



Centre for Forensic
Behavioural Science



Do risk assessments describe a feature of the person, or just their risk?: Why this question matters for research and practice

Dr Benjamin Spivak

A/Prof Caleb Lloyd, Dr Ariel Stone

23 November 2024

42nd ANZAPPL Congress, Melbourne

Sorry!

“Measurement is an incredibly dry and boring topic...” (Fried, 2018)

Fried, E. Measurement Schmeasurement. <https://eiko-fried.com/measurement-schmeasurement/>



What are risk assessments designed for?

- ▲ Risk assessments are designed to estimate the likelihood of offending behaviour (e.g. crime, violence, sexual violence etc.) and to support risk management
- ▲ Come in several different flavours:
 - ▲ Actuarial risk assessments involve mathematically combining responses to items to create an index that estimates probability of re-offending
 - ▲ Structured professional guidelines involve humans considering a range of risk factors and making a judgement about probability of re-offending

Taking risk assessments beyond prediction

- ▲ In recent years, some have suggested that risk assessments can also be used to measure underlying psychological attributes (e.g. deviance, criminality, anti-sociality, etc. etc.)
- ▲ If risk assessments measure psychological attributes we can then use them to make inferences about people's psychological features.
- ▲ This could be useful for scientific development of forensic psychology, risk management and for guiding intervention

Taking risk assessments beyond prediction

“In the absence of a construct, there will not be an explicit direction for the improvement of estimating risk.”

“Incorporating a risk-based construct, which involves hypothesis testing and an explanation of behavior, could provide risk-related areas for clinical intervention”

Kroner, D.G., Mills, F.J., & Reddon, J.R. (2005). A coffee can, factor analysis, and the prediction of antisocial behaviour: The structure of criminal risk. *International Journal of Law and Psychiatry*, 28(4), 360-374, doi: 10.1016/j.ijlp.2004.01.011

What is measurement in psychology

- ▲ We hypothesise the existence of a psychological construct (e.g. pro-criminal attitude)
- ▲ This psychological construct is hypothesised to range along a dimension that can correspond to numbers
- ▲ We can measure the construct through a set of items that reflect differing levels of the construct

Higher criminal attitudes



Lower criminal attitudes

Pro criminal attitude items
Celebrates whenever a crime is committed
Has criminal heroes
Has disdain for the law
Supports the bad guys on police shows
Loves speeding in their car

Purported psychological constructs measured in risk assessments

STATIC-99R/STATIC-2022R (Brouillete-Alarie et al., 2016)



SAVRY (Hilterman et al., 2016)



START (Whittington et al., 2016)



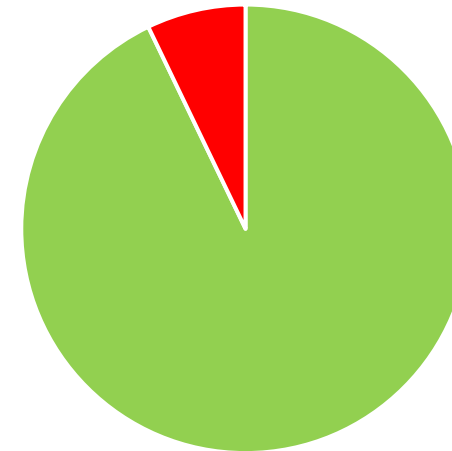
How good are we at finding measures?

My UNsystematic review

- ▲ I examined 14 publications that aimed to find underlying latent constructs from risk assessments
- ▲ 13 out of 14 found evidence to support measurement
- ▲ What a strike rate!

It took physicists 300 years to successfully measure temperature. So why are we so successful?

Proportion of papers successfully finding evidence of construct measurement



■ Success ■ Not success



This is a broader pattern within psychology

“...the success rate of asserted measurement approaches 100%. Even if one takes into account that the available evidence of measurement is affected by a publication bias, this record is remarkable.”

“Latent variables, like attitude or satisfaction, are arguably more complex than length or temperature. And still the measurement of psychological attributes does not seem to be more challenging than measuring physical attributes.”

(Salzberger, 2013)



Why do we keep finding measures?

- ▲ Are we that much cleverer than those in the hard sciences?
- ▲ Are psychological constructs easier to measure than physical constructs?
- ▲ Are we using an approach that is almost certain to guarantee results?



How do we currently establish measurement with risk assessments?

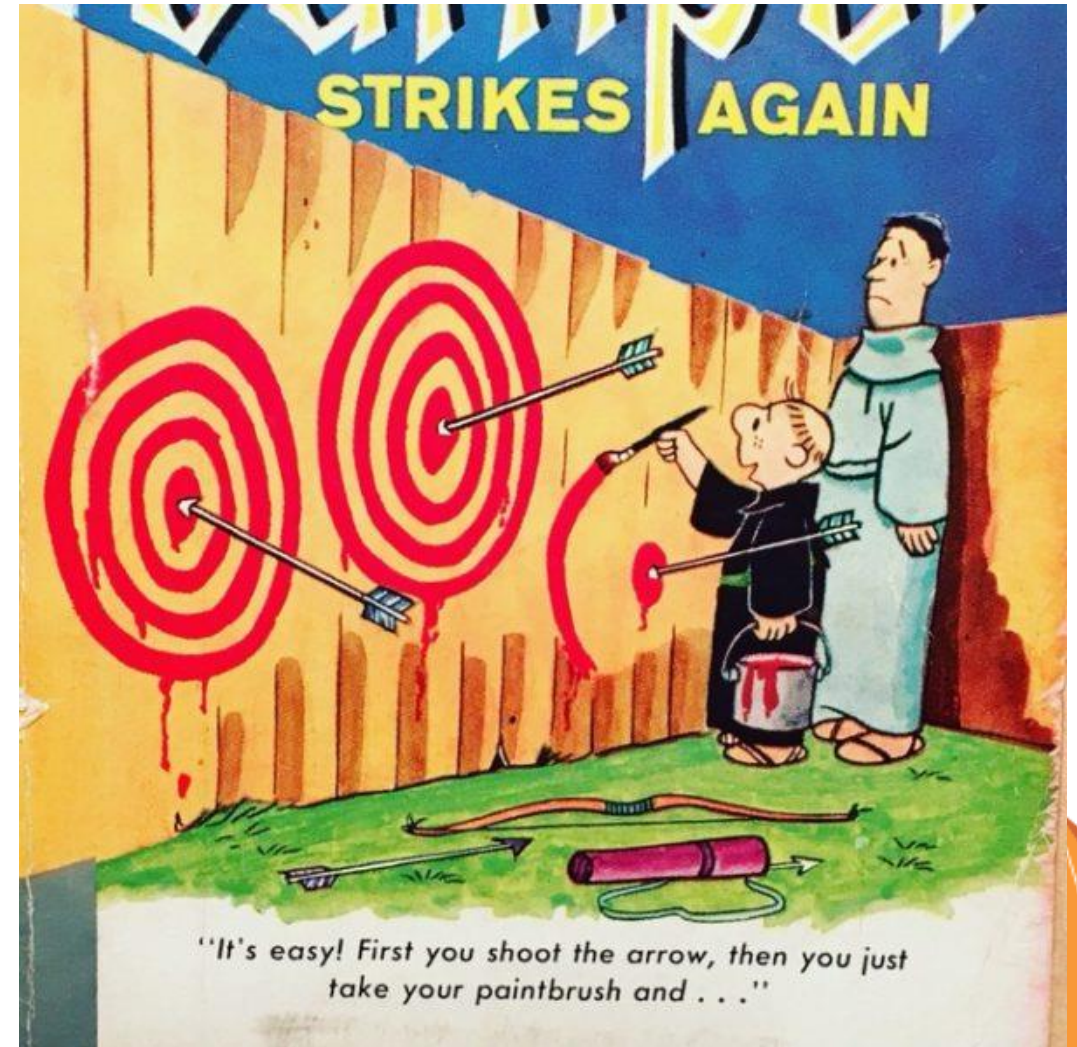
- ▲ Data driven process - Risk assessments are not designed for measurement, so selection of items is not guided by a-priori theories about latent constructs and their likely structure.
- ▲ Instead items on a risk assessment are subject to statistical approaches like factor analysis that find clusters of items that are correlated with one another.
- ▲ These clusters of items are then provided with a name and a claim is made that summing the items together measures a construct.



Why is the data driven approach problematic?

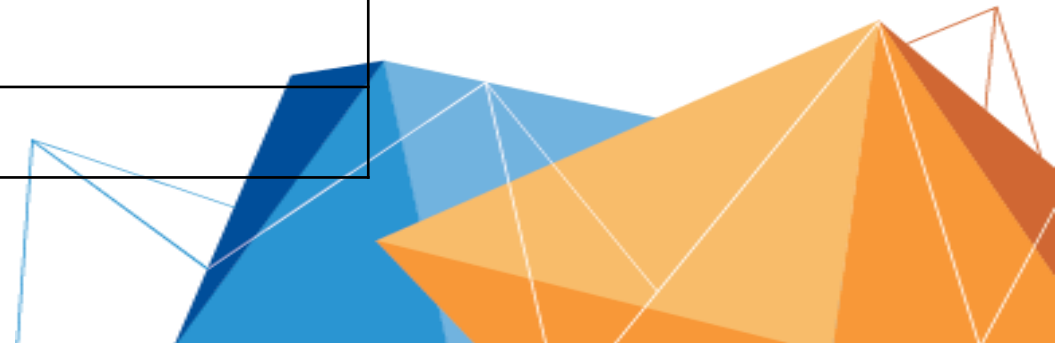
“A factor analysis...of my age, the population of Mexico, the price of swiss cheese, my pet turtle’s weight and the average distance between galaxies during the past ten years will yield a strong [factor]” (Gould, 1994)

Letting the data lead your theory is like painting a bullseye around your arrow.

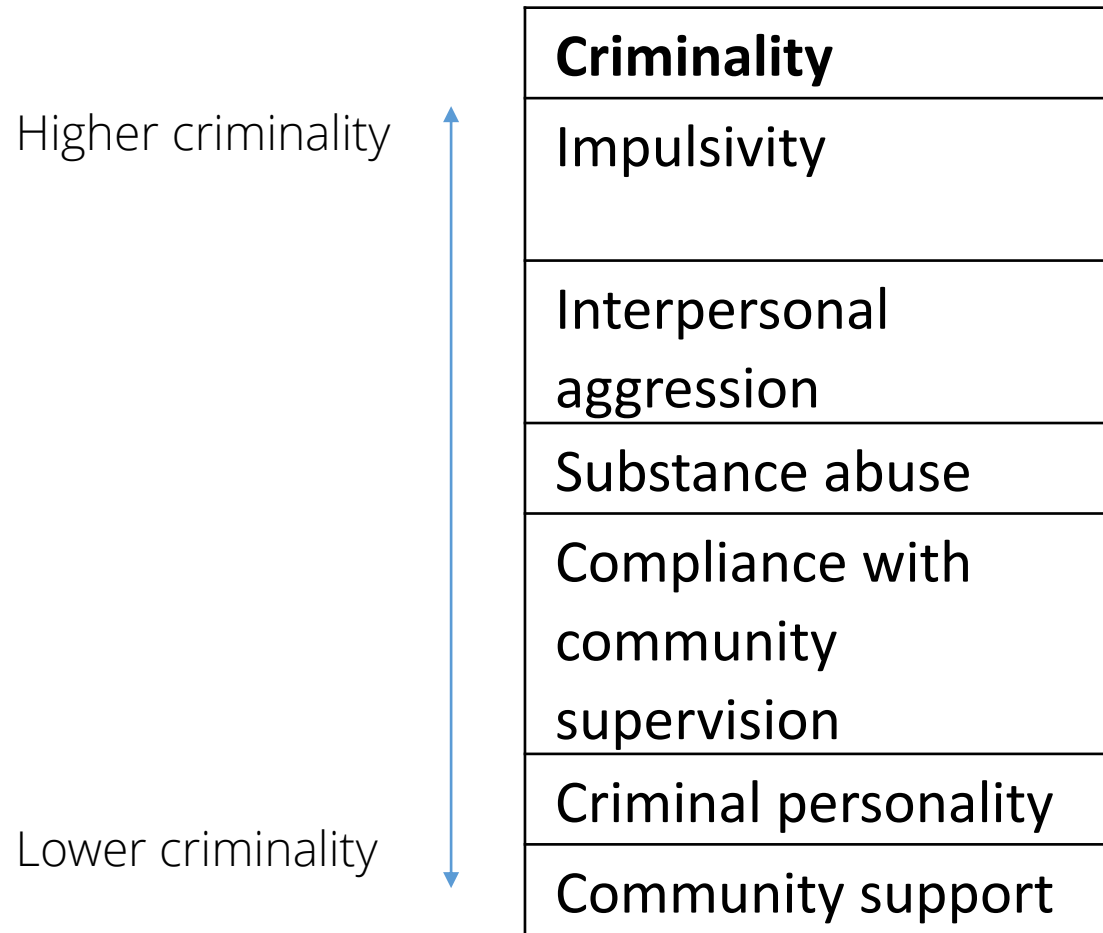


The results of a data driven approach to measurement

Deviance	Criminality	Treatment responsivity
Sexually deviant lifestyle	Impulsivity	Insight
Deviant sexual preferences	Interpersonal aggression	Treatment compliance
Offense planning	Substance abuse	Cognitive distortions
Sexual offending cycle	Compliance with community supervision	Released to high risk situations
Sexual compulsivity	Criminal personality	
	Community support	



The results of a data driven approach to measurement

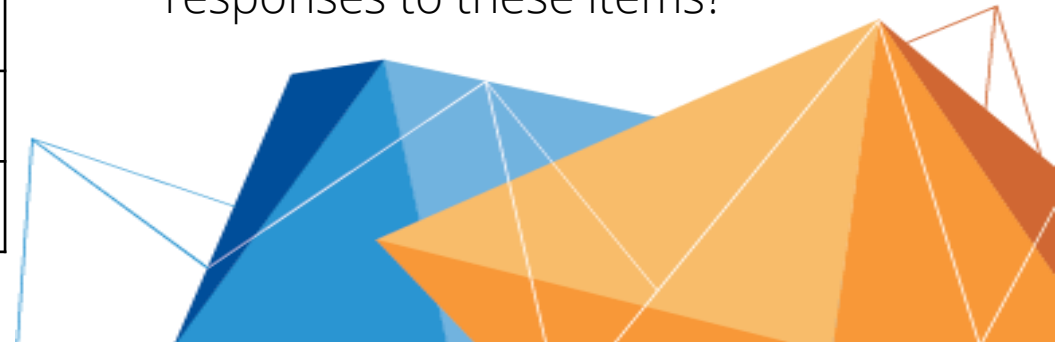


What is criminality?

Do the items reflect the entire range of criminality among the population (sex offenders)?

Which items denote higher vs lower levels of criminality?

How does the latent construct of criminality cause different responses to these items?



The results of a data driven approach to measurement

Deviance	Criminality	Treatment responsivity
Sexually deviant lifestyle	Impulsivity	Insight
Deviant sexual preferences	Interpersonal aggression	Treatment compliance
Offense planning	Substance abuse	Cognitive distortions
Sexual offending cycle	Compliance with community supervision	Released to high risk situations
Sexual compulsivity	Criminal personality	
	Community support	

The results of a data driven approach to measurement

Mathematical proficiency
Algebra
Order of operations (PEMDAS/BODMAS)
Arithmetic (addition, subtraction etc.)
Counting
Number recognition

Higher proficiency

Lower proficiency

Higher criminality

Lower criminality

Criminality
Interpersonal aggression
Compliance with community supervision
Substance abuse
Impulsivity
Criminal personality
Community support



What is wrong with our current approach?

- ▲ A lack of theory development means that we are unsure exactly how items relate to different levels of the construct
- ▲ Intercorrelations between items do not ipso facto support claims of measurement.
- ▲ Resulting measurement scores have no meaning – a score of 5 for ‘deviance’ or ‘criminality’ does not refer to any specific qualitative state.
- ▲ As a result we can provide no substantive meaning to what it means to be relatively low or high on a measure.



Where does this leave us?

- ▲ Decide whether we actually need to measure psychological constructs?

“I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it...your knowledge is of a meagre and unsatisfactory kind” (Lord Kelvin in Thomson, 1981, p 80-81)

“Measurement need not be a universal characteristic of science. Instead what characterizes science is the method of observation” (Michell, 1997)



Where does this leave us?

- ▲ If we want to measure, let's start with theory development:
 - ▲ What psychological constructs are hypothetically quantitative?
 - ▲ How can we use items to elicit different levels of the construct?
 - ▲ What sorts of external phenomena lead to an increase or decrease in levels of the construct?



Where does this leave us?

- ▲ Let's ask ourselves difficult empirical questions about the measures we create
 - ▲ Let's ask how our measures can be disproven rather than asking what evidence would be needed to provide support.
- ▲ E.g. how well does the scale order people on your hypothetical attribute?
- ▲ E.g. Do external factors result in a particular level of change on that scale?



Why does this matter?

- ▲ Research waste
- ▲ Inappropriate reification of test results can lead to poor policy, inappropriate treatment
- ▲ Slows down scientific progress on understanding the nature of psychological aspects of offending behaviour



Risk assessments are still useful for prediction

- ▲ The prospects of uncovering measures through existing risk assessments are poor.
- ▲ BUT
- ▲ We don't need measures for risk assessments to be useful.
- ▲ We can use age and cholesterol to predict heart attacks without resorting to measurement claims.
- ▲ Similarly we can use risk assessment scores to predict CJS involvement without resorting to measurement claims.



If we want to measure then how do we get there?

- ▲ Before theory
 - ▲ Read what experts in measurement have to say about the topic (highly recommend Michell for a non-technical introduction)
- ▲ Take time to develop theory
 - ▲ Develop a theory which allows you to hypothesise about the structure of measurement e.g. order and additivity.
- ▲ Test your theory
 - ▲ Collect data and subject your theory to appropriate hypothesis tests which can falsify your measurement model.

