of the ratings indicated by all the other factors. The override is necessary to cover extraordinary situations that may be unique to the case; for example, a specific threat of violence made toward a known individual may be considered serious enough to override even a composite rating assessment of low risk using the instrument.

In summary, the VRS is a newly developed risk assessment tool that employs a broad range of both static and dynamic factors. It is unique in that the factors are operationally defined in terms of their relationships to violence. Identified high-risk factors can be used as treatment targets and, Part B of the VRS is specifically designed to assess changes in risk as a result of treatment. No other risk assessment instrument is designed specifically to assess the risk of violence and changes in risk as a result of treatment.

## Research Hypotheses

The focus of the present study was to investigate the interrater reliability, internal consistency and construct validity of the VRS.

The interrater reliability of the VRS was evaluated based on the degree of concordance between two raters for each item and the total scores. It was predicted that the VRS would demonstrate 'good' interrater concordance for each of the 29 factors, as measured by the Kappa coefficient of agreement statistic ( $\kappa > .40$ , Fleiss, 1981). Also, it was predicted that the interrater reliability of the VRS, as measured by the total score correlations between Rater 1 and Rater 2 would be significant at the .05 level (Pearson Product Moment Coefficient) and demonstrate good overall concordance.

The internal consistency of the VRS was evaluated using *Cronbach's coefficient alpha*, a measure that estimates the internal consistency of items in a scale. It was predicted that the Cronbach's alpha of the scale would be above the .70 level, indicating that the VRS items are measuring the same or a similar construct.

The construct validity of the VRS was evaluated by investigating the relationship of the VRS (Part A) with validated measures of criminal violence that collectively capture the construct of risk of violent recidivism (i.e., the Psychopathy Checklist-Revised and Criminal Career Profile analysis), criminal risk/need (i.e., the Level of Service Inventory-Self-Report version), and aggression/hostility (i.e., the Aggression Questionnaire and Interpersonal Behavior Survey). It was hypothesized that the ratings attained on the VRS, would be significantly correlated with the comparison measures at the .05 level (Pearson Product Moment Coefficient).

## CHAPTER 3

### METHODS AND MEASURES

#### Subjects

In order to sample a more diverse population with both lower risk and higher risk offenders, both provincial and federal incarcerates were included in the study. Thirty (30) provincial male inmates (mean age = 29.2 years, SD = 7.4) at the Saskatoon Correctional Centre and thirty (30) federal male inmates (mean age = 35.3, SD = 8.9) at the Regional Psychiatric Centre (Prairies) were recruited as volunteer participants in the study. For the provincial sample, subjects were selected based on an inclusion criteria requiring a minimum one-month sentence length. All subjects in the federal sample were serving sentences of greater than two years.

#### Procedures

Inmates selected according to the above criteria, were presented with relevant information concerning the study on an individual basis or through a group presentation presented by the researcher. Potential recruits were asked if they wished to volunteer for a study which would evaluate the reliability and validity of an instrument designed to assess the risk and need areas of offenders. Subjects were advised that the assessment would be based on information obtained during two privately administered semi-structured interviews<sup>1</sup> with separate researchers approximately one week apart, a

---

<sup>1</sup> The semi-structured interview was designed to solicit information to rate the VRS, PCL-R, and LSI-R.

review of information contained in their institutional files, and their responses to a set of self-report questionnaires.

Subjects were advised that each interview would require approximately one hour and that completion of the self-report measures would take approximately one additional hour. Participants were asked to read and sign an "Informed Consent" form (Appendices B and C) for the provincial and federal sample respectively), which specified the conditions of participation. Confidentiality was ensured by advising all participants that their names, including any other identifying information, would not be associated with their responses. Participants were advised that, as delineated in the Informed Consent, researchers may have to release the results of the research if ordered by a Court of Law.

Each subject completed one set of the self-report measures. For the provincial sample, the self-report measures included the Aggression Questionnaire, the short-version Interpersonal Behavior Survey (Parts I and II), and the Level of Service-Self-Report Inventory. Self-reported reading levels were used to determine whether the subject met the grade six reading criteria for completion of the questionnaires. The questions were read aloud to the few participants who failed to meet the minimum reading level requirement and also to those who required or requested assistance in completing the questionnaires.

The provincial sample was collected prior to the federal sample. Two modifications were made prior to data collection for the federal sample. The self-report version of the LSI-R (SRI) was administered to the provincial sample. Post-administration, the SRI was found to have a number of technical scoring problems. In



an effort to resolve the difficulties, the scale developer was contacted who acknowledged that a number of scoring difficulties have been previously reported by other researchers. The SRI was included in this study based on the reported similarities between the LSI-R and SRI. Motiuk (1992) reported a relationship between scores attained on the rated and self-report versions of the LSI-R. As reported above, the rated version of the LSI-R has an established relationship with the PCL-R ( $r=.78$  range). A preliminary data analysis of the provincial sample revealed that the SRI was not demonstrating the expected relationships with the comparison measures (e.g., PCL-R, GSIR) that have been previously established using the LSI-R. Based on this finding and the scoring difficulties, for the federal sample, the SRI was replaced by the LSI-R which has been validated in federal inmate populations (e.g., Andrews, 1995, Motiuk, 1993, Serin, 1996).

The exclusion of the SRI from the self-report battery also allowed for the use of the full scale IBS for the federal sample without extending the required administration time for the self-report battery. Part III adds the Impression Management, Passive Aggressiveness, Conflict Avoidance, Dependency, and Shyness scales to the assessment. In addition, the reliability of the scale is enhanced by implementing the full scale.

Following administration of the self-report measures and completion of the semi-structured interview, subjects were thanked for their participation. Participant's questions were addressed at the time of the interview or at a subsequent appointment. Requested feedback concerning the project was also provided at a subsequent appointment.

Prior to data collection, researchers were trained to administer and rate all measures and to conduct the semi-structured interview. Three practice cases were used for both the provincial and federal sample to facilitate this process. All raters had previously received formal training in administering the PCL-R. One rater remained constant for the two samples. A male acted as a second rater for the provincial sample. A female, similarly trained, acted as the second rater for the federal sample.

Order effects were controlled by randomization of the administration sequence of the questionnaires. For each subject, a random number table was used to determine the administration order of the self-report questionnaires. To minimize researcher bias, administration of the self-report sets was divided between the two raters. Rater 1 administered the self-report measures to one-half of the subjects and Rater 2 administered them to the other half. In order to optimize motivated participation, administration of the self-report measures preceded the semi-structured interview for all subjects.

The rated measures (i.e., the GSIR, LSI-R, PCL-R, and VRS) were scored independently by each rater. The interrater reliability of the VRS was assessed for all 60 subjects. The interrater reliabilities of the LSI-R, PCL-R, and GSIR have been previously established (see psychometric evaluations - Appendix E). Therefore, in the present study, concordance between the two raters for these measures was confirmed by assessing the interrater reliability on 20% of the sample. Rating order bias was minimized by using a random number table to determine the rating sequence (i.e., 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, or 4<sup>th</sup>) for each of the rated measures. The same rating order sequence was used for both raters.

## Measures

The convergent validity of the VRS was established by evaluating the relationships between the VRS and measures that assess various dimensions of the construct of violence. (Refer to Appendix E for psychometric reviews of the instruments and the Reference section for publication information).

### Aggression Questionnaire (Buss & Perry, 1992)

The Aggression Questionnaire (AQ) is the revised version of the Buss and Durkee Hostility Questionnaire (one of the most frequently used questionnaires on aggression, with 242 citations in the Social Science Citation Index between 1960 and 1989). The Aggression Questionnaire is a 19 item self-report questionnaire scored on a 5-point Likert scale ranging from 0-"Not at all descriptive" to 5-"Very descriptive". Williams et al. (1996) report a two-factor structure for the AQ in offender populations. One factor represents Physical Aggression and Anger and the second factor consists of Verbal Aggression and Hostility. The relationship between AQ and VRS scores was evaluated in the present study.

### Criminal Career Profile (Wong and Templeman, 1996)

Violent criminal history is one of the most commonly used indices of past violence among offender populations. However, the criminal careers of repeat offenders are multi-faceted chronologies. The variety, density and severity of violent and nonviolent offending varies greatly from offender to offender, as do the dispositions of the convictions (Wong & Templeman, 1996). Even violent offenses

that are designated in the same category (e.g., assault), can vary widely with respect to the seriousness of the act and degree of violence (Wong & Templeman, 1996). Research suggests that past violence is an important predictor of future violence (e.g., Gendreau et al., 1996; Monahan, 1996). Therefore, an accurate assessment of past violence is essential in violence-risk prediction. Wong and Templeman (1996) demonstrated that two variables, the slope and the Age of First Conviction (AFC) derived from the offender's criminal career profile are valid measures of the offender's history of criminal violence. Using information contained in criminal records, Criminal Career Profiles (CCP) were constructed by plotting the offender's successive lengths of time free and time incarcerated on the x- and y-coordinates of a Cartesian plane. The slope of the CCP and the AFC were highly correlated with the number of past violent convictions and Psychopathy Checklist-Revised scores. In addition, stepwise multiple regression analyses indicated that, for previous violent offenses, the slope of the CCP and AFC together accounted for more of the variance than the PCL-R (Wong & Templeman, 1996). Violent offenders as a group have larger slopes and younger AFCs relative to non- (or less) violent offenders. In the present study, the relationship between VRS scores and the CCP slopes and AFCs (generated by official criminal records) was assessed.

#### Interpersonal Behavior Survey (Mauger & Adkinson, 1980)

The Interpersonal Behavior Survey (IBS) is a self-report measure developed to distinguish assertive behaviors from aggressive behaviors. The IBS full scale consists of 272 items in a T-F format and comprises four sub-scales: a validity scale (3-sub-

scales) aggressiveness (7-sub-scales), assertiveness (8-sub-scales), and relationship styles (3 sub-scales). Parts I and II (items 1 through 133) provide short scales measuring a wide variety of assertive and aggressive behaviors. They are approximately one-half as long as Parts I through III (items 1 through 272), but sample almost as many behavioral subclasses. The short-form is useful in studies when administration time is limited. The relationship between the assertiveness and aggressive sub-scales of the IBS and the VRS was evaluated.

#### The General Statistical Information on Recidivism Scoring System (Nuffield, 1982)

The General Statistical Information on Recidivism Scoring System (GSIR) comprises 15 static risk indicators (e.g., Current offense, Age at admission, Number of Previous Imprisonments) which can be rated based on information in institutional files. Although the GSIR scale has demonstrated efficacy in the prediction of recidivism, this success has been limited to general or non-violent recidivism. It was expected that the GSIR total scores would be most strongly correlated with the Static Factor section of the VRS.

#### Level of Service Inventory - Revised (Andrews & Bonta, 1995)

The Level of Service Inventory (LSI-R) is an assessment instrument designed to assess the risk and need areas of offenders and to aid case managers in supervision decisions. The current version of the LSI-R comprises 54 dynamic and static items denoting specific risk variables grouped into 10 subcategories (i.e., criminal history, education/employment, financial, family/marital, accommodations, leisure/recreation, companions, alcohol/drug problems, emotional/personal, attitudes/orientation). Individual items are scored (using both interview and file information) in a binary

format indicating the absence or presence of the risk/need variable (with "1" and "0" indicating the presence or absence respectively, of the attribute described by the item). When summed, higher scores reflect a greater risk of recidivism and need for clinical intervention. Psychometric evaluations of the LSI-R indicate that the underlying construct measured by the LSI-R is a "propensity for rule violation". The relationship between scores attained on the LSI-R and the VRS was assessed in the federal sample.

#### Level of Service Inventory - Self-Report Inventory (Motiuk, 1992)

The items in the LSI-R were used to generate the Self-Report Inventory (SRI) which comprises self-referent statements regarding personal attributes and situations. The majority of the 78 SRI items are in a true-false or yes-no format with the remainder requiring a specific response. The SRI questionnaire is scored in a 0 to 1 format. In the present study, the relationship between the VRS and SRI was investigated in the provincial sample.

#### Psychopathy Checklist-Revised (Hare, 1991)

The Psychopathy Checklist-Revised (PCL-R) is a rating scale designed for the assessment of psychopathy in forensic and psychiatric populations. The PCL-R measures both behavior and personality traits that are characteristic of the criminal psychopath. The scale contains 20 items that are rated on a 3-point ordinal scale based on information obtained from a file review and the administration of a semi-structured interview. To date, the PCL-R has demonstrated the highest predictive validity with respect to violent recidivism ( $r=.28$ , Serin, 1996). The relationship between VRS total scores and PCL-R Factor 1, Factor 2, and total scores was evaluated.

### **Data Analyses**

The data were analyzed by SPSS-PC (6.0) and CrimeWare (Templeman, 1997), a software program to compute the slope for the Criminal Career Profile.

The interrater agreement of the VRS was assessed based on the degree of concordance between the scores of two raters. Item by item concordance and the total score correlations between the two raters was evaluated using the Kappa coefficient of agreement statistic and Pearson's Product Moment correlation coefficient. The overall concordance of the scale was also evaluated using the Kappa ( $\kappa$ ) and Pearson's ( $r$ ) coefficients based on all pairs of ratings ( $n=1651$ ).

The internal consistency of the VRS was investigated using *Cronbach's coefficient alpha*.

The convergent validity of the VRS (Part A) was assessed by correlating VRS scores with validated measures of different dimensions of criminal violence (i.e., PCL-R scores, and the age of first conviction (AFC) and slope generated using the Criminal Career Profile analysis), criminal risk/need (i.e., GSIR and LSI-R total scores), and aggression (i.e., scores on the AQ and IBS). The correlation between VRS total scores and the number of violent and non-violent convictions was assessed using the Pearson Product Moment Correlation coefficient. The relative postdictive criterion validity of the VRS was assessed using a stepwise multiple regression analyses. VRS total scores and scores attained on the comparison measures were used as independent variables. The dependent variables were the number of violent and non-violent convictions obtained from official criminal records.

## Chapter 4

### RESULTS

The objective of this study was to assess the interrater reliability, internal consistency and convergent validity of the VRS in a sample of provincial and federal inmates. This chapter contains the results from the data analyses for the provincial, federal, and combined samples.

#### Sample Characteristics

##### Provincial Sample

The provincial sample was comprised of 30 male offenders incarcerated at the Saskatoon Correctional Centre. Of the original 30 recruited subjects, two were released prior to the completion of data collection and three voluntarily withdrew from the study and were replaced by new volunteer participants. All subjects had received sentences of less than two years. Most of the subjects were participating in, or had previously received, some form of short-term institutional treatment for substance abuse.

The mean age was 29.2 years with a standard deviation of 7.4 years. The sample comprised both Aboriginal (87%) and Caucasian (13%) offenders. The type of convictions in this sample ranged from non-violent to violent and were determined using the CPIC records. The provincial subjects had received a total of 545 non-violent



convictions and 86 violent convictions. Non-violent offenses were defined as physically non-injurious and included property-related, prior-release violations, and alcohol/substance-use related offenses. Violent offenses included assault convictions (e.g., common assault, assault causing bodily harm, aggravated assault) weapons convictions, robbery, and uttering threats (Wong, 1996).

### Federal Sample

The federal sample comprised 30 male offenders incarcerated at the Regional Psychiatric Centre (Prairies), the RPC. All 30 subjects had been transferred to the RPC from medium, maximum and super maximum, institutions in the Prairie region to participate in intensive treatment programs. At the time of data collection, approximately half of the subjects were nearing the end of their treatment and the rest were in the early phase of treatment.

One-half of the federal sample (n=15) was participating in the Aggressive Behavioral Control program, an intensive six-month program for violent offenders who have committed two or more violent offenses in the community or have demonstrated serious misconduct while incarcerated. The other half of the federal sample (n=15) was participating in the Clearwater Sexual Offender program, an intensive four to five-month program for high-risk sexual offenders.

Of the original 30 recruited subjects in the federal sample, one subject was used as a practice subject (and therefore not included in the data analysis), two subjects were involuntarily discharged due to violent and/or disruptive behavior while in the treatment program, one subject was transferred to another unit for psychiatric care, and

one subject voluntarily withdrew from the study. Five additional subjects were recruited to replace the dropouts.

All federal subjects had received sentences of more than two years. A high proportion of the subjects had previously participated in a variety of institutional treatment programs (e.g., substance abuse programs, anger management, sexual offender programs).

The mean age was 35.3 years with a standard deviation of 8.9 years. The sample comprised Aboriginal (43.3%), Caucasian (46.6%), and a mixed group (10.1%) of offenders. The federal subjects had received a total of 228 convictions for non-violent offenses and 115 violent convictions. Federal offenders have (more serious) and a larger number of violent offenses than the provincial offenders.

The study was designed to sample a wide range of offenders. Rather than sampling only federal offenders at the RPC, both provincial and federal offenders were included to ensure that findings could not be attributed to the sampling of a highly selected group of offenders admitted to the RPC for treatment purposes. The results indicate that collectively, the provincial and federal samples comprised a group of offenders with a range of criminal and demographic characteristics.

### **Descriptive Statistics of the VRS and Comparison Measures**

#### **Descriptive Statistics of the Violence Risk Scale (VRS)**

Due to the more serious nature of offenses mandating a minimum two-year sentence, it was expected that the federal sample would evidence higher scores on the VRS and comparison measures relative to the provincial sample. Table 4.1 presents the

VRS scores for the two samples. The Dynamic Factor scores were found to be significantly higher in the federal sample.

**Table 4.1**  
**Means and Standard Deviations of the VRS for the Provincial, Federal, and Combined Samples**

<b>VRS Scores</b>	<b>Provincial (n=30)</b>	<b>Federal (n=30)</b>	<b>Combined (n=60)</b>	<b><i>t</i><sup>2</sup></b>
Total Static Factor Scores	11.8 (3.1) <sup>1</sup>	10.6 (3.8)	11.2 (3.5)	ns
Total Dynamic Factor Scores	32.7 (10.0)	39.3 (9.2)	36.0 (10.1)	-2.6
Total VRS Scores	45.8 (12.3)	51.4 (12.3)	48.6 (12.5)	ns

<sup>1</sup> Standard Deviation in parenthesis

<sup>2</sup> *t* comparison between provincial and federal samples

\*\*p≤.01

The VRS Static Factor, Dynamic Factor and Total Scores for both the provincial and federal sample were normally distributed according to the Shapiro Wilks' Test for Normal Distribution (see Table 4.2). The Shapiro Wilks' Test is a statistical test of the null hypothesis that the data are from a normal distribution. The test shows good power in many situations compared to others tests of normality (Conover, 1980). As expected, there were no significant differences, therefore the null hypothesis is not rejected.

**Table 4.2****Shapiro-Wilks' Test of Normality for Distribution of VRS Scores**

<b>VRS Scores</b>	<b>Shapiro-Wilks' Statistic</b>	<b>Df</b>	<b>Significance</b>
Static Factor Scores - Provincial	.94	30	.14
Static Factor Scores - Federal	.95	30	.31
Dynamic Factor Scores - Provincial	.96	30	.47
Dynamic Factor Scores - Federal	.95	30	.20
VRS Total Scores - Provincial	.95	30	.25
VRS Total Scores - Federal	.96	30	.47

**Descriptive Statistics for the AQ, GSIR, and PCL-R**

Table 4.3 presents the mean scores for the AQ, GSIR, and the PCL-R, for the federal, provincial and combined samples. There were no significant differences between the provincial and federal sample means on any of the measures. Where available, the normative sample means and standard deviations are also presented for comparison. The means scores of the AQ and the PCL-R approximate the normative sample means. Further information concerning the normative samples for each comparison measure is also presented in Appendix E.

**Table 4.3****Mean Scores and Standard Deviations for the AQ, GSIR, and PCL-R for the Provincial, Federal and Combined Samples**

<b>Measure Total Score</b>	<b>Provincial (n=30)</b>	<b>Federal (n=30)</b>	<b>Combined (n=60)</b>	<b>Normative<sup>1</sup></b>	<b><i>t</i></b>
AQ	90.2 (18.9) <sup>2</sup>	80.3 (22.2)	85.3 (21.0)	77.8 (16.5)	ns
GSIR	-5.1 (6.8)	.8 (8.9)	-2.1 (8.4)	n/a <sup>3</sup>	ns
PCL-R	17.9 (8.2)	18.9 (7.7)	18.4 (7.9)	23.6 (7.9)	ns

<sup>1</sup> The normative sample sizes differed among the measures. Refer to Appendix E for further information.

<sup>2</sup> Standard Deviation in parentheses

<sup>3</sup> The GSIR was normed in terms of risk categories. Refer to Appendix E for further information concerning the scale categories.

**Descriptive Statistics for the IBS**

The full-scale IBS was not administered to the provincial sample. The combined sample sub-scale scores for the IBS and normative sample sub-scale means are presented in Table 4.4. There is no composite score for the IBS. For the combined sample, the mean scores of a number of the subscales tend to be lower than that of the normative sample, for example Impression Management, Expression of Anger, Physical Aggression, Passive Aggressiveness. Further analyses revealed differences between the provincial and federal sample means that accounted for much of the discrepancy. The IBS short version does not contain the IM, PA or SH subscales and

combining the two samples thus resulted in an overall lower mean score for the study sample. When the two samples were evaluated separately, the federal sample scores more closely approximated the normative sample means (e.g., Impression Management, 13.7; Passive Aggressiveness 10.47; and Shyness, 9.8). For the Expression of Anger and Physical Aggression subscales, the federal sample means were significantly higher than the provincial sample means (i.e., EA=8.0 versus EA=2.6 and PH=4.4 versus PH=2.4). As expected, all of the assertiveness and relationship subscale means for the combined sample were lower than the normative sample means suggesting that the study sample was characteristically less assertive than the normative sample.

**Table 4.4****Means and Standard Deviations for the IBS - Combined Sample<sup>1</sup>**

<b>IBS Sub-scales (n=59)<sup>2</sup></b>	<b>Combined Sample</b>	<b>Normative Sample</b>
<b>Validity Scales</b>		
Denial (DE)	3.2 (2.0)	2.8 (1.8)
Infrequency (IF)	2.4 (2.0)	1.9 (2.1)
Impression (IM)	8.1 (6.8)	13.5 (4.0)
<b>Aggressiveness Scales</b>		
General Aggressiveness (GGR)	13.9 (7.1)	13.8 (6.0)
Hostile Stance (HS)	9.2 (4.6)	9.6 (4.0)
Expression of Anger (EA)	5.4 (4.8)	7.2 (4.6)
Disregard for Rights (DR)	2.5 (1.9)	3.4 (2.1)
Verbal Aggressiveness (VE)	4.4 (2.6)	5.0 (2.7)
Physical Aggressiveness (PH)	3.4 (2.2)	4.5 (2.4)
Passive Aggressiveness (PA)	5.3 (7.0)	10.3 (6.2)
<b>Assertiveness Scale</b>		
General Assertiveness (GR)	31.3 (11.9)	35.5 (9.9)
Self-Confidence (SC)	8.2 (4.5)	10.4 (3.5)
Initiating Assertiveness (IA)	9.2 (3.7)	9.9 (3.6)
Defending Assertiveness (DA)	12.1 (3.9)	12.5 (3.6)
Frankness (FA)	5.8 (2.7)	7.0 (2.6)
Praise (PR)	4.4 (2.3)	5.9 (2.1)
Requesting Help (RE)	2.3 (2.0)	4.6 (1.9)
Refusing Demands (RF)	3.2 (1.8)	3.9 (1.3)
<b>Relationship Scales</b>		
Conflict Avoidance (CA)	8.2 (5.0)	9.6 (4.1)
Dependency (DP)	4.0 (5.5)	9.5 (4.5)
Shyness (SH)	5.0 (6.8)	5.5 (5.9)

<sup>1</sup> The IBS Short-Version and Full Version were merged for the combined sample.<sup>2</sup> One subject did not complete the IBS.

The self-report version (SRI) of the Level of Service Inventory-Revised was administered to the provincial sample. As described in the Procedures section, the LSI-R rated version was used for the federal sample. The mean scores and standard deviations are presented in Table 4.5. The mean scores of the LSI-R are much higher in the federal sample than in the normative sample. The mean LSI-R score for the federal sample exceeds the 98th percentile for the normative sample.

**Table 4.5**

**Means and Standard Deviations for the Level of Service Inventory-Revised**  
**Provincial and Federal Samples**

<b>Measure</b>	<b>Study Sample</b>	<b>Normative Sample</b>
SRI - Provincial	28.2 (6.0)	23.7 (9.9)
LSI-R - Federal	48.6 (3.2)	26.9 (9.1)



### **Interrater Reliability of the VRS**

The interrater reliability of the VRS was assessed based on the degree of concordance between the scores of two raters on each item and on the VRS total score correlations.

It was predicted that there would be 'good' interrater agreement (Kappa statistic) for each of the 29 items ( $k \geq .40$ , Fleiss, 1981) and that the total score correlations would be significant at the .05 level (Pearson Product Moment Coefficient).

#### **Interrater Reliability of the VRS (Item by Item Analyses)**

The Kappa and Pearson correlation coefficients for each item for the provincial and federal sample were computed and are presented in Table 4.6.

**Table 4.6**

**VRS Item by Item Kappa and Pearson Correlation Coefficients**  
**Provincial and Federal Samples**

VRS Items	Kappa (k) Statistic		Pearson (r) Coefficient	
	Provincial	Federal	Provincial	Federal
<b>Static Factors</b>				
S1. Current Age	1.0	.77	1.0**	.80**
S2. Age at First V. Conviction	.43	n/c	.60**	.66**
S3. No. of Juvenile Convictions	.50	.55	.64**	.79*
S4. Violence through Lifespan	n/c	.37	.50**	.66**
S5. Release Failures/Escapes	.48	.46	.84**	.75**
S6. Stability of Upbringing	.15	.65	.59**	.88**
<b>Dynamic Factors</b>				
D1. Violent Lifestyle	.28	.30	.64**	.59**
D2. Criminal Personality	.36	.12	.65**	.62**
D3. Criminal Attitudes	.33	.49	.51**	.77**
D4. Work Ethic	.37	.25	.65**	.56**
D5. Respect for Authority	.37	.38	.58**	.74**
D6. Criminal Peers	.20	.55	.50**	.74**
D7. Interpersonal Aggression	n/c	.20	.59**	.33
D8. Emotional Disinhibition	.20	n/c	.68*	.38*
D9. Violence During Incarceration	.19	.56	.45*	.87**

Table continues...

**Table 4.6 (... continued)****VRS Item by Item Kappa and Pearson Correlation Coefficients in Provincial and Federal Samples**

<b>VRS Items</b>	<b>Kappa (k) Statistic</b>		<b>Pearson (r) Coefficient</b>	
	<b>Provincial</b>	<b>Federal</b>	<b>Provincial</b>	<b>Federal</b>
D10. Weapon Use	.43	.39	.79**	.68**
D11. Insight into Violence	n/c	.17	.50**	.42*
D12. Mental Disorder	1.0	.82	1.0**	.89**
D13. Substance Abuse	.14	.60	.70**	.91**
D14. Stability of Relationships	.12	.25	.60**	.72**
D15. Community Support	.17	n/c	.40*	.36
D16. Released to High Risk Sit.	.19	n/c	.27	.29
D17. Violence Cycle	.19	.24	.61**	.46*
D18. Impulsivity	n/c	n/c	.06	-.05
D19. Violent Sexual Behavior	n/c	.89	.44*	.98**
D20. Compliance with Supervision	.33	.37	.80**	.79**
D21. Stages of Change	.30	.22	.58**	.44*
D22. Security of Release Instit.	n/c	n/c	1.0**	.31
D23. Longest Violent Free	n/c	.38	.49	.53**

Note. n/c = Kappa statistic could not be computed for items where row values did not equal column values

\*  $p \leq .05$

\*\* $p \leq .001$

The Pearson coefficients generally demonstrated a stronger relationship between the two raters than did the Kappa coefficient. The Kappa coefficient is a chance-corrected statistic that provides a measure of complete concordance. Unlike the Pearson coefficient that evaluates the relationship between the scores of the two raters, the Kappa statistic is computed based on pairs of ratings that are in complete agreement and, therefore, it is a very stringent measure of concordance. The characteristics of the Kappa statistic are examined more fully in the discussion session.

To assess the item by item interrater reliability for the combined sample, the relationship between the scores of Rater 1 and the averaged scores of the second rater for the provincial and federal samples was evaluated. For the combined sample, there was 'good' ( $k \geq .40$ ) concordance for 11 of the items (S1, S2, S3, S5, D3, D9, D10, D12, D19, D21, D22) and 'fair' ( $k > .30$  to  $.39$ ) concordance for an additional 7 items (S4, S6, D4, D5, D6, D13, D20). There was 'poor' concordance ( $k \leq .30$ ) for 11 items (D1, D2, D8, D11, D14, D15, D16, D17, D18, D23). The range of scores for the two raters was not equal for item D7, therefore a Kappa value was not computed.

To assess the degree of difference in scoring between the raters for each item (i.e., the scores of Rater 1 - Rater 2), a post-hoc analyses for the combined sample was computed. The maximum difference between the ratings for each item was +/- 3 scale points. For example, Rater 1 scored "0" and Rater 2 scored "3" or Rater 1 scored "3" and Rater 2 scored "0". The item by item interrater reliability and the Kappa ( $k$ ) and Pearson ( $r$ ) coefficients for the combined sample are presented in Table 4.7

Table 4.7

**Interrater Reliability of the VRS and Degree of Difference Between Raters in the Combined Sample**

Range of Difference R2-R1 <sup>2</sup> = ----->	3	2	1	Direct Hits <sup>3</sup> 0	-1	-2	-3	% Direct Hits	Pearson	Kappa <sup>4</sup>
<b>Static Factors</b>										
S1. (n=60)										
Current Age		3	2	55				92	.89**	.88
S2. (n=55)										
Age First V. Conv.		1	4	39	9	2		71	.63**	.55
S3. (n=60)										
No. of Juv.Conv.		3	4	43	8	1	1	71	.73**	.56
S4. (n=60)										
Violence Through Lifespan		3	13	36	8			55	.55	.33
S5. (n=51)										
Prior Release		1	9	33	6	2		64	.80**	.47
Failures/Escapes										
S6. (n=60)										
Stability of Family Upbringing		4	10	33	12	1		55	.72**	.39
<b>STATIC FACTOR<sup>1</sup></b>										
OVERALL	0	15	42	239	43	6	1	69	.87**	.50

Table continues...

<sup>1</sup> For the Static Factors there were 346 pairs of ratings ( 6 x 60 less omitted items).

For the Dynamic Factors there were 1305 pairs of ratings (23 x 60 less omitted items).

<sup>2</sup> 'Range of Difference' refers to the absolute difference between the scores of the two raters.

<sup>3</sup> 'Direct Hits' refers to instances where the score of Rater 1 = score of Rater 2.

<sup>4</sup> Kappa statistic could not be computed (nc) for items where row values did not equal column values.

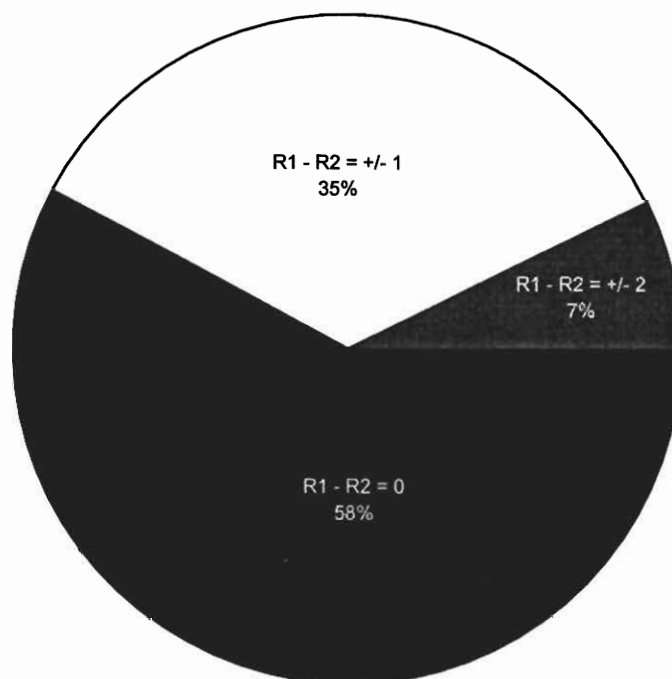
\* p≤.05

\*\*p≤.01

Table 4.7 continued

Range of Difference R2-R1 <sup>2</sup> = ----->	3	2	1	Direct Hits <sup>3</sup>	-1	-2	-3	%Direct Hits	r	k
<b>Dynamic Factors</b>										
D1. (n=60) Violent Lifestyle		3	9	30	18			50	.62**	.29
D2. (n=60) Criminal Personality			5	29	16	10		48	.66**	.27
D3. (n=60) Criminal Attitudes			9	36	13	2		60	.62**	.40
D4. (n=60) Work Ethic		3	20	30	6	1		50	.59**	.30
D5. (n=60) Respect for Authority		1	4	33	12	10		55	.64**	.38
D6. (n=60) Criminal Peers		2	3	33	16	6		55	.62**	.39
D7. (n=60) Interpersonal Aggression		1	4	27	22	6		50	.48**	nc
D8. (n=60) Emotional Disinhibition			7	23	28	2		38	.57**	.11
D9. (n=60) Violence Incarceration		1	11	34	11	2	1	56	.71**	.41
D10. (n=60) Weapon Use		3	11	35	10		1	58	.74**	.43
D11. (n=56) Insight Into Violence		1	10	26	16	3		46	.49**	.22
D12. (n=59) Mental Disorder		1		58				98	.90**	.87
D13. (n=60) Substance Abuse			7	32	18	3		53	.75**	.31
D14. (n=48) Stability of Relationships			4	21	21	2		44	.66**	.19
D15. (n=59) Community Support		2	11	24	18	4		39	.38**	.13
D16. (n=47) Released Back High Risk		3	9	19	13	3		40	.30*	.13
D17. (n=58) Violence Cycle		1	11	29	15	2		50	.55**	.22
D18. (n=60) Impulsivity		2	16	24	13	4		40	.04	.09
D19. (n=57) Violent Sexual Behavior			2	51	2	1	1	89	.90**	.79
D20. (n=46) Compl.with Supervision		1	6	25	13	1		54	.72**	.35
D21. (n=58) Stages of Change			11	33	14			57	.51**	.27
D22. (n=51) Security Release Instit.			3	47	1			92	.89**	.85
D23. (n=46) Longest Violent -Free		1	5	21	13	5	1	46	.49**	.27
<b>DYNAMIC F.OVERALL</b>	0	26	178	720	309	67	5	58	.82**	.39

The total number of pairs of ratings for Rater 1 and Rater 2 was calculated ( $n=1651^1$ ). An analysis of the percentage of pairs of ratings that differed by 0, 1, and 2 or more scale points was also conducted and is presented in Figure 4.1.



**Figure 4.1**  
**Percentage of interrater agreement across the total number of ratings ( $n=1651$ ) expressed as a function of the difference between the scores of Rater 1 (R1) and Rater 2 (R2).**

---

<sup>1</sup> For the Static Factors there were 346 pairs of ratings (6 x 60 less omitted items).  
For the Dynamic Factors there were 1305 pairs of ratings (23 x 60 less omitted items).

The total VRS static and dynamic scores have almost 60% direct hits, highly significant Pearson Correlations of over .80 and Kappa values in the ‘fair to good’ range (e.g., Fleiss, 1981; Landis and Koch, 1977). Inspection of these interrater concordance measures indicates that only one item, ‘Impulsivity’ (D19), has consistently low concordance.

### **Interrater Reliability of the VRS Total Score and Comparison Measures**

Interrater reliability was assessed on the VRS total scores for all subjects for the provincial, federal, and combined sample. To confirm previously established interrater reliabilities for the comparison measures (i.e, GSIR, LSI-R, and PCL-R), two ratings were obtained for 20% of the federal and provincial samples and interrater reliability (Pearson  $r$ ) was computed. The results are presented in Table 4.8. Interrater reliability was high for all four measures.

**Table 4.8**  
**Interrater Reliabilities of the VRS, GSIR, LSI-R, and PCL-R**  
**Provincial, Federal and Combined Samples**

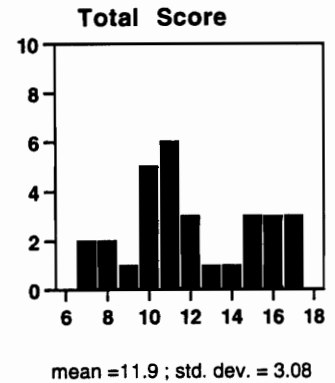
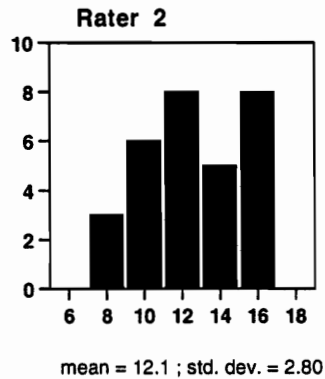
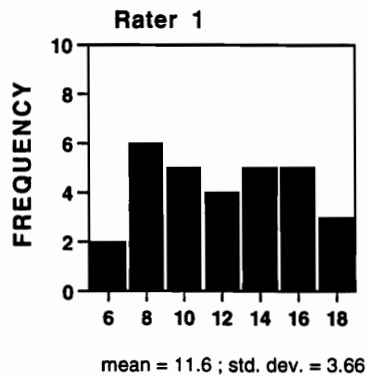
<b>Measure</b>	<b>Provincial</b>	<b>Federal</b>	<b>Combined</b>
VRS Total Score	.86** (n=30)	.89** (n=30)	.87** (n=60)
GSIR	.93** (n=6)	.95** (n=6)	.95** (n=12)
LSI-R	Not Applicable	.93** (n=6)	Not Applicable
PCL-R	.83* (n=6)	.70 (n=6)	.84* (n=12)

\* $p \leq .05$  \*\* $p \leq .01$

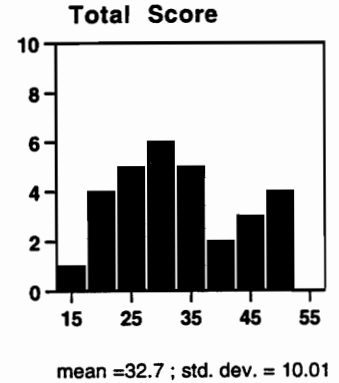
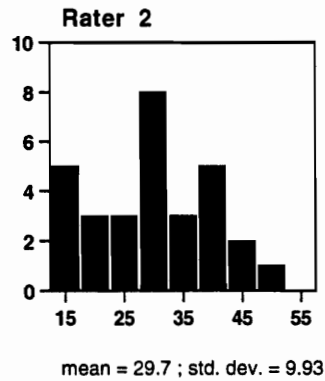
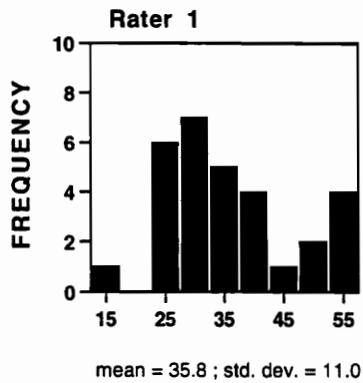


To confirm that the high interrater reliability attributed to the VRS was not the result of a restricted score range, an evaluation of the frequency distributions of the scores for the provincial, federal and combined samples was also conducted. The results are presented in Figures 4.2, 4.3, and 4.4. The distribution of the scores suggests that a wide range of scores were obtained from both the provincial and federal samples.

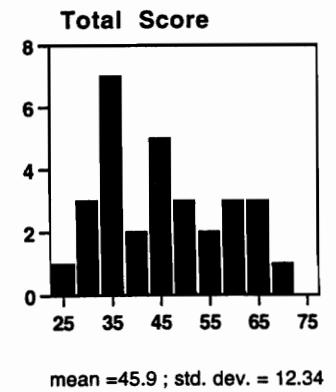
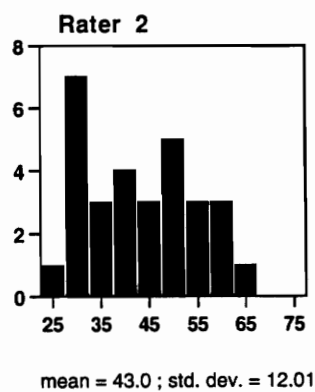
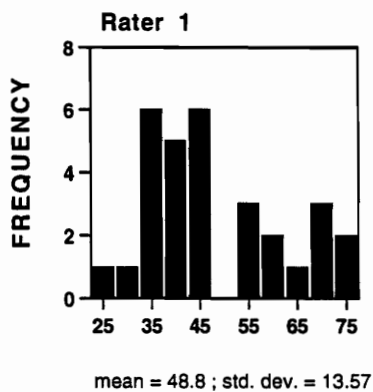
## STATIC FACTOR SCORES



## DYNAMIC FACTOR SCORES

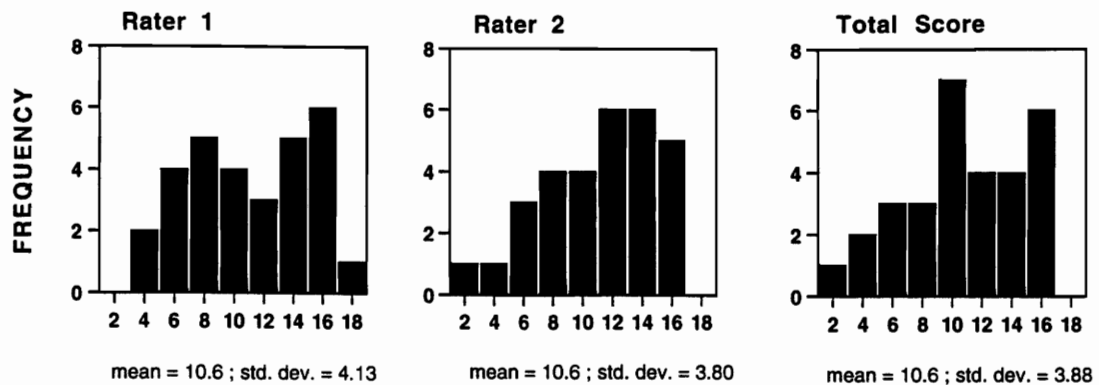


## VRS TOTAL SCORES

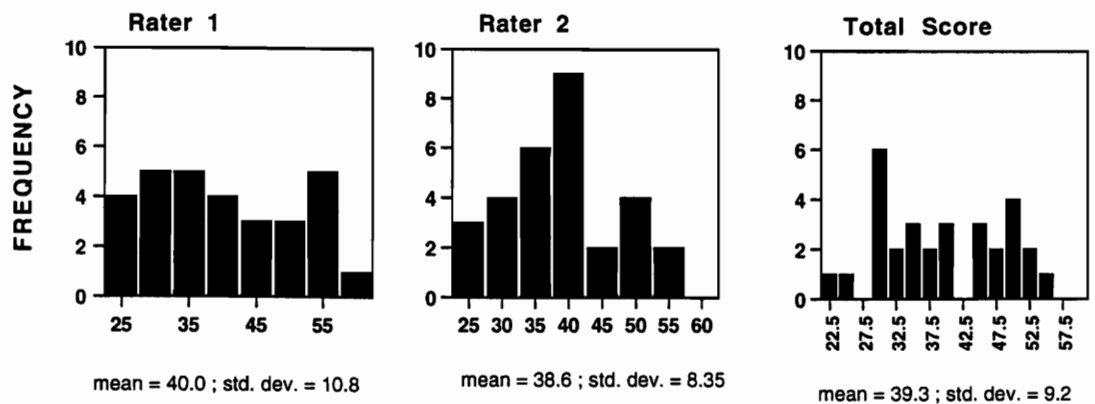


**Figure 4.2**  
Frequency Distributions of VRS Scores - Provincial Sample

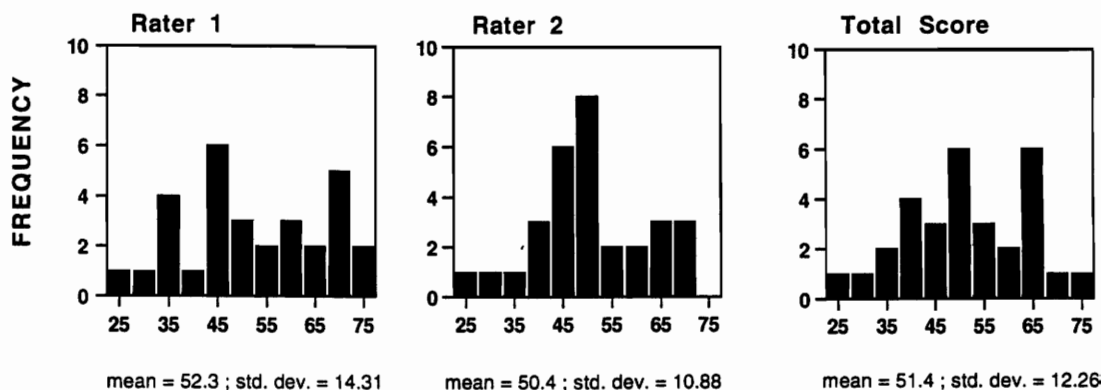
## STATIC FACTOR SCORES



## DYNAMIC FACTOR SCORES

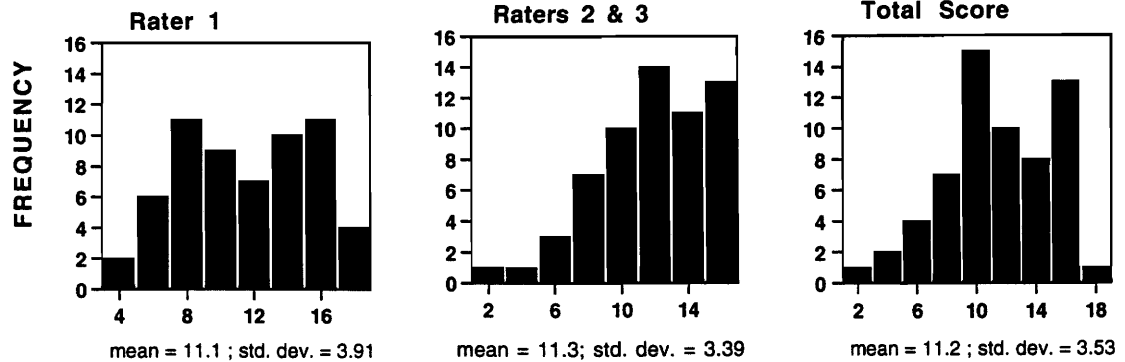


## VRS TOTAL SCORES

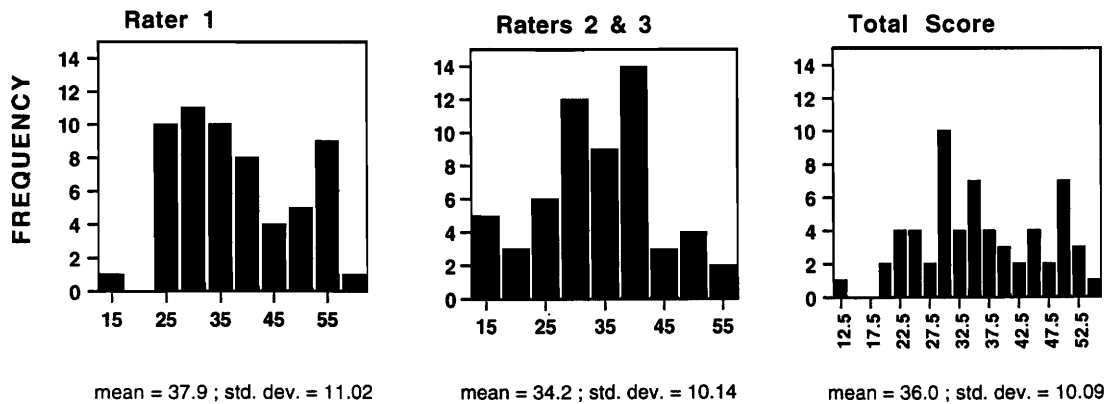


**Figure 4.3**  
Frequency Distributions of VRS Scores - Federal Sample

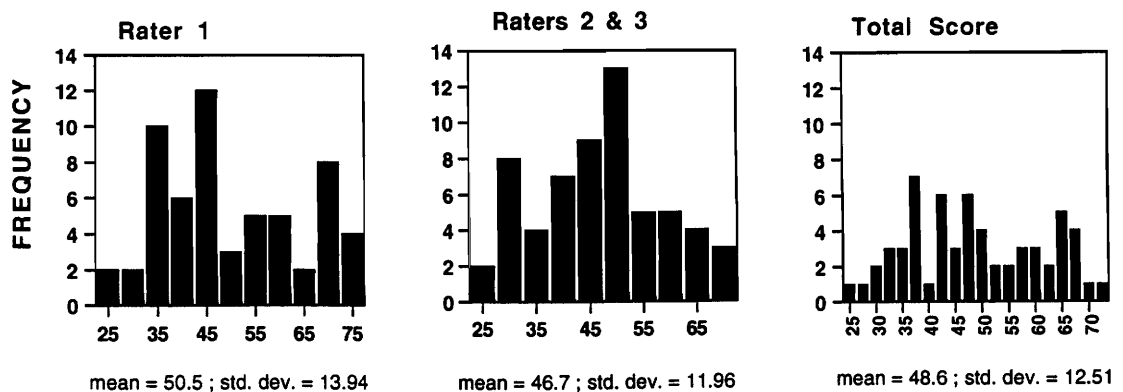
## STATIC FACTOR SCORES



## DYNAMIC FACTOR SCORES



## VRS TOTAL SCORES



**Figure 4.4**  
Frequency Distributions of VRS Scores - Combined Sample

### **Internal Consistency of the VRS**

To assess the extent to which the VRS items are measuring the same or a similar construct, an analysis of the internal consistencies of the scale items for the combined sample (n=60) was conducted using the Cronbach's alpha. A very high alpha coefficient of .92 was obtained suggesting that the scale is highly internally consistent. To assess the presence of weaker items in the scale, that is, items that did not tend to contribute to the overall internal consistency of the scale, the alpha coefficients of the VRS were computed after each item was deleted from the scale. Deletion of a weaker item from the scale should lead to a fairly large increase in the alpha coefficients computed using the remaining items. Such analyses were done using the full scale (29 items) and a split half analyses of odd and even numbered items. The already large full scale alpha (.92) produces a ceiling effect that makes further increases in alpha values difficult to detect. With split-half analyses which give lower alpha values because of smaller numbers of items, the increase in alpha when a weaker item was deleted was easier to detect. Items D12 (Mental Illness) and D19 (Violent Sexual Behavior) were the only slightly weaker items identified in the analyses.

To confirm that the high alpha coefficient was not the result of having many redundant scale items in the VRS that are measuring the same construct, item to item correlational analyses were conducted. The redundancy criterion proposed by Clark and Watson (1995) of 0.6 (Pearson  $r$ ) was employed. Items that correlate higher than 0.6 are considered to be redundant as they tend to measure the same thing. Of the 29

items, only three pairs of items exceeded the 0.6 cut-off criterion. The overlapping scale items were Violence throughout the Lifespan (S4) with Violent Lifestyle (D1) ( $r=.69$ ); Violence Lifestyle (D1) with Violence Cycle (D17) ( $r=.62$ ); and Criminal Attitudes (D3) with Respect for Authority and Social Convention (D5),  $r=.81$ ).

To rule out the possibility that the high alpha coefficient could be the result of the relatively large number of VRS scale items, the split-half reliability analysis was conducted by computing alpha coefficients using odd and even numbered scale items. The Cronbach's Alpha coefficient was used for the full scale and split-half internal consistencies analyses. A high alpha value was maintained using only odd (.82) or even (.86) numbered items. The above results are presented in Table 4.9.

**Table 4.9****VRS Full Scale and Split-Half Alpha Coefficients<sup>1</sup> - Combined Sample<sup>2</sup>**

ITEM	Full Scale	Split-Half Odd Items	Split-Half Even Items
PRES1	.9211	.8206	
PRES2	.9195		.8579
PRES3	.9229	.8172	
PRES4	.9169		.8474
PRES5	.9209	.8096	
PRES6	.9205		.8613
AD1	.9148	.7830	
AD2	.9162		.8431
AD3	.9151	.7856	
AD4	.9175		.8518
AD5	.9157	.7880	
AD6	.9164		.8475
AD7	.9167	.7909	
AD8	.9213		.8655
AD9	.9172	.7991	
AD10	.9162		.8436
AD11	.9213	.8093	
AD12	.9231		.8700
AD13	.9194	.8050	
AD14	.9189		.8557
AD15	.9190	.8017	
AD16	.9164		.8472
AD17	.9171	.7957	
AD18	.9185		.8543
AD19	.9308	.8505	
AD20	.9163		.8481
AD21	.9190	.8009	
AD22	.9213		.8641
AD23	.9192	.8029	
ALPHA COEFFICIENT <sup>3</sup>	.9216	.8153	.8635

<sup>1</sup> Alpha Coefficient if item deleted; <sup>2</sup> n=60; <sup>3</sup> Alpha Coefficient with no item deleted.

### **Validity of the VRS**

The convergent validity of the VRS was evaluated by correlating VRS total scores with scores obtained on the comparison measures. The relationship between VRS scores and the offender's criminal history was also investigated (see Table 4.10).

VRS total scores were significantly correlated with the scores obtained on the GSIR, LSI-R, and PCL-R; 4 of 7 of the aggression sub-scales (i.e., general aggressiveness, expression of anger, physical aggressiveness, and passive aggressiveness) of the IBS, but not with any of the IBS assertiveness sub-scales. There was no significant relationship between the VRS and AQ or between the VRS and self-report version (SRI) of the Level of Service Inventory. VRS total scores were also significantly correlated with number of violent convictions in the offender's official criminal record (FPS) but not with the number of non-violent convictions. VRS total scores were inversely related to Age at First Conviction. There was no significant relationship between VRS total scores and the slope generated by the Criminal Career Profile (CCP).

A stepwise multiple regression analysis was used to evaluate the relative contribution of the VRS, GSIR, LSI-R, IBS, and PCL-R Factor 1, Factor 2, and PCL-R total scores, to the postdiction of violent and non-violent convictions. The VRS contributed significant to the postdiction of violent convictions ( $R=.33$ ) followed by the PCL-R Factor 2 score ( $R=.44$ ). PCL-R Factor 1 and PCL-R total scores did not enter the regression equation. No other measures contributed to the postdiction of violent convictions. For non-violent convictions, the GSIR made the largest contribution to the



postdiction ( $R=.60$ ) followed by the VRS ( $R=.69$ ) which captured a small but significant amount of variance.

**Table 4.10**

**Correlations of the Violence Risk Scale with other Criminal Risk/Violence Measures**

Measure	Pearson $r$ <sup>1</sup>
AQ	.16
GSIR	.49 **
IBS - General Aggression (GGR) <sup>2</sup>	.28*
IBS - Expression of Anger (EA)	.33*
IBS - Physical Aggression (PH)	.40**
IBS - Passive Aggression (PA)	.29*
	( $n=59^3$ )
LSI-R	.83**
	( $n=30$ )
PCL-R	.78**
Postdiction Violent Convictions	.33**
Postdiction Non-Violent Convictions	-.005
Age at First Conviction (AFC)	-.28*
Criminal Career Profile (CCP)	.14

<sup>1</sup> \*\* $p < .01$ , \*  $p < .05$ ,  $n=60$  unless indicated.

<sup>2</sup> Only significant correlations of the IBS are presented.

<sup>3</sup> One subject did not complete the IBS.

## Chapter 5

### **DISCUSSION**

#### **Overview**

The present study investigated the interrater reliability, internal consistency and construct validity of the VRS.

The hypotheses that the VRS would demonstrate fair to good interrater reliability was supported. High interrater reliability was achieved in a sample of provincial and federal offenders comprising a wide range of offender characteristics, indicating that the VRS can be rated reliably by independent raters trained to administer it. The VRS is highly internally consistent. There are very few redundant items and it appears to be a unidimensional scale. The results also indicate that the VRS is a valid instrument for the assessment of risk of violence in an offender population. VRS scores were significantly correlated with validated measures of violent recidivism and criminal risk/need.

The remainder of this chapter will discuss in more detail the interrater reliability, internal consistency and validity investigations. Also examined are the limitations of the present study and suggestions for further research.

### **Interrater Reliability of the VRS**

Overall, the VRS demonstrated good interrater reliability. The total scores of the two raters were highly concordant based on Kappa and Pearson statistics. Good interrater reliability was also obtained for most of dynamic factors which require a considerable degree of clinical judgment. This suggests that the VRS can be rated reliably by independent raters who have been trained to administer it. It is noteworthy that the high interrater reliabilities were achieved using less than five practice cases.

#### **Item by Item Interrater Concordance**

Overall, the item by item interrater concordance was strong. As predicted, the obtained Kappa value using all pairs of ratings ( $n=1651$ ) was in the 'good' concordance range. Over 58% of the ratings were in complete agreement and only 7% differed by more than one scale point.

The item by item interrater concordance for the federal sample was generally higher than for the provincial sample. For example, the concordance between raters for S6, Stability of Family Upbringing, was only  $r=.15$ , ( $n/s$ ), in the provincial sample, compared to  $r=.65$  ( $p<.01$ ) for the federal sample. Other similar examples include D6, D9, and D13. For some of the items, the item by item concordance was slightly higher in the provincial sample relative to the federal sample (i.e., S1., D2., D4., D10., D12, D21).

Inadequate file information for the provincial offenders may have contributed to the lower interrater concordance of some of the items. For the provincial offenders, file information was generally historical in nature and mainly consisted of criminal offense history. Information important in evaluating the dynamic factors was often unavailable

in files (e.g., attitude, present functioning, insight, progress in treatment programs, characteristics and cycles of behavior, etc). As a result, the provincial offender ratings for the dynamic factors relied heavily upon interview-based information.

The interview was semi-structured and independently administered and thus, some differences were expected as a function of the idiosyncratic interview dynamics between the subject and the rater. Post data-collection discussions between the two raters confirmed that there were occasions when the quality and quantity of the information volunteered during the interview varied.

The effects of discrepant self-reported information is minimized when there is additional file information that challenges inconsistent or ambiguous information volunteered during the interview. When there are multi-sources of information, the relative weighting of all of the available information is considered to arrive at the final rating. As a result, the excessive reliance on a single source of information is reduced.

Relative to the provincial sample, the federal sample ratings were generally based on richer and converging sources of file information. The federal sample was participating in intensive treatment programs at the RPC. Reports on the progress of RPC program participants are generated at the beginning, interim, and at the end of treatment. As well, multi-volumed institutional files that depict the offenders criminal, personal, and institutional behaviors were available. These documents provide file information not available for provincial offenders.

However, despite the lack of information for the provincial relative to the federal sample, high interrater reliability was attained for the total sample.

Inadequate information may also have contributed to lower interrater concordance for the VRS items that address community functioning. For example, D14. Stability of Relationships; D15. Community Support; and D16. Released Back to High Risk Situations, evidenced low kappa values ( $k \leq .25$ ). Often times, information concerning community functioning is minimal if the offenders are not expecting upcoming releases, e.g. “lifers”. As a result, the rater is forced to either omit the items or base the rating exclusively on interview-obtained information. The objectivity of some of the self report information is suspect however because many offenders have been away from their community for extended periods of time. This suggests that the interrater reliability of the VRS may increase as a function of the quantity, quality, and relevancy of the available information.

The item by item interrater reliabilities were higher based on the Pearson's Product Moment coefficient than the Kappa coefficient (e.g., S3, S5, S6, D1, D2, D6, D13, D14, D20). Kappa is a chance-corrected coefficient of agreement statistic used for calculating interrater concordance of nominal data, and is a measure of the amount of full agreement (as opposed to degree of association) between two raters. It is defined as the fraction of complete agreement corrected for chance. Kappa values between 0.40 and 0.75 are considered good concordance, 0.75 and above, are considered excellent (Fleiss, 1981).

Another limitation of the kappa statistic is that it cannot tolerate unequal score ranges. For example, if one rater assigned scores consisting of 0, 1, 2, and 3, while the other rater assigned scores of 0 and 1, 2, but not 3, the kappa could not be computed. The Kappa statistic does not consider the degree to which the rater's scores vary from

complete agreement. Disagreement between raters by a rating of 1 produces the same Kappa statistic as disagreement by a rating of 3. The kappa statistic therefore is a very conservative estimate of interrater agreement. As such, the percentage of direct hits and misses, Pearson correlations, and Kappa statistics, are used to determine the interrater reliability of the VRS.

#### Interrater Reliability of Total Score Correlations

The VRS total score correlations between raters was high for the provincial, federal, and combined samples ( $r > .85$ ,  $p \leq .01$ ). Even considering the stringent limitations of the Kappa statistic described above, the VRS Static and Dynamic Factors demonstrated 'good' agreement ( $k = .50$ ,  $p < .01$  and  $k = .39$ ,  $p < .01$ , respectively).

An examination of the frequency distributions of the VRS total scores for Rater 1 and Rater 2 indicates that the scores for the provincial, federal and combined sample are normally distributed based on the Shapiro-Wilk Test for Normal Distribution. For both the provincial and federal samples, VRS total scores ranged from 25 to 75. This finding rules out the possibility that the high VRS interrater reliabilities are the result of a restricted range of scores that could artificially inflate the interrater reliability of the scale.

Overall, the investigations indicate that the high interrater reliability of the VRS is stable across provincial and federal offender samples. Over 58% of the ratings were in complete agreement and 35% differed by only one scale point. Given the stringent requirements of the Kappa statistic, the fair to good kappa values obtained for both the static and dynamic total scores and the significant Pearson  $r$  coefficients, can be interpreted as indicators of strong interrater reliability for the VRS as a whole.

### **Internal Consistency of the VRS**

The high Cronbach's alpha coefficient indicates that the VRS is internally consistent and that the items comprising the VRS are most likely measuring one unidimensional construct. Alpha coefficients of the VRS computed after the sequential deletion of each individual item were used to determine the item's contribution to the scale's internal consistency. Only D12, Mental Disorder, and D19, Violent Sexual Behavior, were considered to be 'weaker items'. These findings may be because no offender with an acute psychiatric illness (psychosis or affective disorders) indicated on file, was included in the study. As well, the sex offenders were restricted to one sub-sample (Clearwater patients) and were not generally distributed in the sample. These two items will be retained however as they will be useful when the VRS is applied to other study samples.

Post-hoc item to item correlational analyses confirmed that all except three items that comprise the VRS are contributing unique variance to the VRS total score. The high internal consistency of the VRS was not a function of the presence of many redundant scale items. As noted previously, only 3 item-pairs showed overlap (i.e., Violence throughout the Lifespan with Violent Lifestyle; Violence Lifestyle with Violence Cycle; and Criminal Attitudes with Respect for Authority and Social Convention). Further reading of the rating instructions revealed that the operationalizations of these items were overlapping. For example, the rating descriptions D3, Criminal Attitudes, directs the rater to rate the offender as high risk if the offender demonstrates no respect or compliance with society's laws and law

enforcement agencies. Similar information is captured in D5, Respect for Authority and Social Convention (i.e., “the offender consistently disregards authority, social conventions and rules”). As a result, Respect for Authority did not contribute unique variance to the VRS total score. These rating descriptions will be edited in the revised edition of the Violence Risk Scale.

Further support for the high internal consistency of the VRS was evidenced by the split-half reliability analyses which indicated that the internal consistency of the VRS was not the result of a scale with a large number of items. Even when the number of scale items were reduced by half, the alpha coefficient remained high (i.e., alpha value  $>.81$  for odd-numbered items and  $>.86$  for even-numbered items).

Overall, the internal consistency analyses indicate that the items comprising the VRS are measuring a unidimensional underlying construct and all except three of the Static and Dynamic Factor items capture unique variance of the underlying construct.

## **Validity of the VRS**

### **Overview of the Validity Investigations**

The validity of the Violence Risk Scale was evaluated by 1) investigating the relationship between VRS total scores and scores obtained on validated measures of criminal risk and violence; 2) evaluating the relationship between VRS scores and the offender's criminal history; and 3) examining the relative contribution of the VRS and the comparison measures to the postdiction of violent convictions.



### Validity Correlations

The extent to which the VRS is measuring the violent recidivism construct was investigated by evaluating the relationship of the VRS with validated risk measures that collectively capture the violence and recidivism constructs. VRS total scores were significantly correlated with most of the comparison measures.

To date, among other risk assessment instruments, the PCL-R has demonstrated the strongest relationship with violent recidivism ( $r=.28$ ). The high correlation between VRS and PCL-R total scores supports the contention that the items comprising the VRS collectively tap the violent recidivism and criminal psychopathy construct. The PCL-R predicts violence but it was not designed to measure changes in the offender's risk. It is expected that the VRS, with the extensive domain of dynamic risk factors, can be used to measure changes in risk.

The high correlation between VRS and LSI-R total scores indicates that the VRS is tapping the underlying LSI-R construct, 'propensity toward rule violation' (Andrews & Bonta, 1991). The LSI-R is the only instrument currently available that has both static and dynamic variables and is strongly linked to general criminal recidivism (Gendreau et al., 1996; Loza & Simourd, 1994). The high correlation between the VRS and the LSI-R indicate that the two are measuring overlapping constructs which is to be expected. Propensity towards rule violations is part and partial of criminal violence.

Noteworthy in the present findings is that there was a ceiling effect in the federal sample's LSI-R scores<sup>1</sup>. Compared to the normative sample mean score of

---

<sup>1</sup> The rated version of the LSI-R was not conducted for the provincial sample. The self-report version of the LSI-R was terminated prior to the federal sample data collection due to technical scoring problems of the SRI (Further detail is provided in the Procedures section).

26.9, (S.D.=9.1), the federal sample mean was 48.6, (S.D.= 3.2, range 42 - 54). The LSI-R Profile Form (Andrews & Bonta, 1995) for inmates, indicates that subjects scoring 41 or higher are categorized as 'High Risk/Needs' offenders. A LSI-R score of 41 corresponds to a percentile ranking of 98. This suggests that the LSI-R loses discriminative power when assessing federal offenders similar to those at the RPC who are among some of the higher risk offenders in the federal system.

The VRS score distribution for the present sample is not positively skewed and will provide better discrimination for higher-risk offenders than the LSI-R. This finding concurs with a recent study that evaluated the predictive power of the most prominent risk instruments (Gendreau et al., 1996). The results of the meta-analysis indicate that in the case of specialized offender populations, for the prediction of violent risk, the LSI-R should be used in conjunction with additional measures (e.g., the PCL-R).

The VRS scores were significantly correlated with GSIR scores which are based solely on criminal history information. The GSIR is one of the most widely utilized static risk assessment instruments in Canada, particularly within Correctional Services of Canada. Although the efficacy of the GSIR as a violent recidivism instrument is limited (Bonta et al., 1996), a strong relationship between GSIR scores and general recidivism is well established ( $r=.42, p\leq.01$ ). The strong correlation between the VRS and GSIR support the contention that the VRS is strongly related to the criminal recidivism construct. Given that previous criminal history is a strong predictor of recidivism (e.g., Gendreau et al., 1996; Monahan, 1981; Mossman, 1994), it is important to capture criminal history variables as a component in a comprehensive

assessment of risk. As expected, VRS static factor scores were more strongly correlated with GSIR scores ( $r=.79, p \leq .01$ ) than with the dynamic factor scores ( $r=.38, p \leq .01$ ) or with VRS total scores ( $r=.49, p \leq .01$ ).

The discriminant validity of the VRS was evidenced by the significant correlations with the aggressiveness domains of the IBS and non-significant correlation with the assertiveness domains. Examination of the validity sub-scale scores indicated that the scores fell within the average range. The profiles appear valid and suggest that the obtained scores were not attributable to denial, infrequent responses, or impression management. The correlations between the IBS aggression sub-scales and the VRS total scores provides support that the VRS captures one of the major constituent components of the violent recidivism construct, namely, aggression. The VRS appears to be able to discriminate between aggressiveness and assertiveness.

An unexpected result was the finding that there was no significant relationship between AQ total score, or any of the AQ subscales, and VRS total scores. This finding was observed in both the provincial and federal samples.

Although the validity of the AQ in an offender population is limited, the inclusion of the AQ in the present study was theoretically and methodologically sound. The AQ represents the revised version of the Buss Durkey Hostility Inventory (BDHI) which is one of the most frequently cited aggression questionnaires. In developing the AQ, Buss and Perry (1992) retained some of the original items, problem items were deleted or modified, and new items were added to enhance clarity and to reduce ambiguity.

The AQ normative means were derived using a sample of college students. At the present time, there are few validation studies of the AQ in offender populations.

Williams et al. (1996) investigated the AQ in an offender population and found that the offender sample means were not dissimilar to the scores of the college sample. This may suggest that the AQ is not discriminatory in offender populations. In the present study, it may be that characteristics unique to the present offender populations limited the utility of the AQ. There are no validity scales in the AQ and therefore, invalid responding is a potential confound. However, the obtained results on the validity subscales of the IBS suggests that the subjects were not engaged in invalid responding. Further investigations of the validity of the AQ in offender populations is warranted.

#### The Relationship Between VRS Scores and the Postdiction of Violence

Support for the construct and discriminant validity of the VRS in assessing criminal violence is further indicated by the significant correlation found between VRS scores and past official violent convictions but not with non-violent convictions. The author is aware that a small number of VRS items are directly linked to violent convictions and may inflate the observed correlation. However, given the high internal consistency of the scale, the deletion of these items will not likely affect the overall VRS-violence correlations.

The unique contribution of the VRS to the assessment of risk of violence is demonstrated by the results of the stepwise multiple regression analyses. The VRS made the largest contribution to the postdiction of violence followed by the PCL-R, Factor 2. These findings indicate that the VRS out-performs the comparison measures in the postdiction of violent convictions. For non-violent convictions, as one would predict, the GSIR made the largest contribution to the postdiction, followed by the VRS which added a small but significant amount of variance.

The relationship between VRS total scores and 'Age at First Conviction' (AFC) and the 'Criminal Career Profile' (CCP) slope was also examined. VRS scores were significantly related to AFC. As predicted, VRS total scores were inversely related to Age at First Conviction. The higher the VRS scores, the earlier the offender received the first conviction which is a good indication of the persistence of criminality.

An unexpected finding was the non-significant relationship between VRS total scores and the CCP slope for the sample. The sample consists of a number of lifers and other offenders who are serving single long sentences with limited past violent criminal histories (e.g., incest offenders). These types of offenders make the calculation of the slope of the CCP technically difficult and do not provide good estimates of the slope of the CCP. Unlike official records on which the CCP is based, the VRS assesses many non-adjudicated criminal activities and violence. This may explain the lack of correlation between the CCP slope and the VRS scores.

### **Conclusions**

Overall, the three study hypotheses were strongly supported. The results provide converging evidence that the VRS is a reliable and valid instrument for assessing the risk of violence in offender populations. High interrater reliability was achieved indicating that the VRS can be rated reliably by trained independent raters. Item analyses confirms that the observed high internal consistency of the VRS is not an artifact of redundant scale items or the length of the scale. The validity of the VRS in the assessment of violence-risk is supported by its relationship to validated measures of criminal/violent risk, the PCL-R, LSI-R, and GSIR. The discriminate validity of the

VRS was evidenced by the significant correlations with aggressiveness and non-significant relationship with assertiveness, and with violent but not with non-violent convictions. The VRS also outperforms the other risk assessment instruments in the postdiction of violence but at the same time, provides assessments of the change in violent risk.

### **Limitations of the Study**

A limitation of the present study is that the subjects were not drawn from a random sample of offenders. Subjects volunteered to participate in the study and this may have introduced a potential confound. The characteristics of the study settings however precluded drawing subjects from a random sample. Time restraints limited the number of subjects used; a larger sample size may have enhanced the significance of the findings.

The quantity and quality of the information used to score the ratings differed between the provincial and federal samples. The federal sample participants were enrolled in an intensive treatment program at the time of data collection and as a result, the effects of treatment may have effected the results.

The provincial sample comprised 87% aboriginal offenders compared with only 43% in the federal sample. The composition of the present sample will have limited the generalizability of the results to other samples with other racial compositions.

### **Recommendations for Future Research**

The few problematic and overlapping scale items identified in this study need to be addressed and revised accordingly. Future evaluative research on the VRS using self-report inventories will be enhanced by the inclusion of validity measures designed to detect systematic response biases. Also, for participants who fail to meet the minimum reading level requirements, or for those participants that request assistance to complete the self-report measures, a standardized delivery format should be implemented to minimize potential biases.

Future research should focus on the assessment of the VRS as a violent recidivism prediction instrument in low, medium and high-risk offender populations. Longitudinal follow-up at various time intervals (e.g., six-month, 1-year, 3-year, 5-year), should be undertaken to evaluate the predictive validity of the VRS. As well, the reliability and validity of pre- and post-treatment VRS scores (Part B) of the Violence Risk Scale requires evaluation. Such an evaluation should include an assessment of the interrater reliability of the post-treatment rating descriptions as well as an examination of the relationship between post-treatment VRS scores and scores obtained on validated measures of treatment outcome and violence. The validity of the full scale should also be specifically assessed in female, adolescent, and aboriginal as well as non-aboriginal minority populations.

## REFERENCES

- Andrews, D. A. (1982). The Level of Service Inventory (LSI): The first follow-up. Toronto: Ministry of Correctional Services.
- Andrews, D.A., Kiessling, J.J., Mickus, S.G., & Robinson, D. (1986). The construct validity of interview-based risk assessment in corrections. Canadian Journal of Behavioral Science, 18, 460-470.
- Andrews, D. A., & Bonta, J. (1994). The Psychology of Criminal Conduct. Cincinnati: Anderson.
- Andrews, D. A., & Bonta, J. L. (1991). The Level of Service Inventory-Revised. Interview and scoring guide. Report on the assessment and evaluation project. Ottawa: Ontario Ministry of Correctional Services.
- Andrews, D.A., & Bonta, J.L. (1995). The Level of Service Inventory-Revised. Toronto: Multi-Health Systems, Inc.
- Baird, C.S. (1981). Probation and parole classification: The Wisconsin model. Corrections Today, 43, 36-41.
- Blackburn, R. (1993). The Psychology of Criminal Conduct. Chichester: John Wiley & Sons.
- Bonta, J. (1993). The level of supervision inventory: an overview. The IARCA Journal on Community Corrections, 4, 6-7.
- Bonta, J. (1996). Risk-needs assessment and treatment. In A. Harland (ed.), Choosing Correctional Options that Work: Defining the Demand and Evaluating the Supply. Thousand Oaks, CA: Sage.



Bonta, J., Harman, W. G., Hann, R. G., & Cormier, R. B. (1996). The prediction of recidivism among federally sentenced offenders: A re-validation of the GSIR scale. Canadian Journal of Criminology, January, 1996, 61-79.

Bonta, J., & Motiuk, L. L. (1985). Utilization of an interview-based classification instrument: A study of correctional halfway houses. Criminal Justice and Behavior, 12, 333-352.

Bonta, J., & Motiuk, L.L. (1992). Inmate classification. Journal of Criminal Justice, 20, 343-353.

Borum, R. (1996). Improving the clinical practice of violence risk assessment. American Psychologist, 51, 945-956.

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. Journal of Personality and Social Psychology, 63, 452-459.

Clark, L.A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. Psychological Assessment, 7, 309-319.

Cornell, D.G., Warren, J., Hawk, G., Stafford, E., Oram, G., & Pine, P. (1996). Psychopathy in instrumental and reactive violent offenders. Journal of Consulting and Clinical Psychology, 64, 783-790.

Douglas, K.S., MacFarlane, E., & Webster, C.D. (1996). Predicting dangerousness in the contemporary Canadian mental health and criminal justice systems. Canada's Mental Health, 43, 4-11.

Farrington, D. P., & Gunn, J. (1985). Introduction. In D. P. Farrington & J. Gunn (Eds.), Aggression and Dangerousness. NY: John Wiley & Sons.

Farrington, D. P., & Tarling, R. (1985). Criminological Prediction: An Introduction. In D. P. Farrington & J. Gunn (Eds.), Prediction in Criminology (pp. 2-33). Albany: State University of New York Press.

Fleiss, J. L. (1975). Measuring agreement between two judges on the presence or absence of a trait. Biometrics, 31, 651-659.

Fleiss, J.L. (1981). Balanced incomplete block designs for inter-rater reliability studies. Applied Psychological Measurement, 5, 105-112.

Fulero, S. M. (1995). Review of The Hare Psychopathy Checklist-Revised. In J. C. Conolley, J. C. Impara, & L. Murphy (Eds.), The twelfth mental measurements yearbook (pp. 453-454). Lincoln: University of Nebraska.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! Criminology, 34, 575-607.

Gordon, A., & Wong, S. (1995). The evaluation of interrater reliability of dynamic risk predictors. Canadian Psychology, 36, 23.

Gottfredson, S. D., & Gottfredson, D. M. (1985). Screening for risk among parolees: policy, practice, and method. In D. P. Farrington & R. Tarling (Eds.), Prediction in Criminology (pp. 54-78). Albany: State University of New York Press.

Hann, R. G., & Harman, W. G. (1989). Release Risk Prediction: A test of the Nuffield Scoring System. Corrections Research, Ministry Secretariat, the Ministry of the Solicitor General of Canada.

Hann, R. G., & Harman, W. G. (1992). Predicting general release risk for Canadian penitentiary inmates. Corrections Branch of the Ministry of the Solicitor General of Canada.

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. Personality and Individual Differences, 1, 111-119.

Hare, R. D. (1991). The Hare Psychopathy Checklist-Revised. Toronto: Multi-Health Systems.

Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P. (1990). The Revised Psychopathy Checklist: Reliability and factor structure. Psychological Assessment, 2, 338-341.

Harris, G. T., Rice, M. E., & Cormier, C. A. (1991). Psychopathy and violent recidivism. Law and Human Behavior, 15, 625 - 637.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders. The development of a statistical prediction instrument. Criminal Justice and Behavior, 20, 315-335.

Hart, S. D., & Hare, R. D. (1989). Discriminant validity of the Psychopathy Checklist in a forensic psychiatric population. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1, 211-218.

Hemphill, J.F., & Wong, S. (1997). Hare's Psychopathy Checklist and Recidivism. Legal and Criminological Psychology. In press.

Hoffman, P.B. (1983). Screening for risk: A revised salient factor score (SFS 81). Journal of Criminal Justice, 11, 539-547.

Kosson, D.S., Smith, S.S. & Newman, J.P. (1990). Evaluating the construct validity of psychopathy on Black and White male inmates. Three preliminary studies. Journal of Abnormal Psychology, 99, 250-259.

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

Lidz, C., Mulvey, E., & Gardner, W. (1993). The accuracy of predictions of violence to others. Journal of the American Medical Association, 269, 1007-1011.

Loza, W., & Simourd, D. J. (1994). Psychometric evaluation of the Level of Service Inventory (LSI) among male Canadian federal offenders. Criminal Justice and Behavior, 21, 468-480.

Mauger, P. A., & Adkinson, D. R. (1980). Interpersonal Behavior Survey. Los Angeles: Western Psychological Services.

Meehl, P. (1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Minneapolis:University of Minnesota Press.

Menzies, R., & Webster, C. D. (1995). Construction and validation of risk assessments in a six-year follow-up of forensic patients: A tri-dimensional analysis. Journal of Consulting and Clinical Psychology, 63, 766-778.

Monahan, J. (1980). The clinical prediction of violent behavior. Maryland: National Institute of Mental Health.

Monahan, J. (1981). Predicting violent behavior: An assessment of clinical techniques. Beverly Hills: CA: Sage.

Monahan, J., & Steadman, H.J. (Eds.) (1994). Violence and mental disorder: Developments in risk assessment. Chicago: University of Chicago Press.

Monahan, J. (1996). Violence prediction: The past twenty years and the next twenty years. Criminal Justice and Behavior, 23, 107-120.

- Morris, N., & Miller, M. (1993). Predictions of dangerousness in the criminal law. The IARCA Journal on Community Corrections, 4, 17-22.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. Journal of Consulting and Clinical Psychology, 62, 783-792.
- Motiuk, L. L. (1993). Using the LSI and other classification systems to better predict halfway-house outcome. The IARCA Journal on Community Corrections, 4, 8-9.
- Motiuk, M. S., Motiuk, L. L., & Bonta, J. (1992). A comparison between self-report and interview-based inventories in offender classification. Criminal Justice and Behavior, 19, 143-159.
- Nuffield, J. (1982). Parole decision-making in Canada: Research towards decision guidelines. Ottawa: Solicitor General of Canada.
- Otto, R. (1992). The prediction of dangerous behavior: A review and analysis of "second generation" research. Forensic Reports, 5, 103-133.
- Pollock, N., McBain, I., & Webster, C. D. (1989). Clinical decision making and the assessment of dangerousness. In K. Howells & C. R. Hollin (Eds.), Clinical Approaches to Violence West Sussex: John Wiley & Sons Ltd.
- Quinsey, V.L., Rice, M., & Harris, G.T. (1995). Actuarial prediction of sexual recidivism. Journal of Interpersonal Violence, 10, 85-105.
- Rice, M.E. & Harris, G.T. (1995). Psychopathy, schizophrenia, alcohol abuse, and violent recidivism. International Journal of Law and Psychiatry, 18, 333-342.
- Serin, R.C. & Amos, N.L. (1995). The role of psychopathy in the assessment of dangerousness. International Journal of Law and Psychiatry, 18, 231-238.

Serin, R. C. (1996). Violent recidivism in criminal psychopaths. Law and Human Behavior, 20, 207-217.

Templeman, R. (1997). Crime Ware. Saskatchewan: Canada. Unpublished Manuscript.

Webster, C. D., & Eaves, D. (1993). The HCR-20: A scheme for the assessment of dangerousness and risk. Mental Health, Law and Policy Institute Department of Psychology, Simon Fraser University.

Webster, C.D., Eaves, D., Douglas, K. & Wintrup, A. (1995). The HCR-20 Scheme: The assessment of dangerousness and risk. British Columbia, Canada: Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.

Webster, C. D., Harris, G. T., Rice, M. E., Cormier, C., & Quinsey, V. L. (1994). The Violence Prediction Scheme: Assessing dangerousness in high-risk men. Toronto: Centre of Criminology.

Webster, C. D., & Menzies, R. J. (1993). Service in the deinstitutionalized community. In S. Hodgins (Eds.), Mental disorder and crime (pp. 22-38). Newbury Park: Sage.

Williams, T.Y., Boyd, J.C., Cascardi, M.A., Poythress, N. (1996). Factor structure and convergent validity of the Aggression Questionnaire in an offender population. Psychological Assessment, 8, 398-403.

Wong, S., & Gordon, A. (1996). Violence Risk Scale - Experimental Version 1 (VRS). Department of Psychology and Research. Regional Psychiatric Centre, Saskatchewan, Canada: Solicitor General of Canada.

Wong, S. (1996). Categorization of violent and non-violent offending.

Unpublished manuscript.

Wong, S., & Templeman, R., Gu, D., Andre, G., & Leis, T. (1996). Criminal Career Profile: A quantitative index of past violent convictions. Unpublished Manuscript.

## APPENDIX A

### VIOLENCE RISK SCALE -EXPERIMENTAL VERSION 1 (VRS-E1)

#### SCALE ITEMS

<b>Static Factors</b>	
S1. Current Age	S4. Violence Throughout the Lifespan
S2. Age at First Violent Conviction	S5. Prior Release Failures/Escapes
S3. Number of Juvenile Convictions	S6. Stability of Family Upbringing
<b>Dynamic Factors</b>	
D1. Violent Lifestyle	D13. Substance Abuse
D2. Criminal Personality	D14. Stability of Relationships with Significant Others
D3. Criminal Attitudes	D15. Community Support
D4. Work Ethic Situations	D16. Released Back to High Risk
D5. Respect for Authority & Social Conventions	D17. Violence Cycle
D6. Criminal Peers	D18. Impulsivity
D7. Interpersonal Aggression	D19. Violent Sexual Behavior
D8. Emotional Disinhibition	D20. Compliance with Supervision
D9. Violence During Incarceration	D21. Stages of Change
D10. Weapon Use	D22. Security Level of Anticipated Release Institution
D11. Insight Into the Cause of Violence	D23. Longest Violent Incident Free Period in the Last Five Years
D12. Mental Disorder	



## APPENDIX B

### INFORMED CONSENT STUDY CHARACTERISTICS

The purpose of this study is to determine the reliability and validity of an assessment instrument.

- \* I understand that my participation is completely voluntary. I may withdraw at anytime without affecting my sentence or my stay at the Saskatoon Correctional Centre. (The information that I provide will not be recorded in my institutional files or discussed with institutional staff. However, group information may be shared with institutional staff.
- \* To guarantee confidentiality, my name will not be associated with my answers. Instead, my answers will be coded with an identification number known only to the investigators, and this information will be secured in a separate location. However, the researchers may have to release the data if ordered by a Court of Law.
- \* The investigators may review my institutional file(s).
- \* I will be asked to complete a three-part questionnaire which will require approximately 45 minutes. I will also be interviewed in private by two separate researchers, each interview will take approximately 1 hour.
- \* If the results are published in a scientific journal, all participants will remain anonymous; no individual results will be used.
- \* I have carefully read this form and have clearly understood what my participation in this study will involve.

I agree to volunteer and participate in the study described above.

\_\_\_\_\_  
Please print name

\_\_\_\_\_  
ID Number

SIGNATURE: \_\_\_\_\_

Witness: \_\_\_\_\_

DATE: \_\_\_\_/\_\_\_\_/\_\_\_\_

- \* I would like to receive a copy of the summarized results of the project. Yes \_\_\_ No \_\_\_

If you have any questions about any aspect of this project, you may ask the researchers or phone Dr. Stephen Wong at 975-4156.

**APPENDIX C**  
**INFORMED CONSENT**  
**STUDY CHARACTERISTICS**

The purpose of this study is to determine the reliability and validity of an assessment instrument.

- \* I understand that my participation is completely voluntary. I may withdraw at anytime without affecting my sentence or my stay at the Regional Psychiatric Centre. (The information that I provide will not be recorded in my institutional files or discussed with institutional staff. However, group information may be shared with institutional staff).
- \* To guarantee confidentiality, my name will not be associated with my answers. Instead, my answers will be coded with an identification number known only to the investigators, and this information will be secured in a separate location. However, the researchers may have to release the data if ordered by a Court of Law.
- \* The investigators may review my institutional file(s).
- \* I will be asked to complete a two-part questionnaire which will require approximately 45 minutes. I will also be interviewed in private by two separate researchers, each interview will take approximately 1 hour.
- \* If the results are published in a scientific journal, all participants will remain anonymous; no individual results will be used.
- \* I have carefully read this form and have clearly understood what my participation in this study will involve.

I agree to volunteer and participate in the study described above.

\_\_\_\_\_  
Please print name

\_\_\_\_\_  
ID Number

SIGNATURE: \_\_\_\_\_

Witness: \_\_\_\_\_

DATE: \_\_\_\_/\_\_\_\_/\_\_\_\_

- \* I would like to receive a copy of the summarized results of the project. Yes\_\_\_No\_\_\_

If you have any questions about any aspect of this project, you may ask the researchers or phone Dr. Stephen Wong at 975-4156.



UNIVERSITY ADVISORY COMMITTEE  
ON ETHICS IN HUMAN EXPERIMENTATION

(Behavioral Sciences)

NAME : D. Saklofske, Department of Educational Psychology      EC# 96-132  
S. Wong, Department of Psychology  
(A. Gordon)

DATE:      October 17, 1997

The University Advisory Committee on Ethics in Human Experimentation (Behavioral Sciences) has reviewed the modifications to your study, "The interrater reliability, internal consistency and validity of the Violence Risk Scale-Experimental Version 1 (VRS-E1)" (96-132).

1.      Your study has been APPROVED.
2.      Any significant changes to your protocol should be reported to the Chair for Committee consideration in advance of its implementation.
3.      The term of this approval is for 3 years.

---

David Hay, Chair  
University Advisory Committee  
on Ethics in Human Experimentation  
Behavioral Sciences

Please direct all correspondence to:

Secretary  
UACEHE, Behavioral Science  
Office of Research Services  
University of Saskatchewan  
Room 210 Kirk Hall, 117 Science Place  
Saskatoon, SK S7N 5C8

## **APPENDIX E**

### **DESCRIPTION AND PSYCHOMETRIC REVIEW OF VRS COMPARISON MEASURES**

#### **Aggression Questionnaire (Buss and Perry, 1992)**

The Aggression Questionnaire (AQ) represents the revised version of the Buss and Durkee Hostility Inventory (BDHI) (one of the most frequently used questionnaires on aggression, with 242 citations in the Social Science Citation Index between 1960 and 1989). The BDHI has been shown to have predictive validity in populations of violent prisoners (Gunn & Gristwood, 1975), delinquent adolescents (Romney & Syverson, 1984), and aggressive men (Barnett, Fagan, & Booker, 1991; cited in Williams et al., 1996). The AQ is a 19 item self-report questionnaire scored on a 5-point Likert scale ranging from 0-"Not at all descriptive" to 5-"Very descriptive". The scale is comprised of four sub-scales that are theoretically and/or empirically related to violence, namely, physical aggression, verbal aggression, anger, and hostility. From a theoretical standpoint, physical and verbal aggression are viewed as different types of aggression, whereas hostility and anger are considered as contributing factors to aggression (Williams et al., 1996). The sub-scale and total scores will be correlated with the VRS total scores.

#### Psychometric Review:

In developing the AQ, Buss and Perry (1992) retained some of the BDHI's original items. Problematic items were modified or deleted and new items were added for greater clarity and to reduce ambiguity. A Likert-scale response format replaced the true-false format of the BDHI.

### Confirmatory Factor Analysis and Correlations Among Factors

A series of exploratory and confirmatory factor analytic studies using three samples of male and female undergraduate students (n=406, n=448, and n=399) yielded a four-factor solution for the 29 items (i.e., Physical Aggression, Verbal Aggression, Hostility, and Anger). Item selection was based on two criteria. An item had to load at least .35 on its own factor but less than .35 on any other factor and, both of these specifications had to be met for all three samples. Of the original 52 items, 23 did not meet these criteria and were excluded (Buss & Perry, 1992). The greatest variation in factor loading across samples was .25 (for the Anger item "I sometimes feel like a powder keg ready to explode"). Most items produced small variations across the three samples (i.e.,  $\leq .13$ ).

All factors were intercorrelated beyond chance levels. Verbal and Physical Aggression were closely related (.45) but only moderately correlated with Hostility (.28 and .25 respectively). Anger correlated strongly with the other three factors (.45 to .48). Buss and Perry's post hoc hypothesis for the moderate correlations found between Hostility and both Physical and Verbal Aggression was due to the items sharing an overlapping relationship with Anger. This interpretation was supported when the correlations between these sub-scales was severely attenuated when Anger was subtracted out (.08 between Hostility and Physical Aggression and .05 between Hostility and Verbal Aggression).

### Internal Consistency

The internal consistency of the four factors and the total score was evaluated using the alpha coefficient for all subjects (n=1253). For the four sub-scales the alphas were as follows: Physical Aggression, .85; Verbal Aggression, .72, Anger, .83; and Hostility, .77. The total score produced an alpha level of .89.

### Norms and Sex Differences

Men evidenced significantly higher scores on Physical Aggression, Verbal Aggression, and Hostility, but not on Anger. To derive quantitative estimates of gender differences, standardized mean differences for each scale were calculated to determine effect size.

Effect sizes were .89 for Physical Aggression, .44 for Verbal Aggression, .05 for Anger, and .19 for Hostility.

### Reliability

One sample of 372 subjects was tested twice at a 9-week interval. The test-retest correlations were as follows: Physical Aggression, .80; Verbal Aggression, .76; Anger, .72; and Hostility, .72 (total score = .80). Buss and Perry (1992) note that these coefficients suggest adequate stability over time for scales with relatively small numbers of items.

### Correlations with Other Personality Traits

The four sub-scales were correlated with various personality traits (i.e., emotionality, activity [by gender], impulsiveness, sociability, assertiveness, competitiveness, public self-consciousness, private self-consciousness [by gender], and self-esteem). Evidence of discriminate and convergent validity of the AQ was demonstrated as the inventory correlated weakly with personality traits such as activity, sociability, public and private self-consciousness, and self-esteem. Strong correlations with Anger were found between Emotionality (.43), Impulsiveness (.42), and Assertiveness (.40). Physical Aggression correlated most strongly with Competitiveness (.36) followed by Assertiveness (.28). Not surprisingly, Verbal Aggression was most related to Assertiveness (.49) followed by Competitiveness (.39). The highest correlations for Hostility were with Emotionality (.52) and Impulsiveness (.37). Evidence of convergent validity of the AQ was also found when undergraduate males rated by their peers as aggressive were more likely to score as more aggressive on the AQ than were men rated as nonaggressive.

### Limitations

The data described above were based on a sample of university students which limits the generalizability to other populations.

### The Factor Structure and Convergent Validity of the Aggression Questionnaire in Offender Populations.

Williams, Boyd, Cascardi, and Poythress (1996) evaluated the AQ in an adult offender sample consisting of men and women. The sample (n=200) was composed of Caucasian (40%), African American (49.5%), and Hispanics (8%), and Others, 2.5%. Participants ranged in age from 17 to 69, with a mean age of 30 (SD=9.20), 38% of the sample was female, and 62% male. Twenty-two percent of the sample did not have a high school education, 38% had received some high-school and 23.5% had completed high school or had obtained a general equivalency diploma (GED). Eight offenders requested that the questionnaire be read to them due to reading difficulties. The offense pattern of the participants ranged from felonies (56%), misdemeanors (39%) and violation of parole (4%).

#### Convergent Validity

Responses on the AQ were correlated with responses on the Novaco's Anger Scale (NAS, Novaco, 1994; cited in Williams et al., 1996). The NAS is a two-part self-report measure that assesses three major domains of anger: cognitive, arousal, and behavioral. Part A contains 48 items regarding how people may experience anger. Part B contains 25 items describing situations that may potentiate anger. Test-retest reliability, internal consistency, and concurrent and predictive validity for the scale are at acceptable levels (Novaco, 1994; cited in Williams et al., 1996).

The correlation between the total score of the AQ and NAS was .79. The pattern of results indicated that in general the AQ Physical Aggression/Anger sub-scale tended to correlate higher with the Behavioral domain of the NAS and its respective sub-scales. The Verbal Aggression/Hostility factor of the AQ tended to correlate highest with the Cognitive domain of the NAS and its sub-scales. Both factors were similarly correlated with the Arousal domain.

To determine whether there were significant differences in the correlations between the NAS sub-scales and each AQ factor, *T*-tests for independent samples for each pair of correlations were performed. The correlations were significantly higher between Physical Aggression/Anger and the Behavioral domain than the correlations between Verbal Aggression/Hostility and these NAS sub-scales. The Verbal Aggression/Hostility factor of the AQ correlated highest with the NAS Cognitive domain ( $r=.70$ ), although this correlation was not significantly different from the correlation between Physical Aggression/Anger and the Cognitive domain ( $r=.62$ ) (Williams et al., 1996).

#### Confirmatory Factor Analysis

Although the four-factor structure of the AQ was replicated (Physical, Verbal, Anger, and Hostility), conventional fit indices suggested that the four-factor structure did not constitute a good fit for this sample of adult offenders (William's et al., 1996). Examination of a test of sampling adequacy revealed that three items were problematic ("I tell my friends openly when I disagree with them", "I am an even-tempered person", and "I can think of no good reason for ever hitting a person". These three items demonstrated the lowest corrected item-total correlations ( $r=-.16$ ,  $-.02$ , and  $.06$  respectively) indicating the lack of shared variance with the scale. All other items had excellent item-total correlations ( $\geq .80-.90$ ). Two, rather than four factors, predominated the analyses: physical aggression/anger and verbal aggression/hostility. Factor 1 consisted of 14 items and Factor 2, 12 items. The percentages of variance explained by the factors were 36.4% and 4.0%, respectively. The factor matrix indicated that the two factors correlated  $.69$  with each other. Physical Aggression/Anger and Verbal Aggression/Hostility were highly correlated with the total score,  $r=.94$  and  $.89$ , respectively (Williams, et al., 1996). The high internal consistency of the AQ total score and the high intercorrelations of the AQ factors may indicate a unidimensional construct of aggression.



## Summary

Overall the Williams et al. (1996) study suggests that in offender populations, those who are high in hostility may be more likely to be verbally aggressive, whereas those high in anger may be more likely to be physically aggressive. These results are consistent with aggression theory that divides the construct of aggression into two discrepant components (i.e., physical vs. verbal, direct vs. indirect, impulsive vs. consciously controlled). The pattern of associations supports previous research showing anger to be an important risk factor for violence.

Mean aggression scores for the offender population were not dissimilar to the scores of the college sample used by Buss and Perry (1992). Contrary to predictions, mean AQ scores for the offender population were not significantly higher than those for the student population. Williams et al. (1996) suggest that this finding may reflect the relatively small percentage of offenders with charges categorized as aggressive (22%) compared with those coded as nonaggressive (75%). In addition, offenders with aggressive crime charges did not evidence significantly higher AQ scores than those classified as non-aggressive. The lack of concordance between AQ scores and aggressive crime charges may be indicative of the fact that adjudicated criminal charges may not adequately represent an offender's criminal history or pattern of aggressive behavior. Those offenders classified as nonaggressive may in fact have a history of non-adjudicated violence (Williams et al., 1996).

Although further research is required, Williams et al. (1996) suggest that the Aggression Questionnaire may be used in clinical research and in the assessment of violence in criminal populations.

### **Interpersonal Behavior Survey (Mauger & Adkinson, 1980)**

The Interpersonal Behavior Survey (IBS) is a self-report questionnaire developed to distinguish assertive behaviors from aggressive behaviors. The scale is sensitive to changes over time and/or treatment, all 272 items are written in present tense.

The IBS was developed based on the assumption that assertive and aggressive behaviors are independent response classes. The test developers define assertiveness as directed behavior aimed at reaching some desired goal. According to Mauger and Adkinson (1980), assertive individuals tend to be goal directed in spite of obstacles. If goals are blocked, the assertive person aims at eliminating the interference and not at attacking the offending individual. The assertive person may be competitive but they will abide by social conventions. They typically play to win but observe the rules in doing so. They would only resort to violent behaviors for self-defense. Mauger and Adkinson (1980) posit that aggressive behavior originates from attitudes and feelings of hostility toward others. Often the purpose of aggressive behavior is to attack or exert power over others. Attainment of the goal may be used as a rationalization for the aggressive actions. Aggressive people may deliberately violate or simply disregard the rights of others in pursuing their goals (Mauger & Adkinson, 1980).

Assertive and aggressive behaviors are multidimensional response classes. The IBS differentiates subclasses of assertive and aggressive behaviors rather than just providing single global scores for each of the two broad response classes. It identifies more specifically behavioral deficits or excesses and therefore should be helpful in planning individualized interventions. The IBS consists of 272 items in a T-F format and comprises four sub-scales: a validity scale (3-sub-scales) aggressiveness (7-sub-scales), assertiveness (8-sub-scales), and relationship styles (3 sub-scales).

### Psychometric Review:

#### Test Development

Three strategies were employed in the construction of the IBS scales. An initial set of scales was developed through internal consistency item analysis. Each IBS item was correlated with the Denial scale. In order to minimize the impact of socially desirable responding, items having significant correlations with the Denial scale were deleted from the item pool. The scales were then evaluated by item analysis procedures based on a multi-trait model. The resulting scales have no overlapping items and adequate internal consistency (Mauger & Adkinson, 1980).

The 272 IBS items are divided into three parts. Part I can be completed in approximately 10 minutes and consists of 38 items that provide a short general sample of both assertive and aggressive behaviors, as well as the Denial scale. Part I can be used as a screening test or as a sampling procedure to measure change over time as a result of intervention. Parts I and II (items 1-133) provide short scales measuring a wide array of assertive and aggressive behaviors. They are approximately one-half as long as Parts I through III (items 1-272) but sample almost as many behavioral subclasses. The short-version of the IBS (Parts I and II) is therefore useful in situations in which the administration time is limited. Part III adds the Impression Management, Passive Aggressiveness, Conflict Avoidance, Dependency, and Shyness scales to the assessment. The longer scales were developed to increase the reliability of brief scales. Not surprisingly, due to the increased number of items, the longer scales demonstrate greater reliability compared with the short scales. The longer version scales were developed using internal consistency item analysis procedures in which each potential new item was correlated with the existing short form of the scale. To ensure comparability between the IBS long and short versions, items that had correlations greater than .30 with the intended scale and correlations  $\leq .30$  with other scales were retained for the IBS long version. The interrelation matrix of the IBS scales is similar

for the short and long scales. The correlations between the short and long scales are high but Mauger & Adkinson caution that since they are part-whole correlations the comparability of the two versions of these scales is not conclusive. Mauger & Adkinson report that in some cases the longer scales have higher correlations with relevant scales on other inventories. For example, the Expression of Anger scale - long version correlated .46 with the Assault scale on the Buss-Durkee Hostility Inventory, whereas the same scale on the IBS short version correlated only .28 with it (Mauger & Adkinson, 1980). The test developers contend that research results based on the long scales should apply to the short scales as both appear to measure the same behavior.

### Normative Samples

The General Reference Norm Group modeled the demographic distributions of the 1970 United States census and consisted of 400 male and 400 female residents from the southern United States. The means and standard deviations for the normative sample were presented in Table 5 (see Results).

### Reliability

Test-retest reliability's over both a 2-day and 10-week period ranged from .71 to .96 and .80 to .93, respectively (SEMs ranged from 2.98 to 4.91 for the 2-day and from 2.18 to 4.05 for the 10-week test-retest format). The coefficient alpha is in the .60 to .88 range for both the construction and cross-validation samples.

### Validity

Factor analytic studies support the assumption that assertive and aggressive behaviors form distinct response clusters. The relationships between IBS scales were explored via a principal factor method. The findings for two samples (i.e., community residents and college students) are highly similar. The first factor is defined by substantial loadings from

all of the assertiveness scales. The second factor is defined by loadings from the two aggressiveness scales.

The validity scales measuring defensiveness - Denial (DE) and Impression Management (IM) - both load moderately on the second factor with negligible loadings on the other factors. Mauger and Adkinson (1980) report that as measured by the IBS, social desirability appears to have little relationship to assertiveness but does influence responses to aggressiveness scale items.

### Convergent and Discriminant Validity

Convergent and discriminant validities have been assessed by noting correlations of the IBS scales with other scale inventories. For example in a sample of college students, the General Assertiveness, Rational (SGR) scale correlated .47 with the Dominance scale of the California Psychological Inventory (CPI) (Gough, 1975; cited in Mauger & Adkinson, 1980). Discriminant validity of the General Assertiveness, Rational (SGR) scale is evidenced for example by minimal correlations with the Buss-Durkee Hostility Inventory. The Buss-Durkee Hostility Inventory however demonstrated moderate to high correlations with the General Aggressiveness, Rational (GGR) scale ( $r=.65$ ).

### Clinical Utility of the IBS

Mauger and Adkinson (1980) suggest that the behaviors sampled by the IBS scales are more in the normal range of personality functioning than in the pathological range. An interpretative strategy for interpreting IBS profiles is included in the test manual.

## **General Statistical Information on Recidivism Scale (Nuffield, 1982)**

The GSIR scale comprises 15 static risk indicators (e.g., Current offense, Age at Admission, Number of Previous Imprisonments, etc.) which can be rated based on information in institutional files. Although the GSIR scale has demonstrated efficacy in the prediction of recidivism, this success has been limited to general or non-violent recidivism.

### **Psychometric Review:**

#### **Construction Sample**

Data for constructing and assessing the predictive validity of the GSIR scoring system were derived from a *random* sample of roughly 2500 male releasees from federal penitentiaries from 1970 to 1972. All subjects were incarcerated for either violent and/or non-violent convictions. The follow-up period was three years with re-arrest for any indictable offense as the major outcome variable.

#### **Scale Development**

Fifteen variables were found to be related to recidivism in the sample (e.g. age at admission, current offense, interval at risk). Nuffield (1982) used a weighted Burgess method such that the risk factors were assigned weights depending upon their deviation from the base rate of success. For every difference of 5%, a score of +/-1 was assigned. First incarcerates had a success rate of 79% compared to the 56% base rate of success for the general offender population. For example, an inmate with no prior record of incarceration would be assigned a score of -4 for this factor. Inmates with five or more imprisonments, had a success rate of only about 43% and would receive a score of +2. Nuffield (1982) reported scores on the scale ranging from -24 to +19 which were used to define one of 5 probability of risk categories. 'Very good risk' (scores of -6 to -27; 4 out of 5 will not recidivate), 'Good' (scores of -1 to -5; 2 out of 3 will not recidivate), 'Fair' (scores of 0 to +4; 1 out of 2 will not recidivate), 'Fair to Poor' (+5 to +8; 2 out of 5 will not

recidivate), 'Poor Risk' (+9 to +30; 1 out of 3 will not recidivate). Each category contained at least 16% and no more than 25% of the sample (Hann & Harman, 1992) .

### Validation Study

#### Sample:

The sample was comprised of 3,267 inmates, released from federal penitentiaries during the fiscal year 1983-84, who were incarcerated for having committed an indictable offense. The average age of the sample was 27.2 years (SD=8.4) serving average sentences of 48.7 months (27% were serving sentences of two years, 33% three years or less, and 84% with sentences of five years or less). The sample consisted of Caucasian (85.8%), Aboriginal (8.6%), female (3%) and others (2.5%).

#### Scale Validation

Two demographic items (number of dependents and employment status) contained in the original GSIR scale were excluded in the validation study due to incompleteness in the available information required to score the items. Therefore, a slightly modified version of the GSIR scale was used. Hann and Harman (1992) replicated the 1970-1972 results. The only difference noted is that the system's ability to differentiate between 'Good' and 'Fair' risks (using the original five categories) had decreased significantly from 1970-72 to 1983/84. The difference in the success rates between these two categories was 14% for the 1970-72 sample and only 3% for the 1983/84 releases (Hann & Harman, 1992) . The 1970-72 groups labeled as 'Fair' risks had a success rate (53%) slightly below the overall average rate (56%) whereas the 1983/84 releases had a success rate slightly above the overall average for those releases (i.e., 58% vs. 53%). Overall, the 1983-84 sample was similarly divided with each of the five score groupings containing at least 16% but no more than 27% of the cases. The differences for overall success rates between the two samples were relatively minor (56% vs. 53%) and the 1983/84 sample was considered as comparable to the 1970-72 sample (Bonta et al., 1996) .

## **Level of Service Inventory-Revised (Andrews & Bonta, 1995)**

The Level of Service Inventory (LSI-R) is an assessment instrument designed to assess the risk and need areas of offenders and to aid case managers in supervision decisions.

### Psychometric Review:

The first LSI-R validation study comprised 598 male probationers for whom in-program and post-program outcome was monitored. Although not specifically specified, it does not appear that minority groups were excluded. The LSI-R was later validated on both provincial and federal male-inmate populations.

### Reliability Estimates

Interrater reliability ( $r=.94$ ) and 3-month test-retest temporal stability ( $r=.80$ ) were assessed when either/or both rater and time were varied. The reliability estimates were generally high, but decreased with increasing time intervals when different raters were sampled. However, there were few disagreements with respect to the level of supervision assignment, and the absolute difference between the LSI-R scores was always five or less (Andrews, 1982). It is important to note that the LSI-R utilizes both static and dynamic (or changeable) variables that account for some of the variability between test-retest scores. Assessments of the internal consistency of the LSI-R were derived from examinations of the subcomponent total correlations and *alpha* values. The LSI-R demonstrated moderate internal consistency ( $r=.72$ ). Several studies have since reported on the internal consistency of the LSI-R. As noted previously, the results indicate that the LSI-R items measure the same underlying dimension which has been labeled "propensity for rule violations" (Andrews & Bonta, 1991).

A review of eight independent studies indicated that overall *alpha* coefficients ranged from .64 to .90 (6 of the 8 studies reported *alpha* values of less than .83). A



summary of research findings indicate that *alpha* values for each of the sub-scales are moderate to high ( $r \geq .50$  to  $r \geq .80$ ).

### Factor Analyses

Factor analyses for the LSI-R has produced inconsistent results. Andrews and Robinson (1984; cited in Andrews, 1991) reported a three-factors structure. The first factor accounted for 75% of the variance. Factor score coefficients were Companions (.45), Leisure/Recreation (.28), and Attitudes (.21). Factor score coefficients for the second factor which accounted for 14% of the variance, were Reward at School or Work (.40), Money problems (.34), Accommodation (.20), and Family problems (.15). For the third factor, accounting for 11% of the variance, the dominant subtotals were problems with Alcohol/Drugs (.36), Emotional/Personal Disturbance (.15), and Criminal History (.40). A later factor analysis again reported three factors, although the factor loadings were inconsistent with the first findings (Bonta & Motiuk, 1985; cited in Andrews, 1991). In a subsequent factor analysis only two factors were revealed as significant, with the majority of the items loading on one factor (Bonta et al., 1985). Andrews (1991) suggests that the emphasis should be kept on the LSI-R total score and the subcomponent scores considered as independent sources of information. Andrews hypothesized that some of the discrepancies could be explained by differences in sample size and procedures used for factor extraction. In addition, factors structures for the LSI-R may vary between settings and populations (Andrews, 1991). However, other scales (e.g., The Psychopathy Checklist-Revised) report consistent loadings of subcomponents independent of the population type (e.g., Hare, 1991).

### Face Validity:

The face validity of LSI-R was derived from an extensive review of the recidivism literature as well as consultations with probation officers (Loza & Simourd, 1994). As stated above, most of the LSI-R items pertain to information commonly used in the probation and parole decision-making process.

### Construct Validity - (Convergent and Divergent Validities):

As discussed previously, the LSI-R total scores purport to measure the construct "propensity for rule violation", with the total score related to level of risk. Andrews, Bonta, Motiuk and Robinson (1984) administered the LSI-R, as well as alternative measures of the same target domain (i.e., rule violation), to a sample of prisoners<sup>1</sup> and a sample of probationers. The total LSI-R scores were most strongly associated with measures of generalized rule violation and skill deficits. In both prison and probation samples, intake LSI-R scores were positively correlated with indices of official levels of supervision. For probationers, a 90% agreement was reported between LSI-R and level of supervision assigned at intake, the LSI-R correlated at  $r=.40$  ( $n=341$ ) with the number of contacts with the probation officer, and  $r=.30$  ( $n=561$ ) with early discharge/closure of the probation file. For prisoners, the LSI-R correlated at  $r=.34$  ( $n=144$ ) with level of security assigned at intake, at  $r=.25$  ( $n=98$ ) with level of service while in prison, and  $r=.43$  ( $n=119$ ) with early release from prison (e.g., parole) (Andrews, 1991).

Andrews, Kiessling, Mickus, and Robinson (1986) administered the LSI-R to a sample of 192 probationers. Andrews et al. (1986) reported mild to moderate correlations between the LSI-R subcomponents and other measures purported to measure the same construct (correlations ranged from  $\sim .20$  to  $\sim .50$ , with the highest reported correlation being that between the LSI-R subcomponent Criminal History with self-reported conviction and official recidivism measures). These findings were interpreted as support for the construct validity of the LSI-R.

### Criterion and Predictive Validity:

Most of the research on the LSI-R has focused on the relationship between the LSI-R and correctional outcome. Andrews (1991) reports that numerous outcome criterion have been predicted above chance levels by LSI-R scores, for example, early termination of

---

<sup>1</sup> Whether this particular sample comprised provincial or federal inmates is not specified.

probation, early closures of probation files, in program recidivism, reincarceration, success in correctional halfway houses, etc. Correlations between LSI-R total scores and recidivism for both probationers and inmates at 1-year follow-ups are in the .29 to .41 range. In three separate studies, LSI-R scores were predictive of institutional misconducts and assaults (Bonta & Motiuk, 1992). A recent study reported that violent offenders, defined as those having committed at least one major offense (murder, manslaughter, assault, kidnapping, forcible confinement) or serious offense (robbery, sexual offense), obtained statistically significant greater mean scores on the LSI total score relative to non-violent offenders (Loza & Simourd, 1994). Violent offenders had statistically significant greater mean scores on the LSI total score ( $M=27.4$ ,  $SD=9.1$  versus  $M=23.9$ ,  $SD=10.9$ ,  $F[1,159]=4.95$ ,  $p\leq.05$ ). Violent offenders also attained statistically higher subtotal scores on Alcohol/Drug ( $M=4.2$ ,  $SD=2.6$  versus  $M=3.1$ ,  $SD=2.4$ ,  $F[1,140]=6.69$ ,  $p\leq.05$ ); Family/Marital ( $M=1.9$ ,  $SD=1.3$  versus  $M=1.3$ ,  $SD=1.4$ ,  $F[1,126]=5.11$ ,  $p\leq.05$ ); Leisure/Recreation ( $M=1.5$ ,  $SD=0.8$  versus  $M=1.2$ ,  $SD=0.9$ ,  $F[1,156]=5.08$ ,  $p\leq.05$ ); and Emotional/Personal ( $M=2.7$ ,  $SD=1.6$  versus  $M=1.4$ ,  $SD=1.6$ ,  $F[1,147]=24.01$ ,  $p\leq.001$ ).

### **Self-Report Version of the LSI-R (Motiuk, Motiuk, & Bonta, 1992)**

The SRI was developed based on the LSI-R. The 78-item questionnaire will be included as a validity measure for the present study.

#### **Psychometric Review :**

##### **Sample:**

The SRI was evaluated on 100 male volunteer provincially-sentenced inmates (i.e., those serving sentences of  $\leq 2$  years).

##### **Reliability Estimates:**

All of the SRIs for the study sample were hand scored using templates derived from the LSI scoring guide, achieving strong interrater agreement .

### Concurrent Validity:

The SRI subcomponents demonstrated good correlations ( $r = .41$  to  $.80$ ) with their LSI-R counterparts (with the exceptions of financial ( $r = .22$ ,  $p \leq .05$ ) and attitude/orientation ( $r = .12$ , ns).

### Convergent Validity:

The correlational relationships between SRI items and a variety of measures that purport to tap the same or similar sub-components were moderate (ranging from  $.30$  to  $.60$ ).

### Predictive Validity:

SRI total scores were examined for predictive criterion validity in relation to institutional performance measures. Institutional adjustment was measured by the number of misconducts and the number of assaults. Significant partial correlations, controlling for number of days in custody, were found for prison misconduct ( $r = .17$ ,  $p \leq .05$ ) and assaults ( $r = .19$ ,  $p \leq .05$ ). The relationship between post-release recidivism and SRI total scores was significant for both parole violation ( $r = .29$ ,  $p \leq .05$ ) and reincarceration ( $r = .26$ ,  $p \leq .01$ ). Both the SRI and LSI were associated with assaults ( $X^2 = 6.26$ ,  $p \leq .05$ ;  $X^2 = 8.51$ ,  $p \leq .01$ ), and reincarceration ( $X^2 = 4.91$ ,  $p \leq .05$ ,  $X^2 = 6.24$ ,  $p \leq .05$ ). The Relative Improvements Over Chance measure (RIOC) for the prediction of misconduct were 44.5% and 22.7%, for the SRI and LSI, respectively.

Multiple regression analyses were used to explore the incremental predictive criterion validities of the SRI and LSI, with prison misconduct, assault, halfway house outcome, parole violations, and reincarceration as separate criterion variables. Only for halfway house outcome (success/failure) did the LSI demonstrate any incremental validity relative to the SRI.

### **The Psychopathy Checklist-Revised (Hare, 1991)**

The PCL-R is a rating scale for the assessment of psychopathy in forensic and psychiatric populations. The PCL-R measures both behavior and personality traits that are characteristic of the criminal psychopath. The scale contains 20 items which are rated on a 3-point ordinal scale, based on information obtained from file review and administration of a semi-structured interview. The PCL-R is included in this study because empirical findings indicate that PCL-R scores are predictive of violent re-offending. VRS total scores were correlated with PCL-R scores.

#### **Psychometric Review:**

Psychometric evaluations of the two versions indicate that the PCL and the PCL-R have virtually the same psychometric properties and external correlates, are highly correlated, and can be considered as measures of the same construct (Hare, 1991). Parallel/alternate forms reliability between total score on the PCL and the PCL-R is  $r=.88$ , although when corrections are made for the inherent unreliability of the scales the correlation is found to lie between .95 and 1.0 (Hare, Harpur, Hakstian, Forth, Hart, & Newman, 1990; cited in Fulero, 1995).

#### **Demographics of Validation Samples:**

Although few in number, studies suggest that the PCL-R retains its reliability and validity across racially distinct groups (e.g., Kosson et al, 1990; Wong, 1984). The PCL-R has not been validated with female offenders.

#### **Validation Samples**

The PCL-R validation samples comprised volunteer offender participants who were incarcerated in a variety of institutional settings ranging from minimum to maximum-security classifications ( $N=1632$ ).

### Interrater Reliability - PCL-R Total Scores

Inter-rater reliabilities were high ( $\sim .90$ ) using the intraclass correlation coefficient across samples (Hare, 1991).

### Internal Consistency

Cronbach's *alpha* coefficients were high for both the pooled prison and forensic patient samples ( $r=.87$  and  $r=.85$ , respectively).

### Test-Retest Reliability

There have been few investigations of the PCL-R's test-retest reliability. The only available data comes from a very small sample ( $N=10$ ) of opiate addicts attending a methadone clinic. The subjects were assessed at intake, and subsequently reassessed by a second rater (blind to the intake assessment) after a period of one month. The test-retest reliability coefficient was high ( $r=.94$  - Cacciola, Rutherford, and Alterman, 1990; cited in Hare, 1991).

### Validity Estimates:

#### Content and Concurrent Validity:

The manual provides an extensive literature review attesting the claim that the items contained in the PCL-R are generally consistent with traditional and current clinical views on the personality traits and behaviors that define the construct of psychopathy (e.g., International Classification of Diseases, Sartorius, Jablensky, Cooper, & Burke, 1988, cited in Hare, 1991; Cleckley criteria, Cleckley, 1976).

Psychopathy and anti-social behavior assessment instruments were used to evaluate the concurrent validity of the PCL-R (e.g., DSM-III-R criteria for ASPD, MMPI, Pd, CPI, MCMI and MCMI-II antisocial scales). Hare (1991) reported that measures that purport to tap personality traits associated with psychopathy correlate higher with Factor 1 whereas anti-social behavior measures are more strongly associated with Factor 2. Global ratings of psychopathy produced PCL-R total score correlations ranging from  $r=.80$  to  $r=.90$  and produce higher Factor 1 values ( $r\sim .80$  -  $.87$ ) relative to Factor 2 values ( $r\sim .65$  -  $.74$ ).

PCL-R Factor 2 items are more strongly correlated with DSM-III-R diagnosis of Anti-Social Personality Disorder than are Factor 1 items (.83 vs. .40 respectively). There is an asymmetric relation between PCL scales and APD, at least in forensic populations: A PCL-R diagnosis of psychopathy is more predictive of APD than APD is of psychopathy. Hare et al. (1990) compared the effect size of the PCL-R and diagnoses of APD in studies using institutional behavior, violence (in prison), parole outcome, and violent recidivism as dependent variables. The mean effect size was considerably larger for PCL-R total scores ( $r=.46$ ) and diagnoses of psychopathy ( $r=.44$ ) than it was for diagnoses of APD ( $r=.28$ ).

In addition psychopathy-related self-report scales that have APD sub-scales were correlated with the PCL-R (e.g., MMPI sub-scales PD, MA, PD+MA, PD-So) and, although the correlations were in the low to moderate range, their differential relationships with PCL-R- Factor 1 and Factor 2 items were as expected, namely higher correlations with Factor 2 than Factor 1 were reported . It must be noted that self-report inventories are of limited usefulness in the diagnosis of psychopathy, particularly in forensic populations (Hare, 1991).

### Predictive Validity

The predictive efficacy of the PCL-R was assessed using official recidivism as the dependent variable within a two-year follow-up (Serin, 1996). The sample comprised 81 minimum and medium risk offenders<sup>1</sup> who had been granted unescorted temporary absences (UTAs) in 1984-85 and subsequently released. Non-violent offenses included property crimes. Violent offenses comprised robbery, assault, manslaughter, sexual assault, and murder. The mean PCL-R score was 22.1 (SD=6.7). PCL-R Factor 1 scores correlated with both general and violent recidivism ( $r=.14$  and  $r=.26$ ,  $p\leq.01$ , respectively). There was a stronger relationship between PCL-R Factor 2 scores and general recidivism

---

<sup>1</sup>The demographic characteristics of the sample were consistent with the region in which the study was conducted.

than violent recidivism ( $r=.36$ ,  $p\leq.05$  and  $r=.22$ ,  $p\leq.01$ , respectively). Total PCL-R scores correlated  $r=.31$ ,  $p\leq.01$  with general recidivism and  $.28$ ,  $p\leq.01$  with violent recidivism.

The overall failure rate of the sample was 57%. General failure rates were 40% for nonpsychopaths ( $n=20$ ), 51.2% for the mixed group ( $n=41$ ) and 85% for the psychopaths ( $n=20$ ). Overall the violent recidivism rate was 10%. None of the nonpsychopaths recidivated violently however, 7.3% of the mixed group, and 25% of the psychopaths recidivated violently (Serin, 1996).

To date, with respect to the prediction of violent recidivism, the PCL-R has demonstrated the strongest predictive power with respect to the prediction of violent recidivism. The use of the PCL-R as a violent risk instrument however is an extension of its original intent, namely to assess the construct of psychopathy. As such, by design the use of the PCL-R as a violence-risk instrument is limited.