Carleton University

A Field Study on Static-99R Interrater Reliability by Canadian Police Personnel

By Heather Burke, MA, PhD Candidate

Department of Psychology

Date: 2024-09-25

Acknowledgments

The points of view expressed in this presentation do not necessarily represent the views of the Edmonton Police Service. This presentation draws on research supported by the Social Sciences and Humanities Research Council.

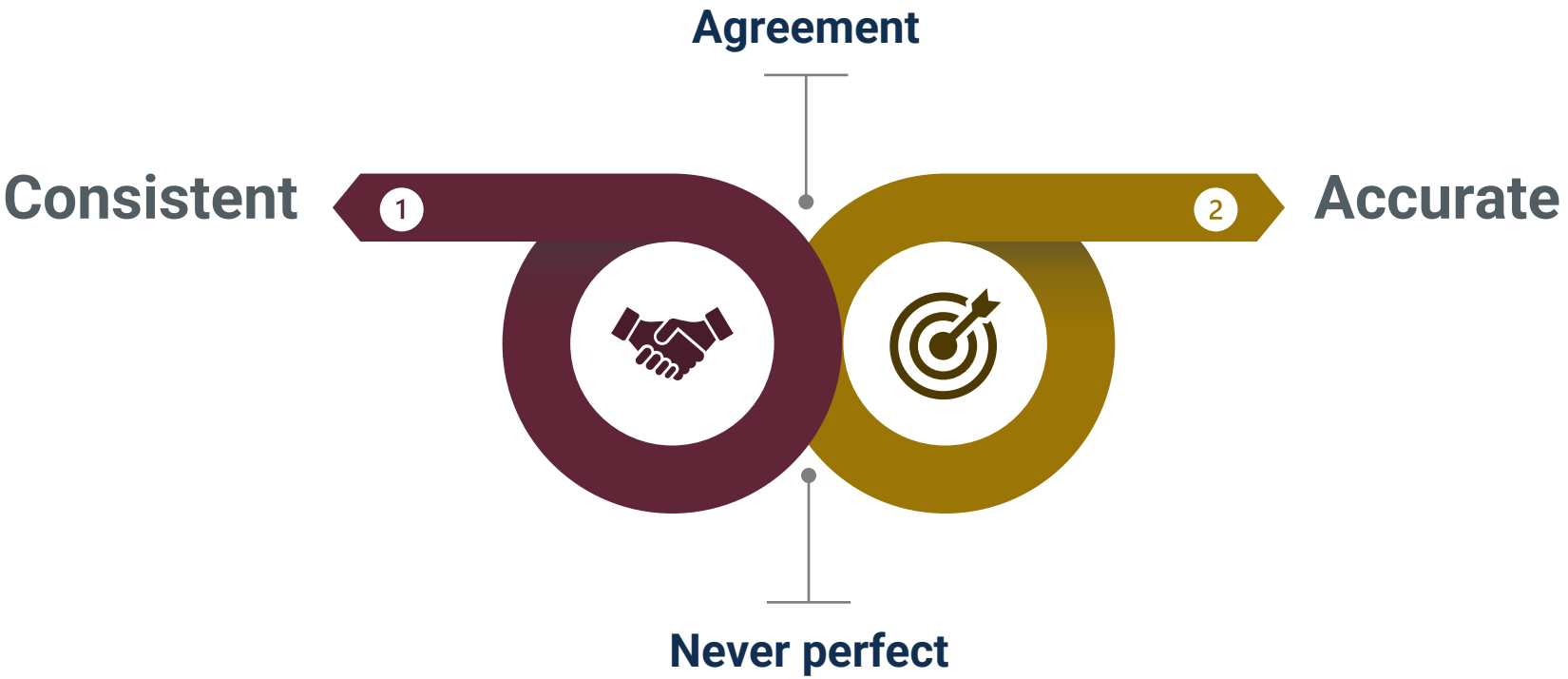
Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

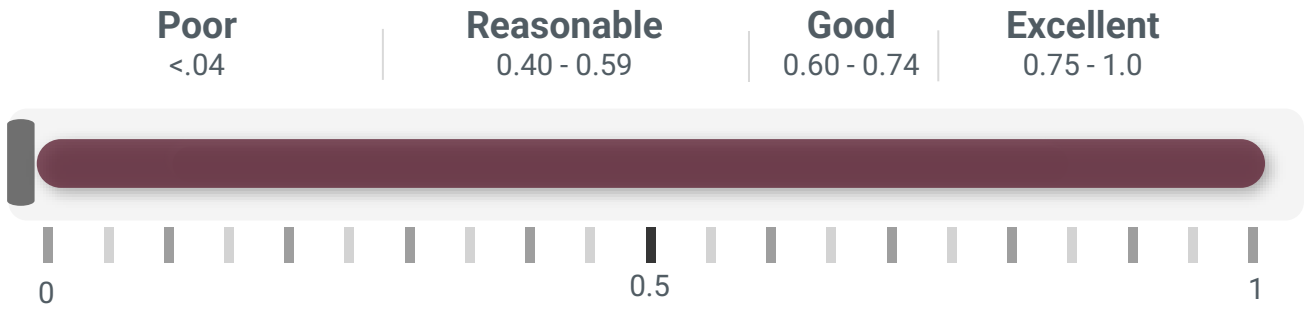
Introduction

Interrater Reliability



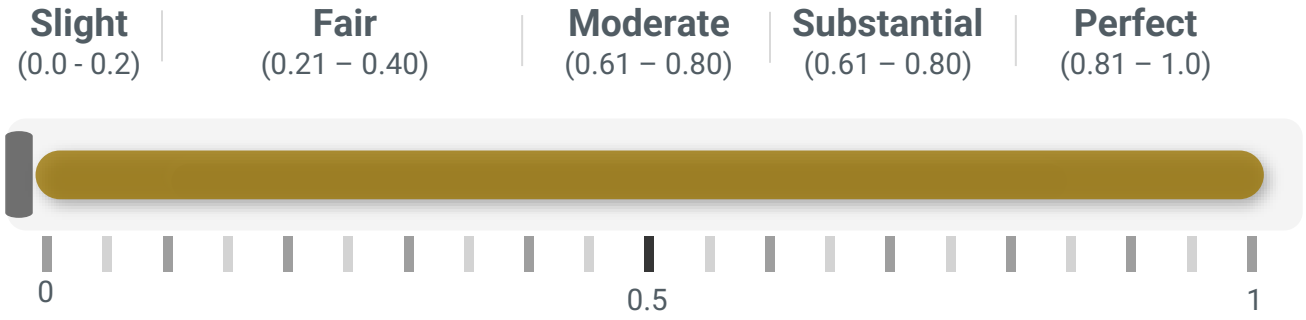
Interrater Reliability Statistics

Total Score



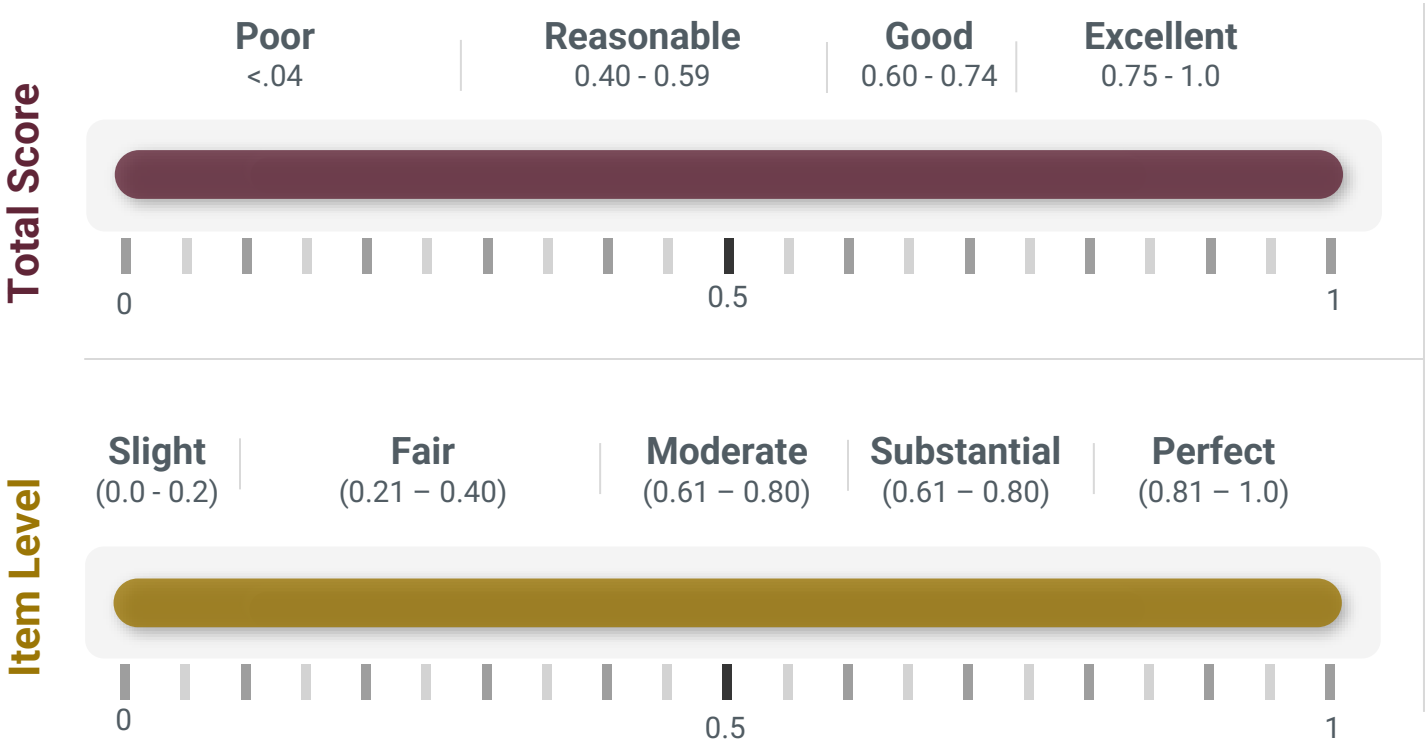
Intraclass Correlation Coefficient (ICC)

Item Level



Cohen's Kappa

Interrater Reliability Statistics



Percent Agreement

SEM (Precision)

Risk Assessment in Policing

01

Traditional Practice

02

Gap in Policing

03

Tools in Policing

04

Potential Solution

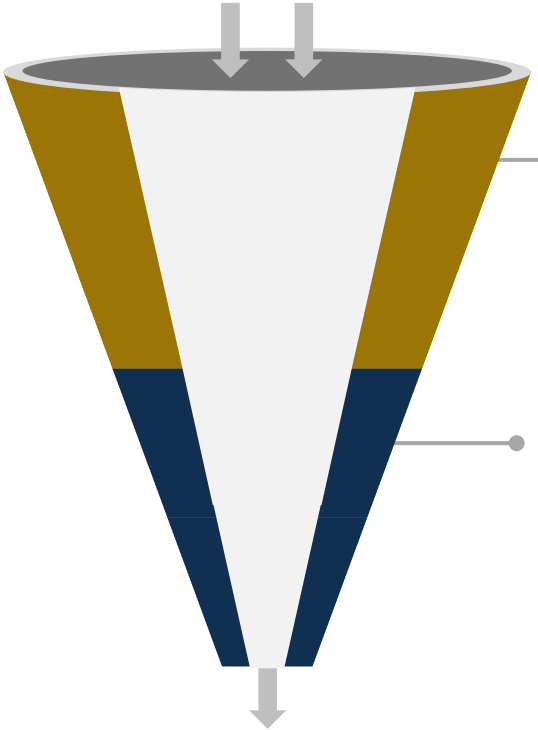
Static-99R

- **File information**
- **No advanced degree**
- **Total score:** -3 to 12, reflecting five nominal risk levels

Total Score	Risk Level
-3, -2	I – Very low
-1, 0	II – Below average
1, 2, 3	III – Average
4, 5	IVa – Above average
6+	IVb – Well above average

Static-99R Scoring

Index Clusters

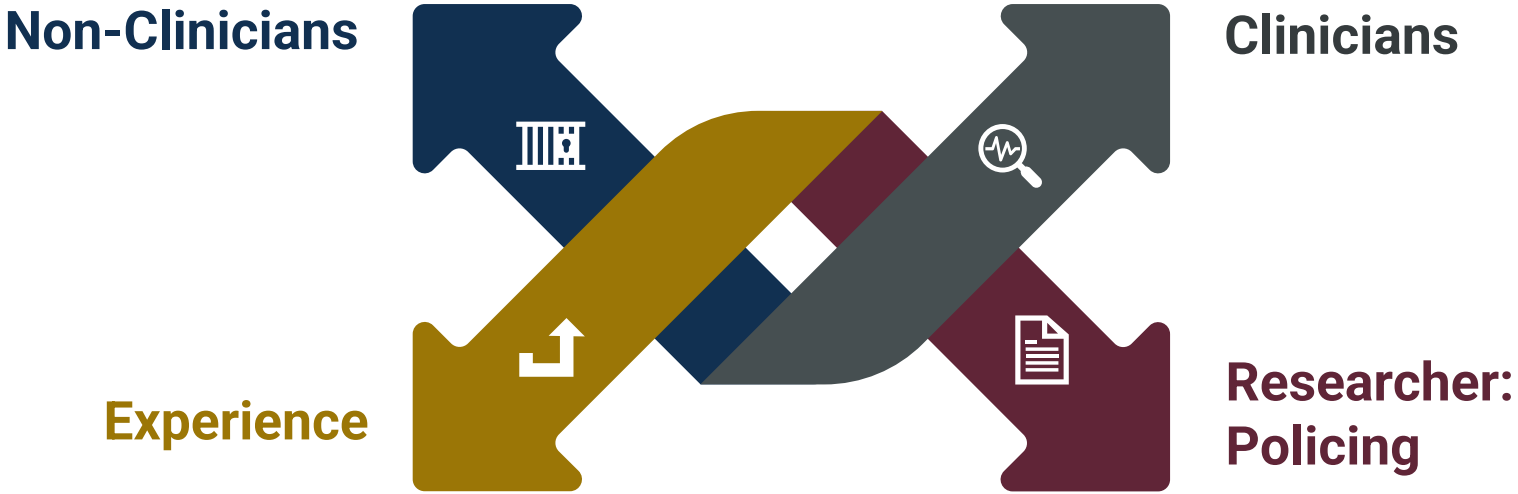


Offences before detection

Pseudo-recidivism

Consolidated Index Cluster

Static-99/99R Interrater Reliability Field Studies



Research Objectives

01 **Interrater reliability**

02 **Score discrepancies on risk categorization**

03 **Training comparison**

04 **Post-training analysis**

Evaluators and Static-99R Training

Formal Training

- Led by Static-99R co-creator R. Karl Hanson
- **Attendees:** 11 police personnel

1

- 1 graduate level
- 1 undergraduate level

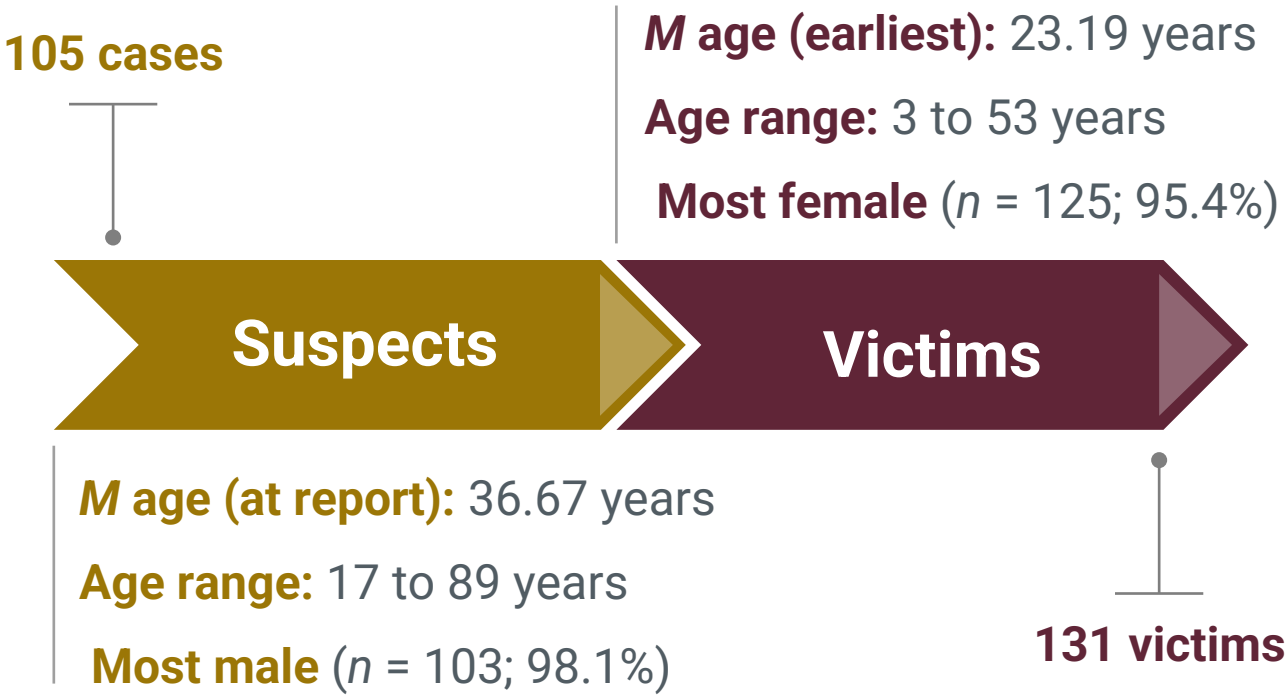
Researchers

2

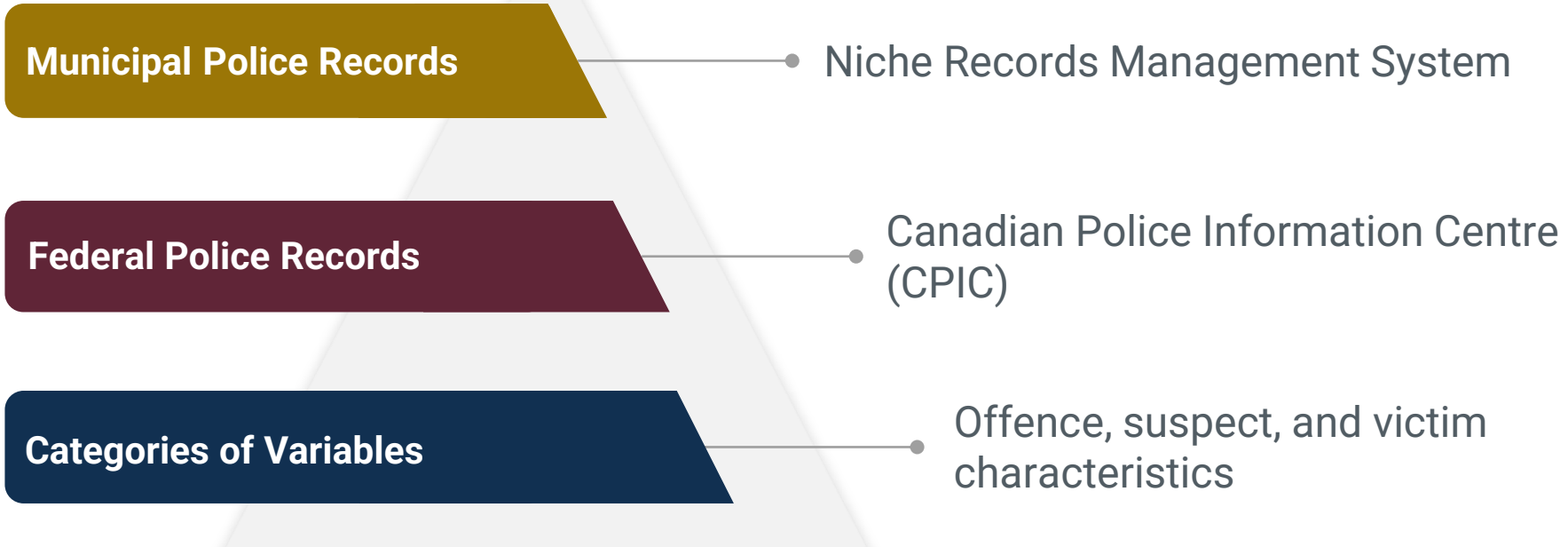
- 8 **formally** trained officers
- 2 **informally** trained officers
- 1 formally trained administrator

Researcher/ Police

Case Sample


105 cases
M age (earliest): 23.19 years
Age range: 3 to 53 years
Most female (*n* = 125; 95.4%)
Suspects
Victims
M age (at report): 36.67 years
Age range: 17 to 89 years
Most male (*n* = 103; 98.1%)
131 victims

Sources of Information



Procedure

07 – Researcher B

06 – Researcher A

05 – PDF Retrieval

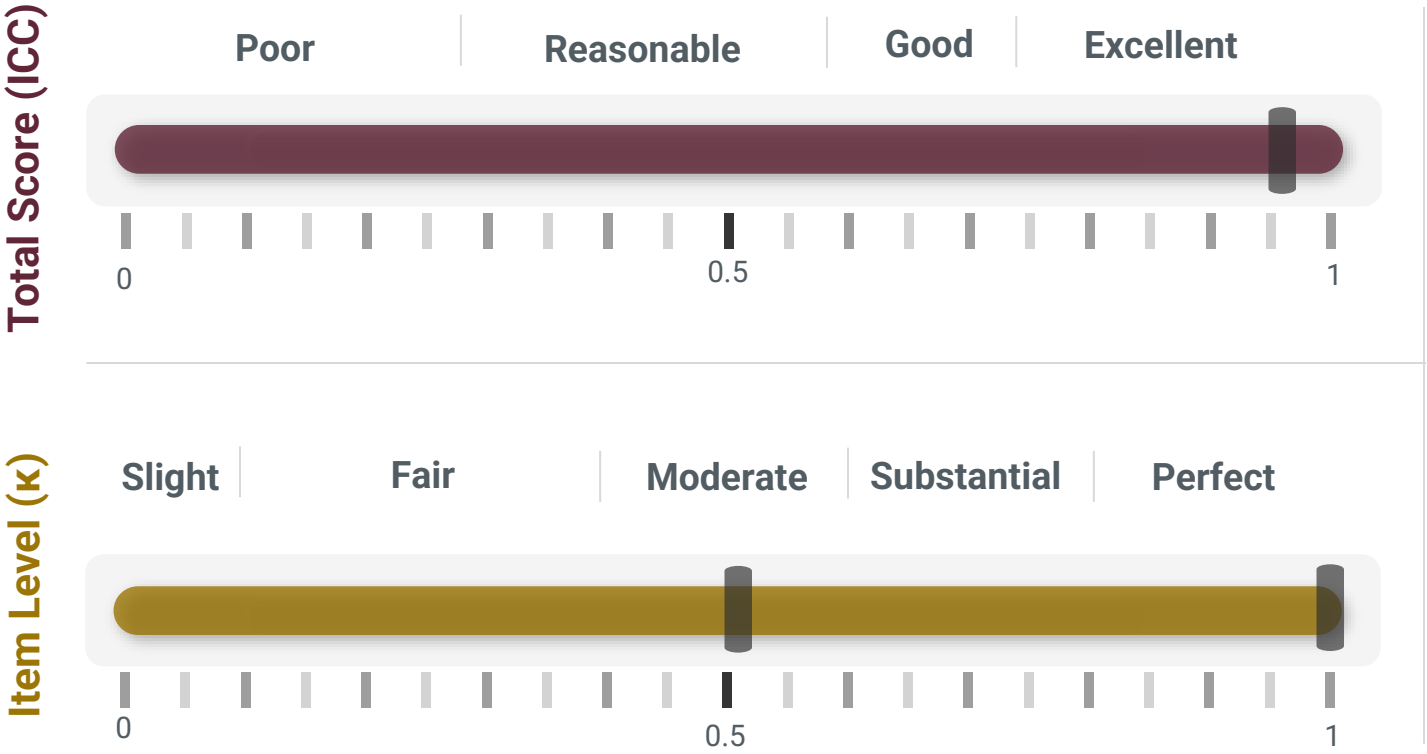
01 - Ethics

02 – Training

03 – Police Coding

04 – Case Logging

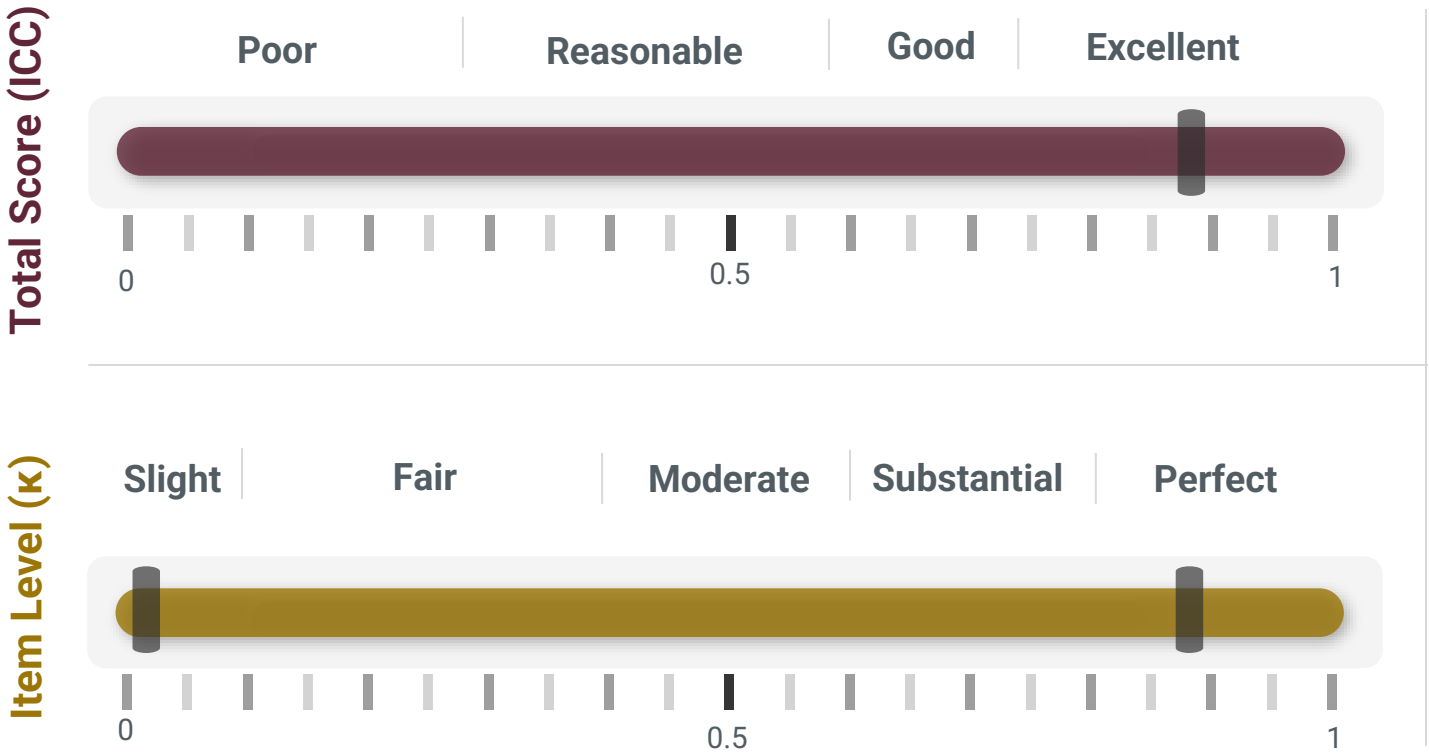
Agreement Between **Researchers**



- **80%** exact agreement
- **SEM** = .493

- **90% – 100%** agreement

Agreement Between a **Researcher and Police**



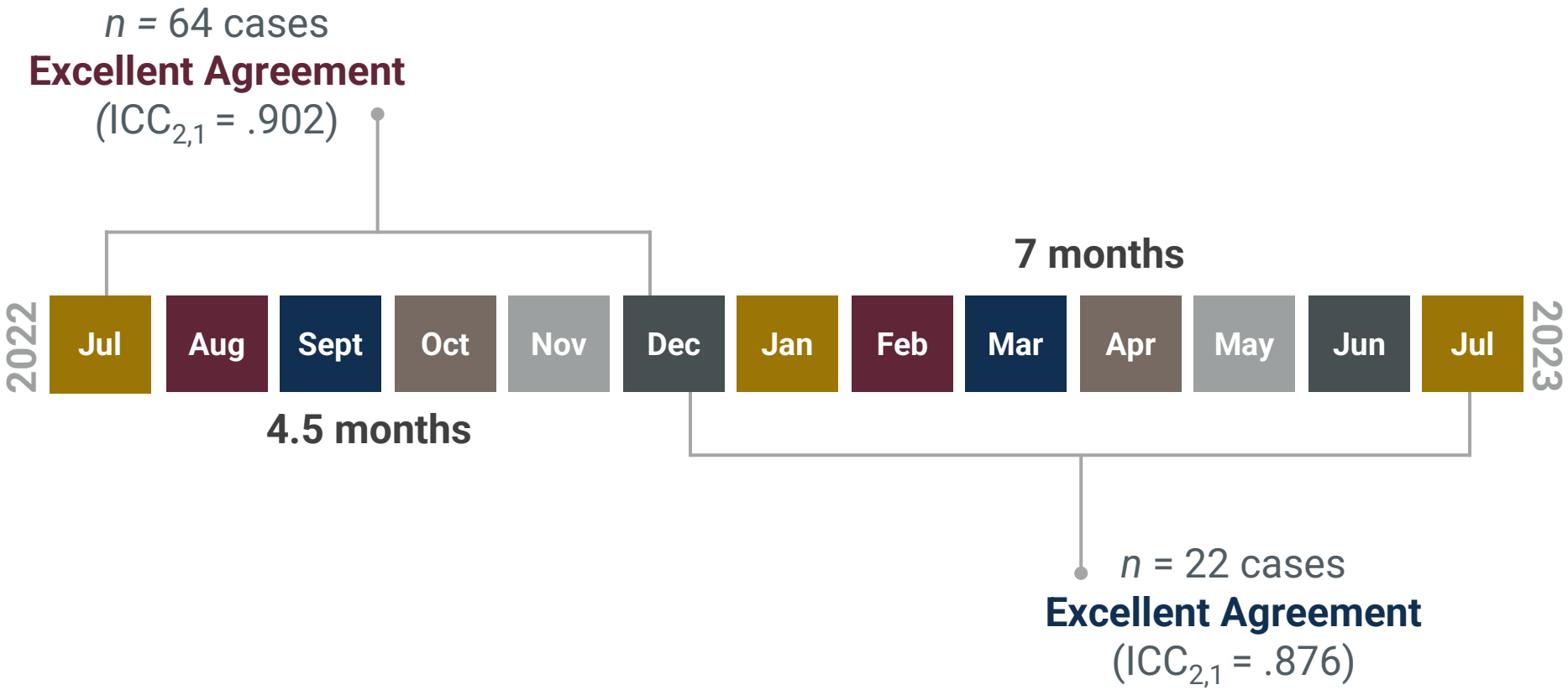
- **47%** exact agreement ↓
- **SEM** = .752 ↑
- **24%** risk difference

- **72%** discrepancy
- **60% — 96%** agreement ↓

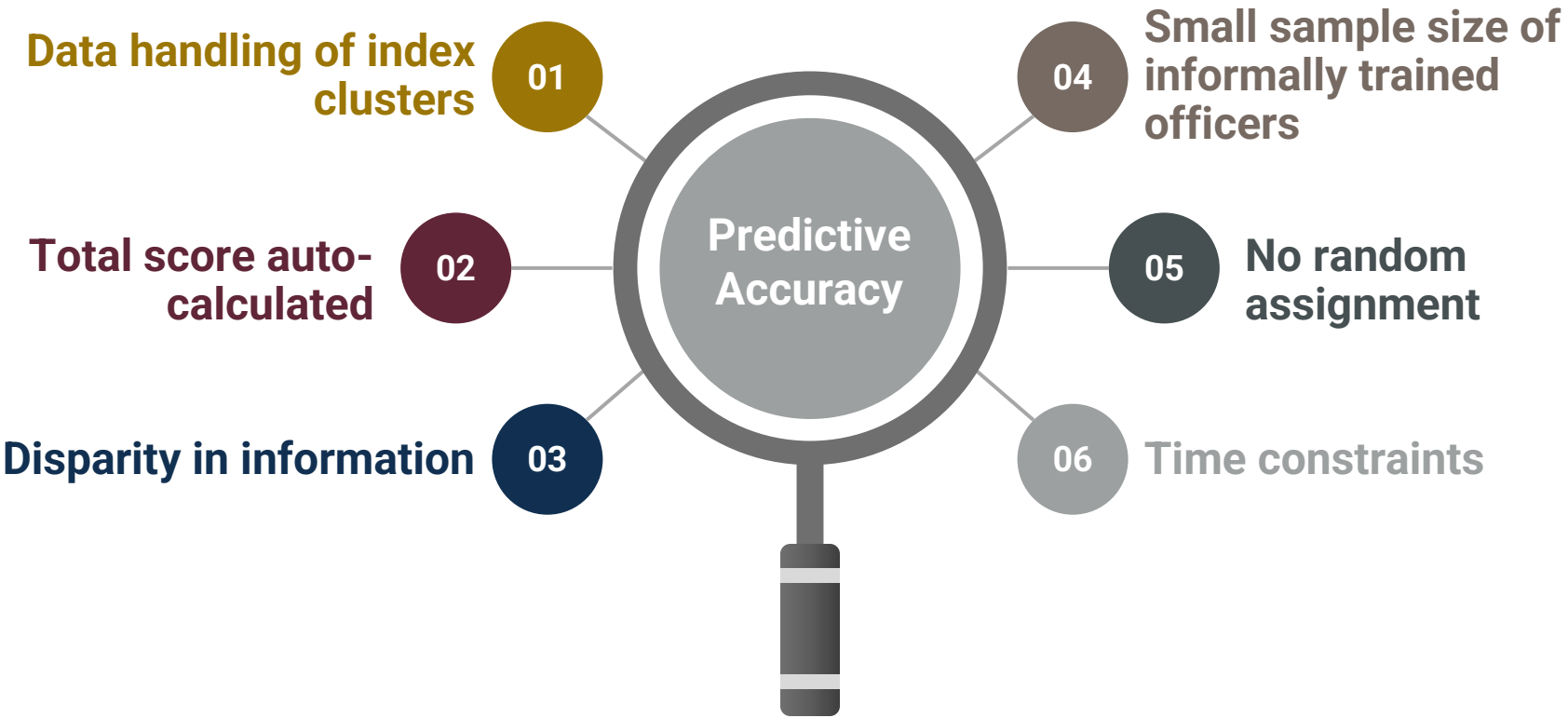
Comparing Training Methods

Formally Trained (84 cases)	VS	Informally Trained (18 cases)
Excellent ($ICC_{2,1}$ = .913)	Total Score Agreement	**Excellent** ($ICC_{2,1}$ = .884)
44%	Identical Total Score	67%
75% (*n* = 63)	Item Discrepency	56% (*n* = 10)
44%: 21 of 47 cases	Risk Discrepecy	67%: 4 of 6 cases
SEM = .679	Precison	SEM = .785

Total Score Reliability Among Police After **Formal Training**



Limitations and Future Directions



Conclusions

- Crucial **first step**
- Findings **generally support** the reliable use of the Static-99R
- Possible **cost-effective** training methods
- **Foundation for** future research on **predictive accuracy**

THANK YOU!

Heather Burke, Carleton University

heatherburke3@cmail.carleton.ca