

Developing a Faking-Resistant Measure of Corporate Psychopathy for Use in Employment
Selection: The CRT-WP

by
Ryan Cook

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Masters of Science in Applied Psychology –
(Industrial/Organizational)

August, 2019, Halifax, Nova Scotia

© Ryan Cook, 2019

Approved by: Dr. Nicolas Roulin
Supervisor

Approved by: Dr. Skye Stephens
Internal Examiner

Approved by: Dr. James O'Brien
External Examiner

Date: August 13th, 2019

Abstract

Developing a Faking-Resistant Measure of Corporate Psychopathy for Use in Employment
Selection: The CRT-WP

by Ryan Cook

Abstract: The issue with using self-report measures of psychopathy in selection is that they are highly susceptible to faking. Conditional Reasoning Tests (CRTs) are a rarely used method for measuring implicit personality that could be a solution to this problem. The Conditional Reasoning Test for Workplace Psychopathy (CRT-WP) is conceptualized in this paper. The current research includes four studies which develop and validate the CRT-WP. Sixty items were generated for six justification mechanisms deemed inherent to psychopathic thinking. Study 1 used SMEs to refine the item list, Study 2 conducted a preliminary EFA, and Studies 3 and 4 validated the remaining items in MTurk and student samples. Overall, there is support the CRT-WP as a reliable measure with sufficient construct validity. However, more evidence is needed for predictive validity. It appears that the CRT-WP could overcome the issues with assessing psychopathy in selection. Application, design, and potential issues are discussed.

August 13th, 2019

Developing a Faking-Resistant Measure of Corporate Psychopathy for Use in Employment
Selection: The CRT-WP

by
Ryan Cook

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Masters of Science in Applied Psychology –
(Industrial/Organizational)

August, 2019, Halifax, Nova Scotia

© Ryan Cook, 2019

Approved by: Dr. Nicolas Roulin
Supervisor

Approved by: Dr. Skye Stephens
Examiner

Approved by: Dr. James O'Brien
Examiner

Date: August 13th, 2019

Table of Contents

Table of Contents	4
Developing a Fake-Resistant Measure of Corporate Psychopathy for Use in Employment	
Selection: The CRT-WP	7
Psychopaths in the Workplace	7
A Brief History of Psychopathy’s Measurement	12
Psychopathy Checklist-Revised and Self-Report Psychopathy scale.....	13
Levenson Self-Report Psychopathy scale.....	14
Psychopathic Personality Inventory – Revised	14
Triarchic Psychopathy Measure	15
Newer Measures	15
Measuring Exclusively Corporate Psychopathy.....	16
Psychopathy Measure – Management Research Version.....	16
Business Scan 360	17
Logical Inference Exercise	18
The Overall Issue – Faking	19
Implicit Measurement through Conditional Reasoning Tests.....	23
Identifying Psychopaths in the Selection Process with the CRT-WP.....	28
Development of the CRT-WP.....	28
Justification Mechanisms	29
Externalization.....	30
Carefree Impulsivity.....	30
Social Superiority	30
Fearlessness	33
Ruthless Self-Interest.....	33
Insensitivity	34
Item Generation.....	35
Study 1 – Item and Scale Revision with Subject Matter Experts	37
Methods.....	37
Results	39

Discussion	41
Study 2 – Exploratory Factor Analysis and Initial Reliabilities	43
Method.....	44
Sample	44
Procedure and Analyses.....	47
Results	49
Discussion	52
Study 3 – Two-Wave Study with Confirmatory Factor Analysis and Validity Assessment	54
Methods	57
Sample	57
Measures	59
Results	61
Wave One	61
Wave Two.....	62
Discussion	65
Study 4 – Additional Validation with a Student Sample	68
Methods	70
Sample	70
Measures	71
Results	73
Discussion	75
General Discussion.....	76
Implications	82
Limitations and Future Research.....	84
Conclusion.....	88
References	89

List of Tables

Table 1 - Measures of Psychopathy and their Issues with Use in a Selection Context	20
Table 2 - List of Justification Mechanisms Produced for the Development of the CRT-WP	31
Table 3 - Content Coverage of the Justification Mechanisms for the CRT-WP in Relation to Main Measures	32
Table 4 - Agreement among SMEs for the Original List of 60 CRT-WP Items	42
Table 5 - Percentages of Response Selections for 43 Items in Study 2 and their Result After the EFA Process.....	48
Table 6 - Fit indices for 1-5 Factor Models in Final EFA of Study 2 (27 items)	51
Table 7 - Factor Loadings for 27-Item, 2-Factor Model Resulting from Study 2	51
Table 8 - Descriptive Statistics, Reliability Coefficients, and Correlations for the 27-Item CRT-WP in Study 2	52
Table 9 - Descriptive Statistics, Reliability Coefficients, and Correlations for Study 3 – Wave 1	61
Table 10 - Fit indices for Four Competing CFA Models in Study 3 – Wave One (27 items)	62
Table 11 - Descriptive Statistics, Reliability Coefficients, and Correlations for Study 3 – Wave 2	63
Table 12 - Hierarchical Regressions Testing CRT-WP Scores Predicting CWBs Beyond the SRP-III and TriPM.....	65
Table 13 - Descriptive Statistics, Reliability Coefficients, and Correlations for Study 4 Variables	74
Table 14 - Summary of Hypotheses.....	78

List of Appendices

Appendix A - Expanded Details of Psychopathy Measures	110
Appendix B - CRT-WP Initial Item List.....	117
Appendix C - Subject-Matter-Expert Task	118
Appendix D - Revised Item List Following Study 1 SME Revision	121
Appendix E - Items for Mahmut et al.'s (2011) adapted version of the SRP-III.....	122
Appendix F - Items for the Brief Triarchic Psychopathy Measure (TriPM; Patrick, 2010)	124
Appendix G - Items for the Counterproductive Work Behavior Checklist (CWB-C; Spector et al., 2006)	126
Appendix H - Items for the HEXACO-60 (Ashton & Lee, 2009).....	128
Appendix I - Adapted Items for Academic Dishonesty/Cheating based on McCabe and Trevino (1993)	130

Developing a Fake-Resistant Measure of Corporate Psychopathy for Use in Employment Selection:
The CRT-WP

Manipulative, deceitful, arrogant, insensitive, remorseless, cold-hearted, egocentric, risky, and parasitic; these are some of the unpleasant adjectives commonly used to describe psychopaths (Boddy, Ladyshevsky, & Galvin, 2010). Psychopathy, which is a stable trait that differs between individuals, has been studied by researchers in psychology, sociology, criminology, and other academic disciplines. Psychopaths in the workplace have been of particular interest to the field of psychology over the past two decades (Babiak, 2007; Smith, Watts, & Lilienfeld, 2014). These individuals are normally referred to as corporate psychopaths, industrial psychopaths, successful psychopaths, professional psychopaths, or workplace psychopaths. “Corporate” and “industrial” are typically only used to describe those in white-collar professions, though only “workplace” and “corporate” will be used further in this paper.

Psychopaths in the Workplace

Perhaps the very first to acknowledge the psychopath as a businessman was Cleckley (1941) in his classic book, *“The Mask of Sanity,”* which is generally credited with the modern-day conceptualization of psychopathy. However, the interest in workplace psychopathy truly began with Babiak in 1995, who is largely considered the “father” of modern workplace psychopathy just as Hare is to modern forensic psychopathy. Babiak (1995) gives an academic case study of one successful psychopath, a 30-year old male, who stood out as being different from the criminal psychopath. This enterprising, white-collar male was described as charismatic, influential, and giving very favourable first impressions (Babiak, 1995). Whilst studying him in-depth, Babiak noted that this psychopath was manipulative, coercive, oppressive, and generally willing to do whatever needed to get his way. This corporate psychopath was not able to remain in a positive light with everyone, as

perceptions of him varied depending on how frequently people interacted with him and how “useful” they were to him (Babiak, 1995). Those who interacted with him more frequently had more negative perceptions of him, and those who were of least “usefulness” to him also had very unfavourable perceptions probably due to there being no purpose for him to be nice or charming to those individuals. Babiak (1995) deduced that successful psychopaths still have the same fundamental traits as other psychopaths, but they lack most of the anti-social deviant behaviour commonly associated with forensic psychopaths (i.e., violent crime, lack of control) *or* they behave these ways in different channels (i.e., behind office doors) which permits them to get caught less often. In fact, five years later, he suggested a five-phase process explaining how psychopathic individuals find success in business (Babiak, 2000).

Since the turn of the century, there has been an incredible surge in the popularity of corporate psychopathy, both in the literature and the media. Smith and Lilienfeld (2013) demonstrate that the amount of media articles about these “corporate monsters” and “snakes in suits” is beginning to dwarf the number of research articles being published about workplace psychopathy. The main issue with this is that there is an increasing risk of the construct being contaminated by the pop-culture conceptualization of psychopathy (Caponecchia, Sun, & Wyatt, 2011). For example, in a study where subordinates rated their supervisors’ levels of psychopathy, 26% of supervisors were categorized as psychopaths based on the measure used (Boddy, 2011). This is problematic considering that the field-wide estimated percentage of workers who are psychopaths is 1% to 3%. This overestimate was likely a result of very low threshold for psychopathy among the general population. Any supervisor who is perceived as a bully, or any supervisor who fires employees “too” quickly, may be classified as a psychopath by laypersons (Caponecchia et al., 2011). Researchers have illustrated that regardless of actual differences in supervisor behaviour, perceptions of being bullied led to higher use of the “psychopath” label when describing supervisors. This is an issue as

perceptions of supervisor psychopathy alone are associated with decreased job satisfaction and other important variables (Boddy, 2011; Boddy et al., 2010; Caponecchia et al., 2011; Mathieu, Neumann, Hare, & Babiak, 2014).

In response to Babiak's original work, other researchers in the field wondered how and why these corporate psychopaths differed from the forensic type of psychopath, which is still a contested topic to this day. There are currently three competing models which attempt to answer why these are two unique types (Lilienfeld, Watts, & Smith, 2015). First, the moderated-expression model states that the "successful" type of psychopath results from an atypical manifestation of psychopathy, where the core traits and behaviours have been mitigated by factors such as high levels of intelligence or effective parenting. In supporting this model, more than one study has provided evidence that education and a good "family background" can influence someone being a successful psychopath instead of an "unsuccessful" criminal psychopath (Boddy et al., 2010; Blicke & Schütte, 2017). Steinert, Lishner, Vitacco, and Hong (2017) state that the moderated-expression model is the most promising of the three because it does not suffer from the theoretical or empirical issues that the other two models do. Next, the differential-configuration model posits that workplace psychopaths have some different/additional traits compared to forensic psychopaths. In support of this model, researchers have shown that high conscientiousness and boldness are critical to successful psychopathy (Lilienfeld et al., 2015), and another recent study has provided evidence that the differential-configuration model is empirically superior in comparison to the other two models (Vergauwe, Wille, Hofmans, & De Fruyt, 2019). Note that these two models already mentioned are not mutually exclusive. Finally, the third is the differential-severity model, which essentially states that corporate psychopathy is just a mild "sub-clinical" expression of psychopathy. This model appears to have the least empirical support (Lilienfeld et al., 2015). Alternatively, some researchers suggest that the difference between successful and criminal psychopathy results from

neurobiological differences, in that the “unsuccessful” criminal psychopaths have cognitive or emotion regulation deficits (Gao & Raine, 2010).

More is known about what corporate psychopaths prefer and what about their psychopathy is adaptive for them. Henley (2002) examined non-criminal psychopaths and their career interests. Henley found that psychopathic individuals prefer risky work activities, a solitary work style, a desire to control or lead, and enjoy collecting wealth or prestige. More recently, Smith et al. (2014) provided additional support for successful psychopaths being drawn to law enforcement and military professions, along with business, politics, and contact sports. Henley (2002) acknowledges that successful psychopaths are not at all interested in careers with high dependency on others, careers that are objectively boring, or careers that are stereotypically female such as teaching or nursing. Concerning interest in politics, Lilienfeld et al. (2012) analyzed records and personality data for 42 Presidents of the United States and found that important traits of corporate psychopathy were correlated with better ratings of characteristics such as, performance, persuasiveness, crisis management, and leadership ability. This is just one example of a finding that some aspects of psychopathy can clearly be adaptive, and even beneficial, in some work contexts. It is unsurprising that superficial charm, risk-taking, and strategies for coercion could be associated with success in professions in where one is required to sell something (e.g., real estate agent), for example. Overall, it seems that boldness and fearlessness are the adaptive characteristics with the most generalizability (Blickle & Schütte, 2017; Hall et al., 2014).

Despite potential benefits for successful psychopaths, there are many negative outcomes associated with their stereotypical attitudes and behaviour. One intuitive shortcoming of psychopathic traits in the workplace is ethical decision-making. Stevens, Deuling, and Armenakis (2012) explains the process by which most psychopaths, regardless of their success, struggle with

responding to ethical dilemmas in the workplace, which usually results in unethical choices. A more objective performance deficiency for corporate psychopaths was demonstrated very recently when psychopathy levels were linked with consistent annual fund losses for hedge fund managers (ten Brinke, Kish, & Keltner, 2018). Beyond their own performance, the behaviours of corporate psychopaths are associated with many detrimental organizational outcomes. Unfair and abusive supervision is consistently related to job satisfaction and turnover intentions (Boddy, 2011; Mathieu & Babiak, 2015). Corporate psychopaths are also known to cause conflict and bully colleagues or subordinates, which is related to lower affective well-being, increased stress, and increased counter-productive work behaviours for employees (Boddy, 2014; Scherer, Baysinger, Zolynsky, & LeBreton, 2013). The relationship between psychopathy and bullying is mentioned, although the recent claim of psychopathic bullying being overestimated is acknowledged (Boddy & Taplin, 2017).

It is evident that although some specific aspects of psychopathy are beneficial for individual performance in some specific contexts, corporate psychopaths are generally associated with negative outcomes for coworkers and the organization as a whole. During the hiring process, organizations often select applicants based on cognitive ability and personality (e.g., conscientiousness) tests, which are valid predictors of job performance (Schmidt & Hunter, 1998). However, psychopathy has been consistently associated with negative consequences discussed in the literature and the media, and yet recent research reveals that psychopathy is rarely considered when making hiring or promotional decisions (Mathieu & Babiak, 2016b). It would be clearly beneficial to organizations if they could simply give job applicants a self-report psychopathy measure along with the other tests they give during the selection or promotion process. At minimum, there could be a significant reduction in many of these daily occurrences such as abusive supervision, unethical decision making, and manipulation. At maximum, we would be able to completely avoid hiring or promoting people with high levels of psychopathic traits into leadership roles or positions of power or authority over

the public (i.e., law enforcement) or clients (i.e., health care). Unfortunately, currently we cannot do this reliably and accurately at anything other than a high financial cost. The reason for this is that there is no self-report measure of workplace psychopathy that meets all the needs of the selection context (Lilienfeld et al., 2015). What is considered by most to be the best measure of psychopathy, requires a clinical professional and a multifaceted assessment which is too costly for most organizations in practice, and would likely lead to negative reactions from most applicants. Ideally, we would want a self-report measure of psychopathy to minimize costs and it would best serve selection (methodologically and legally) to have a measure specific to workplace/job-related psychopathy. There are many self-report measures that exist, the problem is that even most laypeople can and will engage in impression management and faking to present themselves a certain way to an employer (Levashina & Campion, 2007). This ability to fake becomes much more of a problem when the target of your measurement is characterized as deceitful, manipulative, coercive, and risk-taking, with a blatant disregard for rules and procedures; a psychopath.

A Brief History of Psychopathy's Measurement

Since researchers began measuring psychopathy, most of the work has focused on criminal offenders. As a result, many of the following scales were developed using criminal samples to measure forensic psychopathy. For a more detailed discussion and review of these criminal- and forensic-based psychopathy measures, please refer to Appendix A which includes information on how each measure was developed, greater detail on the psychometric structure, and the degree to which each has been supported by further research. However, when it became more common knowledge that psychopaths existed in workplaces, the measurement of non-criminal psychopathy became a separate focus of research. In the last decade, there have been measures devoted specifically to the workplace manifestation of psychopathy (Smith & Lilienfeld, 2013).

Psychopathy Checklist-Revised and Self-Report Psychopathy scale. Hare (1980) created the Psychopathy Checklist (PCL) which had 22 criteria. The psychopathy checklist requires professional clinicians to make ratings of each subject. The PCL later became the *Psychopathy Checklist – Revised* (PCL-R) and removed two of the criteria leaving a maximum score of 40. The recommended score for labelling someone a psychopath is 30/40 (though some recommend 25/40) which has been implicated as the cut-off score for greater risk of re-offense. It was later found that there is a clear two-factor structure in both the PCL and PCL-R. The first factor contains items that describe psychopaths as being selfish, remorseless, manipulative, and having certain attitudes and behavioural tendencies (i.e., interpersonal and affective deficits). The second factor consists of items that detail the psychopath's history, such as chronic instability, social deviance, and others (i.e., antisocial behaviour). A version of the PCL-R was adapted for both criminal *and* non-criminal use and was called the Psychopathy Checklist – Screening Version (PCL:SV; Hart, Cox, & Hare, 1995). The PCL:SV takes 12 of the 20 PCL-R items (eliminating the ones about deviant history) but still requires a professional to conduct the assessment.

Acknowledging the popularity of self-report measures, Hare (1985) translated the exact PCL factors and items into the *Self-Report Psychopathy scale* (SRP). The SRP has been revised multiple times over the years with the current adaptation being called the SRP-4 which has a long and short version (Paulhus, Neumann, & Hare, 2015). Since their origin, the PCL-R and SRP have been the most frequently used measures of psychopathy (Evans & Tully, 2016). However, they have many criticisms (Boduszek & Debowska, 2016; Boduszek, Debowska, Dhingra, & DeLisi, 2016; Drislane, Patrick, & Arsal, 2014; Forouzan and Cooke, 2005; Hall, Benning, & Patrick, 2004; Lilienfeld, 1994; Lilienfeld & Andrews, 1996; Skeem & Cooke, 2010).

Levenson Self-Report Psychopathy scale. The *Levenson Self-Report Psychopathy scale* (LSRP; Levenson, Kiehl, & Fitzpatrick, 1995) was created around the same time as Hare's SRP. The LSRP is modelled after the PCL-R as well, and contains 26 items covering the two factors of "primary" and "secondary" psychopathy. Similar to the PCL-R and SRP, the "primary" factor consists of the core fundamental psychopathy traits such as manipulation and selfishness, while the "secondary" factor measures antisocial behaviours. The LSRP was also developed with mainly male criminals (Levenson et al., 1995). It has remained almost completely unchanged from the original version, although it is still the focus of recent research which most seems to indicate that the psychometric properties have remained valid and reliable (Falkenbach, Poythress, Falki, & Manchak, 2007; Walters, Brinkley, Magaletta, & Diamond, 2008).

Psychopathic Personality Inventory – Revised. The *Psychopathic Personality Inventory* (PPI; Lilienfeld, 1990) was the first self-report measure of psychopathy, as it pre-dated the SRP and LSRP. Lilienfeld criticized the PCL-R's dependence on deviant/criminal behaviour as part of the assessment of psychopathy, as he argued that although these two types of behaviour were highly related, the former was not required by the latter. Since he believed that psychopaths exist in all contexts, the PPI measures the "core" psychopathy traits such as a lack of empathy, superficial charm, and others, without considering the illegal behaviours which are mainly associated with the forensic version of psychopathy. Thus, the PPI was designed specifically to measure psychopathy with non-criminal contexts in mind, though it has been used only second to the PCL-R in clinical contexts as well. The PPI later became the PPI-R (Lilienfeld & Widows, 2005) which consists of 154 items. The PPI-R has two factors, the first (fearless dominance) consists of more adaptive traits such as fearlessness and stress immunity, while the second (self-centered impulsivity) represents the dishonourable traits such as cold-heartedness and Machiavellian egocentricity. Total scores are used to determine levels of psychopathy. There is also a short version of the PPI-R which consists of 40

items. Researchers have found support for the psychometric properties of the PPI-R (Edens, 2004; Falkenback et al., 2007; Tapscott, Vernan, & Veselka, 2012). However, there have been others who have raised concerns about the structure of the PPI-R, its application, and its built-in detection scales (Kelley et al., 2016; Marcus, Church, O'Connell, & Lilienfeld, 2018; Tsang, Salekin, Coffey, & Cox, 2017). Specifically, Hall et al. (2014) recently fit the PPI-R into a three-factor structure that was consistent across criminal and student samples. This was further validated by Sellbom, Wygant, and Drislane (2015). This new three-factor structure has been generally approved of by the main PPI-R researchers (Lilienfeld et al., 2016).

Triarchic Psychopathy Measure. The *Triarchic Psychopathy Measure* was created based on a three-factor structure which persisted within the literature (TriPM; Patrick, 2010). The TriPM is a 58-item self-report inventory. The three factors inherent to the Triarchic model are Boldness (associated with more of the adaptive traits such as fearlessness), Disinhibition (associated with indices of deviant behaviours that are not necessarily criminal), and Meanness (associated with the more typical traits such as Machiavellianism, cold-heartedness, etc.). The TriPM is credited with integrating the different conceptualizations of psychopathy into one measure, with good construct validity (Drislane et al., 2014). The TriPM appears to be the most universally favoured measure of psychopathy in the current literature (Evans & Tully, 2016; Lilienfeld, 2016; van Dongen, Drislane, Nijman, Soe-Agnie, & van Marle, 2017).

Newer Measures. There are other new psychopathy measures that have not yet received much research, such as the *Elemental Psychopathy Assessment* (EPA; Lynam et al., 2011), the *Comprehensive Assessment of Psychopathic Personality* (CAPP; Cooke, Hart, Logan, & Michie, 2012), the *Psychopathic Personality Traits Scale* (PPTS; Boduszek et al., 2016), and the

Psychopathic Processing and Personality Assessment (PAPA; Lewis, Ireland, Abbott, & Ireland, 2017). Again, a more detailed look at these measures can be found in Appendix A.

Measuring Exclusively Corporate Psychopathy

Although some of these previous measures appear to remain valid for assessing psychopathy in non-criminal samples, they were not developed for a work context specifically. Further, there is uncertainty as to whether they would remain valid when there is something “on the line” like there is in the hiring process (Mathieu & Babiak, 2016b). Additionally, using most of the measures mentioned above to choose between applicants would have questionable legal defensibility due to them not having support for use in a non-criminal population. This number may be reduced even further when it is considered that any assessment using PCL-R-based measures, or even the PPI-R, could be construed as a medical or clinical examination (for mental health) which is prohibited unless given as part of an employment offer (Catano, Wiesner, & Hackett, 2016). Thus, if there were to be a selection tool for psychopathy to be used in the hiring process, it would most likely have to be developed for the work context specifically, or at least have some demonstrated predictive validity for job-relevant behaviours. Although the number of workplace-specific measures is low, some do exist. Smith and Lilienfeld (2013) recently reviewed psychopathy in the workplace and identified only three, and a current EBSCO literature search with “*psychopath or psychopathy or psychopathic*” and “*measure or measurement or assessment or scale or inventory or tool*” in research titles yielded no newer measures.

Psychopathy Measure – Management Research Version. Boddy (2011) wanted to identify corporate psychopaths using a workplace-specific model of the PCL-R, resulting in the *Psychopathy Measure – Management Research Version* (PM-MRV). The PM-MRV is an 8-item, observer report measure, which uses similar criteria to factor 1 of the PCL-R. It uses the same rating

procedure as well, in that the observers (usually coworkers) rate each criterion from zero to two based on how much the target displays that behaviour. A total score of 12/16 is recommended as a cut-off score to classify someone as a psychopath. It could be argued that this is as close as one can get to a PCL-R type measure tailored to the corporate psychopath, but perhaps it is *too* close. Jones and Hare (2016) mentioned that the measure unjustly uses items that are proprietary to the PCL-R. Jones and Hare also allege that the PM-MRV selected items from the PCL-R that do not match any particular factor structure without any theoretical grounds, and that the PM-MRV simply does not measure psychopathy more than it does any other “negative” personality construct. The PM-MRV also allows for over-classification as laypeople often use the term “psychopath” too generally in reference to their supervisors and managers (Boddy, 2011; Caponecchia et al., 2011). The PM-MRV should be used and interpreted with extreme caution (Smith & Lilienfeld, 2013).

Business Scan 360. In response to a lack of measures for psychopathy at work, two of the most notable names in psychopathy research created the *Business Scan 360* (B-Scan 360; Babiak & Hare, 2012). Since Hare is one of the developers, it is not surprising that the B-Scan 360 is based on the PCL-R factor structure. First, the researchers created 200 items describing psychopathic behaviours based on critical incidents and stories. Using data from an online Mechanical Turk (MTurk) sample, these items were then rated and only those relevant to psychopathy in the workplace were kept (Mathieu, Hare, Jones, Babiak, & Neumann, 2013). A factor analysis indicated that 113 items divided into six factors in previous analyses. However, the researchers decided to eliminate two of the six factors because they were inconsistent with the PCL-R conceptualization, which is questionable considering that these two factors were arguably the most relevant to workplace outcomes (defined as “disruptive behaviour” and “ability” in reference to one’s ability to function at work). Additionally, the researchers thought it would help the perception of the B-Scan to reduce it to 20-items. Thus, the final version of the B-Scan 360 kept the “best” five items for the four

factors that the researchers decided to keep. Again, this is an observer report measure where subordinates or colleagues rate the target on a scale of zero to two for each item. In order to better understand why the developers of the B-Scan 360 made these decisions to reduce the scale, Smith and Lilienfeld (2013) asked them for the data which was used to create the B-Scan. The authors of the B-Scan 360 did not provide Smith and Lilienfeld with their data.

The B-Scan has received some validation, however, it could be considered concerning that it only comes from the creators of the measure. For example, Mathieu and Babiak (2015) were able to link ratings on the B-Scan 360 with increased turnover intentions, job satisfaction, and abusive supervision. Mathieu and Babiak (2016b) also developed the B-Scan Self, a self-report variant of the B-Scan 360. The B-Scan Self was designed to be “business friendly” and “non-clinical” so that it would be less intensive and more appropriate for organizations to use in the hiring process. Mathieu and Babiak state that the B-Scan Self consists only of items that will be accepted by organizations (no questions about personal history), but that it retains the same rigour of the PCL-R. They also provide preliminary evidence for reliability and validity of the B-Scan Self with online samples. The B-Scan measures have received minimal coverage by other researchers thus far.

Logical Inference Exercise. The *Logical Inference Exercise* (LIE; Gustafson, 2000) is unique for psychopathy measurement. The LIE is an implicit measure that is designed to detect “aberrant self-promotion” which is a facet of psychopathy that may be particularly salient in corporate psychopaths. The LIE shows brief vignettes to respondents who then answer a series of questions based on what they viewed. It appears to be a test of reasoning ability, but it implicitly measures biases in thinking that are inherent to psychopathy (i.e., superiority over others). When this measure was reviewed by Smith and Lilienfeld (2013), it was still in the preliminary stages of its development. The LIE seemed to have promise for measurement in the business world, as it was

inexpensive and close to self-report, yet still mostly protected from faking. For whatever reason, the LIE has not been mentioned in the literature since then.

The Overall Issue – Faking

As mentioned before, there is demonstrated utility in a tool that can assess work-related psychopathic personality in the selection process, and researchers have been calling for one (Mathieu & Babiak, 2016b; Smith & Lilienfeld, 2013). For any position that requires managing other employees or access to sensitive materials and resources, assessing psychopathic personality in the selection process would help organizations avoid many work-related issues and behaviours that research has associated with the corporate psychopath. Consider the selection process for a top leadership position within law enforcement or the military; although some adaptive characteristics of psychopathy may lend themselves positively to performance in these positions, the presence of a corporate psychopath is a net negative overall (Boddy, 2011; Smith & Lilienfeld, 2013). It seems that even after considering all the non-criminal and workplace-specific measures produced to date, there is not one that is best suited for use in the hiring process (Mathieu & Babiak, 2016a). Table 1 briefly summarizes the issues associated with each measure. Note that the LIE and those included under the “newer measures” section are not included in the table as there is minimal published research on them, to my knowledge.

There are two evident trends in the issues presented in Table 1. First, the measures that require someone to rate the target are either too costly in terms of resources for most positions, or are seemingly questionable considering that untrained and unstandardized laypeople are used to assess a trait in others that they likely do not fully understand. The B-Scan 360 and the PM-MRV also have limited utility in screening and selection since the designated rater would not have enough familiarity with a new job applicant’s behaviours and tendencies to give a valid assessment.

The second issue that surrounds the self-report measures is that all of them are susceptible to faking. The PPI-R is the only of one these self-report measures to include validity scales that measure virtuous and deviant responding, while the rest present each item at face value with the hope that no one is attempting to artificially lower their scores by faking. The clear problem with this is

Table 1

Measures of Psychopathy and their Issues with Use in a Selection Context

<i>Measure</i>	<i>Target Use</i>	<i>Issues for Use in Selection</i>
1. PCL-R	Forensic	- Conceptualized and validated using criminals, which have demonstrated differences from corporate psychopaths - Requires expert rater and lengthy/invasive assessment
2. SRP	Forensic and Non-Forensic	- Conceptualized and validated using criminals, which have demonstrated differences from corporate psychopaths - Susceptible to faking and social desirability
3. PPI-R	Non-Forensic and Forensic	- Conceptualized with mainly students and “community” samples, which likely do not capture workplace specifics - Susceptible to faking, although detects social desirability
4. LSRP	Forensic and Non-Forensic	- Conceptualized and validated using criminals, which have demonstrated differences from corporate psychopaths - Susceptible to faking and social desirability
5. TriPM	Non-Forensic and Forensic	- Conceptualized with mainly students and “community” samples, which likely do not capture workplace specifics - Susceptible to faking and social desirability
6. PM-MRV	Corporate	- Uses untrained raters to make assessments of others - The rater must have familiarity with the target, therefore it is unlikely it could be used to rate a new applicant
7. B-Scan 360	Corporate	- Uses untrained raters to make assessments of others - The rater must have familiarity with the target, therefore it is unlikely it could be used to rate a new applicant
8. B-Scan Self	Corporate	- Conceptualized using PCL-R framework of psychopathy which was based on criminals - Susceptible to faking and social desirability

Note. The ordering of “forensic and non-forensic” or “non-forensic and forensic” signifies which target population the measure was designed for first, and who it later became used for secondarily.

that an item such as “I enjoy watching my coworkers cry” can easily be recognized as measuring undesirable behaviour. Thus, it is unlikely that anyone would respond to this item at the high end of a Likert-scale when there is something to gain, like there is in a selection context (Mathieu & Babiak, 2016b). This problem inherent to face-valid self-report measures becomes magnified when it is psychopaths we are attempting to detect, as they are partially defined by their ability to lie, deceive, and manipulate (Kelsey, 2016; Kelsey, Rogers, & Robinson, 2015). Across different measures, psychopathy scores have been positively correlated to faking behaviour and intentions to fake in the future (Fisher, Robie, Christiansen, & Komar, 2018; Grieve, 2012; Roulin & Krings, 2016). Even the PPI-R is still vulnerable to faking regardless of its built-in validity scales. Marcus et al. (2018) recently analyzed the functionality of the three validity scales of the PPI-R and found that only the Deviant Responding scale (attempting to look bad) is actually associated with lower validity coefficients consistently, while the Virtuous Responding (VR) scale (attempting to look good) had relatively little validity (Marcus et al., 2018). It is important to note that other research has found that the VR scale of the PPI-R does in fact detect respondents who are responding in the most positive light (Anderson, Sellbom, Wygant, & Edens, 2013). However, the VR scale measures virtuous responding, which is similar to socially desirable responding and impression management, but explicit *faking* is altogether different from these concepts (for an in-depth explanation, see Burns & Christiansen, 2011).

Imagine a criminal offender is already in prison for the next 30 years. It is unlikely that he will care about how he is perceived by the researchers, so if he believes that he really is “*very likely*” to bully others to get his way, he will have no problem responding that way on the self-report measure. Alternatively, think about a criminal who has the chance to get moved from solitary confinement to minimum security based on his psychopathy score. This individual also believes that

he is “*very likely*” to bully others to get their way, but do you think he will be just as candid in his response?

Researchers have already provided evidence for the claim that most of the self-report measures of psychopathy can be faked at will. For example, Edens, Buffington, Tomicic, and Riley (2001) established that respondents could significantly lower PPI scores when asked to “fake good,” and Edens (2004) found similar results with *total* PPI scores although it was noted that factor 1 was only minimally altered by faking. Around the same time, Rogers et al. (2002) gave youth offenders an incentive for faking on the SRP and found that they could easily lower their psychopathy scores in comparison to their honest scores which they gave beforehand. The LSRP has also been recognised as fakeable, and total psychopathy scores measured by the LSRP were positively correlated with participant ability to successfully avoid detection on another screening measure (Book, Holden, Starzyk, Wasylkiw, & Edwards, 2006). In two more recent, comprehensive reviews, criminal offenders with moderate to high levels of psychopathy (as measured by a clinician with the PCL-R) were given the SRP, LSRP, PPI-R, and an imaginary scenario in which they could shorten their jail time if they scored low on these measures (Kelsey, 2016; Kelsey et al., 2015). On the PPI-R, these offenders lowered their scores from the 82nd percentile to the 27th percentile on average. With the LSRP, they could lower their scores equal to average student levels. Finally, on the SRP, these moderate-to-high psychopaths could lower their scores *below* the average student level (Kelsey, 2016; Kelsey et al., 2015). There is considerably less research on the TriPM as it relates to faking. However, one recent study by Kelley, Edens, Donnellan, Mowle, and Sörman (2018) found that positive impression management was significantly correlated with lower meanness and lower disinhibition when participants were rating themselves using the TriPM. To my knowledge, the B-Scan self has not yet garnered any published research surrounding its susceptibility to faking, though since it is essentially a workplace adaptation of the SRP it may prove to be just as fakeable.

Despite their vulnerability to faking, self-report measures are still the most commonly used method to assess psychopathy overall, and in the workplace specifically (Kelsey, 2016; Mathieu & Babiak, 2016b; Verschuere et al., 2014). As a result, some researchers are advocating for change in the measurement of psychopathy. Robinson and Rogers (2015) state that all self-report psychopathy measures without faking detection scales imbedded within them are of questionable utility. Others have said that the validity for all of these self-report measures is weak because respondents are ready, willing, and able to fake low psychopathy whenever there is any incentive present (despite the tendency to not fake and the inverse relationship with impression management under non-incentivized research conditions; Verschuere et al., 2014). Finally, some extreme opinions exist that self-report measures should stay away from the construct of psychopathy (Kelsey, 2016; Kelsey et al., 2015). The current research is proposing a measure that has an answer for all the major concerns as it is a self-report measure, with low cost and no expert raters, developed using non-criminal samples, and prevents the ability to fake even when presented with an incentive. This can be accomplished through an implicit Conditional Reasoning Test.

Implicit Measurement through Conditional Reasoning Tests

A conditional reasoning problem gives a set of logical premises in the context of a scenario, and a series of response options for test-takers to select the most logical one. A set of these problems make up a Conditional Reasoning Test (CRT), which measures exactly that, conditional reasoning and logical decision making. Below is an example conditional reasoning problem from the U.S. Federal Air Marshal Written Test. The correct answer is E), because the other response options can all be inferred from sentences 2 to 4.

In order to ensure a fair trial for any defendant there are certain rules that defense attorneys must obey. A defense attorney must advocate for his client to the best of his ability even if he knows that his client is guilty. A defense attorney must never share confidential, attorney-client information with the prosecuting attorney. A defense attorney must not give his client bad advice that would make the defense attorney's job easier. Failure to follow these rules can result in the defense attorney being sued for misrepresentation. Suppose Ken G. is a defense attorney who has been sued for misrepresentation.

Which of the following cannot be reasonably inferred about Ken G.?

- A) Ken G. may have shared confidential, attorney-client information with the prosecuting attorney*
- B) Ken G. may not have advocated for a client of his to the best of his ability*
- C) Ken G. may not have followed all the rules that defense attorneys must obey*
- D) Ken G. may have given a client bad advice in order to make Ken G.'s job of representing him easier*
- E) Ken G. followed all the rules that defense attorneys must obey but his client sued him since the end result of the case was not to the client's liking*

Conditional reasoning problems require many cognitive processes to be used at the same time and are usually quite challenging due to their complexity. This, in part, is why we use them to measure conditional reasoning and logical decision-making abilities which are important to many occupations (James, McIntyre, Glisson, Bowler, & Mitchell, 2004). CRTs are used as part of the selection process for many positions, and do not have to mimic scenarios that one would find in the particular position being applied for. This is the main difference between CRTs and Situational Judgement Tests (SJTs), although they are somewhat similar. It appears that newer conceptualizations of SJTs are becoming increasingly similar to classic CRTs (Catano et al., 2016).

James (1998) hypothesized that these CRTs could be used in an innovative way. He argued that differences in one's personality are reflected in attitudinal tendencies and biases to favour, adopt, and choose certain options in everyday life. For example, consider two individuals who are having issues understanding some of the material that is on an upcoming physics test. The first person, high in trait Achievement Motivation, is likely to have an implicit tendency to believe that hard work and

continued effort pays off, so they continue to study. The second person, low in Achievement Motivation, may implicitly start to distance themselves from the importance of the test and doubt their ability to improve, resulting in them choosing to give up studying. Thus, if we design conditional reasoning problems that present scenarios such as this, with response options that are associated with different attitudinal tendencies and biases, we should be able to implicitly measure the personality trait of interest without the respondent knowing that this is what is being measured.

The first step to creating a CRT which implicitly measures a personality construct is to develop a set of “justification mechanisms” (JMs; James, 1998; James et al., 2004). JMs are descriptions of these reasoning processes, tendencies, and biases that are inherent to the personality construct of interest. JMs are usually established by consulting all relevant literature on the construct to consider all theories, conceptualizations, and previous measurements. James’ initial CRT for achievement motivation decided on six JMs, one of which is given below:

Personal responsibility inclination: tendency to favor personal factors such as initiative, intensity, and persistence as the most important causes of performance on demanding tasks.

James then created the Conditional Reasoning Test for Aggression (CRT-A; James et al., 2004) which has become more popular in the literature and has been used in a personnel selection context. Thus, the following discussion will use aggression for examples. Once JMs have been finalized, conditional reasoning problems that measure the JMs need to be constructed. Each problem on the CRT measures one or more JM, similar to how a personality inventory item usually measures one factor of the personality construct (James et al., 2004). It seems relatively straightforward, except there is a trick with the response options that are offered to test-takers. For every conditional reasoning problem there are four possible response options, and the goal is to measure levels of aggression with an individual’s response selections. Essentially, this is done by

implicitly directing respondents to make a dichotomous choice for each problem, between two meaningful options. Meaning that, two of the four response options are designed to be easily recognized as “incorrect” solutions to the problem by respondents. The remaining two options are designed to be equally logical and “correct” solutions to the problem, however, one is written in a manner that aligns with high aggressive tendencies and beliefs, while the other is written in a manner that aligns with low aggressive tendencies and beliefs. To make this clearer, an example item from the Conditional Reasoning Test for Aggression (CRT-A; James et al., 2004) is given below:

The old saying, “an eye for an eye,” means that if someone hurts you, then you should hurt that person back. If you are hit, then you should hit back. If someone burns your house, then you should burn that person’s house.

Which of the following is the biggest problem with the “eye for an eye” plan?

- a. It tells people to “turn the other cheek.”*
- b. It offers no way to settle a conflict in a friendly manner.*
- c. It can only be used at certain times of the year.*
- d. People have to wait until they are attacked before they can strike.*

This is one of the easier problems to tell which response option is which, making it a clear example. It is fairly evident that options *a* and *c* do not serve as a logical response at all, guiding respondents to the other two choices. It is important to remember that respondents are told that this is a conditional reasoning test, so they are truly trying to use their best analytical skills to select *the* correct response option, though really we have just guided them into choosing an *equally logical* aggressive option (*d*) or non-aggressive option (*b*). If the aggressive option is selected the respondent receives a +1, if the non-aggressive option is selected they receive a -1, and if one of the two nonsensical options are selected they receive a zero. This is the same for every item on the CRT and in the end, respondents have a total score that measures their implicit aggressive tendencies and biases (James et al, 2004; James & LeBreton, 2010).

Although the CRT method of measuring implicit personality may seem dubious and abstract, the CRT-A has received almost unanimously positive support. When James et al. (2004) originally put forth the CRT-A, they did so using 10 validity studies which established a consistent factor structure, consistent scale reliabilities, and consistent correlations with performance measures as expected. Additionally, the researchers failed to find any connection between CRT-A scores and gender or race. The construct validity of the CRT-A and its ability to predict counter-productive work behaviours have both been more recently supported (Berry, Sackett, & Tobares, 2010; DeSimone & James, 2015; Galić, 2016; Galić, Scherer, & LeBreton, 2014). The CRT method of implicit measurement has only minimally been used with other constructs, however it has been used very recently to create an implicit CRT for creative personality (Schoen, Bowler, & Shilpzand, 2018) and power motive (Galić, Ružojčić, Trojak, & Zeljko, 2018). Most of the other CRT based measures that have been developed are only master's theses or doctoral dissertations. For example, Patton (2000) developed a CRT for reliability in the workplace, Wright (2012) created the CRT-L for leadership, Fine and Gottlieb-Litvin (2013) attempted to measure integrity through a CRT approach, Rasmussen (2016) developed a CRT to measure extraversion and agreeableness, and Clark (2017) recently put forth a CRT intended to measure hostile sexism. There has been no additional research on any of these CRTs, to my knowledge.

Unlike all of the self-report measures of psychopathy listed earlier, CRT-based measures prevent faking through their implicit nature. Researchers have consistently been able to back-up the inability to fake on implicit CRT-based measures (Bowler & Bowler, 2014; Bowler, Bowler, & Cope, 2013; Galić et al., 2014b; LeBreton, Barksdale, Robin, & James, 2007; Lee, 2014; Rasmussen, 2016; Wiita, Meyer, Kelly, & Collins, 2017). Even when given an incentive, such as being told that results would influence chances of being hired, CRTs have still proved resistant to faking as respondents cannot distort their scores if they are oblivious to what is truly being measured

(LeBreton et al., 2007; Wiita et al., 2017). However, this is all conditional upon the fact that the implicit nature remains intact. Multiple studies have shown that participants *can* fake their way to lower scores on CRTs if the researchers inform them how the test actually works beforehand (Bowler et al., 2013; LeBreton et al., 2007; Lee, 2014; Rasmussen, 2016; Wiita et al., 2017).

Identifying Psychopaths in the Selection Process with the CRT-WP

It is based on all of the information presented up to this point that the Conditional Reasoning Test for Workplace Psychopathy (CRT-WP) is proposed. The CRT-WP uses conditional reasoning methodology to implicitly measure psychopathic attitudes, tendencies, and biases, just as the CRT-A does for aggression. The CRT-WP is developed using the recommendations of James who is the creator of this implicit measurement technique. The first step in developing implicit CRTs is to develop the justification mechanisms integral to psychopathic thinking. This step uses the wealth of literature surrounding the topic as mentioned earlier, which in this case is workplace psychopathy. Some of the recent conceptualizations of psychopathy like the Triarchic model serve as a starting point, based on the evidence that its three factors have persisted across other conceptualizations as well (Drislane et al., 2014; Hall et al., 2014). James et al. (2004; 2005) recommend that four to six overarching justification mechanisms are identified in the development of implicit CRTs. Then, conditional reasoning problems with response options that require the application of those justification mechanisms are created. The content of all the conditional reasoning problems in the current study are contextualized in work-related behaviours and decisions, so that the connection between performance on the CRT-WP and actual on-the-job performance is theoretically stronger.

Development of the CRT-WP

As mentioned in the previous section, development of the CRT-WP began with consulting a wealth of literature on psychopathy measurement. Over 60 peer-reviewed research articles and scale

manuals were studied with notes being taken on factor structure, item structure, scoring system, and content coverage. Included in this literature were articles focused on each of the different measures referenced earlier including articles in support of them, challenging them, and comparing them to one another. This way, all of the articles consulted to develop the CRT-WP should not be biased toward any one particular measure or conceptualization of psychopathy more than any other. As noted above, most of the newer measures (i.e., CAPP, PAPA, etc.) have had little peer-reviewed research focused on them at this point, so these measures had minimal influence on the development on the CRT-WP. Additionally, although the EPA has existed for a longer period of time compared to some of the newer measures, there is not much validation or support for its use. Therefore, due to their continued popularity in both research and practice, and a greater understanding of their structure and application, the main measures used as a knowledge base for the creation of the CRT-WP were the PCL-R and SRP, PPI, LSRP, and the TriPM. Thus, by reviewing the literature of these measures and others, all theories, conceptualizations, and previous measurements were consulted, as recommended by James (1998) and James et al. (2004).

Justification Mechanisms

Remember that justification mechanisms (JMs) are descriptions of tendencies in thinking, reasoning processes, and biases that are intrinsic to the construct, which in this case is psychopathy. However, the JM development process was focused more specifically on the workplace manifestation of psychopathy, as the justification mechanisms and item content for criminal psychopathy may be different based on how some of the core traits (i.e., fearlessness) are viewed in different contexts (Lilienfeld et al., 2015). After an iterative process, a list of six JMs was generated which intend to comprehensively cover all of the factors from the main existing measures of psychopathy. Six JMs is within the four to six recommendation of James et al. (2004) and James and

LeBreton (2010). The list of six JMs used for the CRT-WP along with their descriptions is presented in Table 2. Before explaining each JM in detail, it is important to understand how they relate to the other measures. Table 3 displays the content coverage of the six JMs of the CRT-WP in relation to the other main measures of psychopathy. There is a wealth of literature which exists for each of the following JMs, however, only a brief explanation and the most relevant literature is presented in this section.

Externalization. As Table 2 describes, the externalization JM involves the tendency to blame others or other factors for anything which goes wrong for the individual, and not taking responsibility for their own actions. In criminal psychopaths, some research suggests that blame externalization is the facet of psychopathy which is most predictive of consistent antisocial behaviour (DeLisi et al., 2014). Failure to accept responsibility or blame has also been linked specifically with corporate psychopathy (Pardue, Robinson, & Arrigo, 2013). Blame externalization is one of the eight factors of the PPI, and the definition of this JM also covers factors from other existing measures.

Carefree Impulsivity. Largely a combination of the carefree non-planfulness and impulsive non-conformity of the PPI, the carefree impulsivity JM refers to the predisposition for actions to be guided by spontaneous impulsivity rather than careful planning or deliberation. Carefree non-planfulness has been associated with increased acceptance of toxic waste dumping (Ray & Jones, 2011). Impulsiveness has also been suggested to predict the perpetration of counter-productive work behaviours (Cohen, 2015). Impulsive and carefree actions have been included as a factor in almost every measure of psychopathy.

Social Superiority. In some professions, extreme confidence is beneficial and scoring high on this justification mechanism alone may not be a bad thing. Social superiority refers to the

Table 2

List of Justification Mechanisms Produced for the Development of the CRT-WP

<i>Name</i>	<i>Description</i>
1. Externalization	A propensity to blame other people or external factors for negative occurrences. This bias appears as a “global irresponsibility” for actions and outcomes that clearly resulted from choices under the control of the individual. Individuals with this bias will deflect blame and absolve themselves of any wrongdoings.
2. Carefree Impulsivity	A predisposition for actions and decisions to be guided by impulsivity instead of reasoning, deliberation, or long-term planning. Actions will often seem to have a disregard for socially accepted norms and behaviours. The excitement of spontaneity combined with a lack of consideration for potentially harmful outcomes results in this predisposition.
3. Social Superiority	A persisting belief that one’s social status and social skills are superior to generally everyone around them. The individual will believe that they can charm and persuade others in any situation. The individual also believes that he/she is a dominant, alpha social personality that should be considered above others.
4. Fearlessness	An inclination toward risk-taking behaviours along with a high tolerance/resilience for the uncertainty in the outcomes. This is accompanied by an abnormal disregard for, and lack of, the fear or anxiety that most people experience in high stress situations. The individual is not satisfied with just being content.
5. Ruthless Self-Interest	The tendency to actively seek out opportunities for self-promotion with complete disregard for anyone or anything other than the self. The individual strives to achieve their own goals and advancement at any cost, and will find a way to justify exploitation and other behaviours that negatively effect others as a result. There is a survival-of-the-fittest mentality.
6. Insensitivity	A disinclination to feel concern, guilt, remorse, or give any consideration to the feelings of others. This is a complete lack of empathy in any situation. This differs from ruthless personal gain in that this insensitivity is present even in situations where there is nothing to gain for the individual.

Note. It is acceptable for the JMs to have some degree of overlap according to the other implicit CRTs which have already been developed (James et al., 2004; Schoen et al., 2018). Definitions for Carefree Impulsivity and Fearlessness were re-worded following Study 1 (see below).

Table 3

Content Coverage of the Justification Mechanisms for the CRT-WP in Relation to Main Measures

<i>Justification Mechanism</i>	<i>Tri-PM</i>	<i>PCL-R / SRP</i>	<i>PPI</i>	<i>LSRP</i>
Externalization	- Disinhibition	- Irresponsibility - Failure to accept responsibility	- Blame externalization	- Primary
Carefree Impulsivity	- Disinhibition	- Poor behaviour control - Lack of long-term goals - Impulsivity	- Carefree non-planfulness - Impulsive non-conformity	- Secondary
Social Superiority	- Boldness	- Glib/Superficial charm - Grandiose estimation of self - Pathological lying **	- Social potency	- Primary & Secondary*
Fearlessness	- Boldness	- Need for stimulation ** - Sexual promiscuity	- Fearlessness - Stress immunity	- Secondary
Ruthless Self-Interest	- Meanness	- Cunning and manipulateness - Callousness - Parasitic lifestyle	- Machiavellian egocentricity	- Primary
Insensitivity	- Meanness	- Lack of remorse/guilt ** - Shallow affect - Lack of empathy	- Cold-heartedness	- Primary
<i>(Not Covered)</i>	-	- Early behaviour problems - Short-term marital relationships - Juvenile delinquency - Revocation of conditional release - Criminal versatility	-	-

Note. The bottom row indicates any factors on the other measures which are not covered by the six JMs for the CRT-WP. Only the factors of the PCL-R and SRP related to criminal behaviour are not covered, which is by choice. * = Social Superiority covers both primary and secondary factors of the LSRP, ** = These factors are somewhat covered by other JMs as well.

persistent belief that one is superior to everyone around them. The individual will believe that he or she has superior social skills which they can use to can charm and persuade others in any situation. In corporate professionals, a grandiose estimation of self and overall psychopathy scores have been significantly correlated with charisma and charm, and all of these variables were negatively associated with performance ratings by others (Babiak, Neumann, & Hare, 2010).

Fearlessness. Similar to social superiority, fearlessness may not always be a maladaptive trait in a corporate setting. In fact, these two JMs encompass the boldness factor of the triarchic model which is generally believed to be an adaptive factor in many cases (Lilienfeld et al., 2012). The fearlessness JM describes the inclination toward risk-taking behaviours in psychopathic thinking, which is accompanied by an abnormal disregard for, and lack of, the fear or anxiety that most people experience in high stress situations. As mentioned earlier, fearless dominance was associated with overall psychopathy scores and with higher approval ratings in U.S. presidents (Lilienfeld et al., 2012). Fearlessness was another factor correlated with more lenient attitudes toward toxic dumping (Ray & Jones, 2011).

Ruthless Self-Interest. Perhaps the most salient characteristics of psychopathy are meanness, manipulateness, and callousness. These, along with the PPI factor of Machiavellian egocentricity, are covered by the Ruthless Self-Interest JM. This refers to the tendency to seek opportunities for self-promotion with complete disregard for anyone or anything other than the self, and striving to achieve one's own goals and advancement at any cost. Boddy et al. (2010) states that this callousness and ruthlessness is what makes corporate psychopaths terrible corporate citizens and harmful to subordinate employees. This concept of meanness at the expense of others has been included in every measure of psychopathy (Smith & Lilienfeld, 2013).

Insensitivity. This JM describes a complete lack of empathy. It differs from ruthless personal gain in that this insensitivity is present even in situations where there is nothing to gain for the individual. It is also described as cold-heartedness, or a lack of remorse in other measures. This insensitivity could be viewed as a benefit in some professions, however, it is generally a negative and is critical to the conceptualization of the corporate psychopath (Smith & Lilienfeld, 2013). Cold-heartedness, for example, has been associated with higher annual incomes and higher corporate ranks in financial employees (Howe, Falkenbach, & Massey (2014).

As Table 3 shows, these six justification mechanisms cover the necessary biases in thinking for all of the factors of the PPI (and PPI-R), TriPM and LSRP. Note that for the TriPM, there are actually two justification mechanisms for each of the three factors, which was a methodological choice because of the research cited earlier which seems to indicate that the triarchic model actually underlies other models such as the PPI and LSRP (Drislane et al., 2014). Another methodological choice was for the JMs to **not** cover all of the factors on the PCL-R and SRP. The factors such as “revocation of conditional release,” “juvenile delinquency,” and even “short-term marital relationships” simply are not inherent to the corporate brand of psychopath, and also cover content that is illegal, or at least unethical, to ask in most employment interviews (Catano et al., 2016). Of course, it is true that corporate psychopaths may be more likely to engage in criminal behaviours more than non-psychopathic workers, but these factors of the PCL-R/SRP are not *required* to be a corporate psychopath according to the depth of literature on the typology (Smith et al., 2014). As a result, the CRT-WP is aligned with conceptualizations of the PPI-R and TriPM in that criminal or anti-social behaviours are considered more of an outcome of psychopathy, rather than part of the construct itself. This choice was made due to the recent literature seemingly supporting these conceptualizations more so than the PCL-R framework which includes criminal and anti-social behaviours as a defining part of psychopathy measurement.

Item Generation

The item generation process followed a similar process to that of the Conditional Reasoning Test for Aggression (James et al., 2004), and the very recent Conditional Reasoning Test for Creative Personality (Schoen et al., 2018). In both of these existing measures, it is suggested that the final number of items fall within the 18 – 25 range. Knowing this, and knowing that the items created would be going through a multiple phases of revisions, it was decided that double the desired number of items on the final version should be generated. This is common procedure in the scale development and item generation process (Hinkin, 1998). Since there were six JMs to produce items for, it was decided that ten items would be written for each JM, resulting in a final total of 60 items going into the scale development process. This number leaves even more room for the deletion of unwanted items. Multiple item formats were used so that they were similar to the scenarios and problems found in the items of the conditional reasoning test for aggression, while also maintaining the appearance of a true conditional reasoning problem. The 60 items written by myself went through three phases of revisions with the faculty supervisor for this project who specializes in assessment and faking. During the three phases of initial item revisions, some items were completely re-worked while other items remained unchanged. The final 60-item list for the CRT-WP heading into the initial phases of scale development can be found in Appendix B. Note that the items are categorized by JM and the psychopathic and non-psychopathic options are labelled for easier understanding.

In the other implicit CRTs that have been developed, individual items could be based on more than one JM and they still produced an acceptable factor structure. For the CRT-WP, it was decided that items would be written so that they only intended to measure one JM, which would hopefully improve the ability to find a clear factor structure that is more well-defined than previous CRTs. Thus, when reading the items and response options, the “bad” and “good” responses may take

the form of “fearlessness” and “non-fearlessness” responses (i.e., they measure the specific JM), rather than the “psychopathic” and “non-psychopathic” response options (i.e., each item measuring overall psychopathy). For this reason, sometimes the response option which indicates psychopathy may not necessarily be “bad” or negative considering that some of the factors (ex. fearlessness) can be advantageous in certain situations. However, when all of these individual measurements of separate JMs are taken together, they should measure overall workplace psychopathy the same way that a regular scale with multiple factors measures one construct.

Finally, all of the items are written in work-specific scenarios for two reasons. First, having all of the items written in work-specific contexts enhances the face validity of the measure to the respondents. This will make it more believable that the CRT-WP is measuring conditional reasoning and problem-solving ability in the workplace instead of psychopathy. Again, this believability should also increase respondents’ desires to select *the right answer* so that they get a higher score of reasoning ability, when in reality this will distract them from the true nature of the measure. The second reason that the items are worded in work-specific contexts is so that it measures work-related psychopathic tendencies as intended, instead of broader psychopathy for which these JMs *may* be unsuited. Moreover, having the CRT-WP not be workplace specific in its’ items would give it some of the same criticisms as most of the other self-report measures mentioned earlier.

The proposed method for this thesis research involves four separate studies each with their own unique structure. As a result, the following sections are separated by study instead of providing the methodology for all studies and then all of the results, for example. A general discussion section is provided after all four studies are detailed. First, Study 1 involved a comprehensive panel of subject-matter experts (SMEs) who sorted and reviewed the preliminary 60 items. That was followed by an initial data collection where the remaining items were given to an online sample, and the factor

structure and reliability of the items were assessed. After this, many items were removed from the original list so that only the items which functioned the best remained. This new and refined version of the CRT-WP was then given to another online sample (Study 3) and a student sample (Study 4). These two studies aimed to validate different properties of the CRT-WP, which involved the two samples completing other measures and two different waves of data collection for Study 3.

Study 1 – Item and Scale Revision with Subject Matter Experts

Methods

Study 1 was a construct validation and item refinement study. The CRT-A underwent many item and scale refinement analyses prior to the first time the CRT-A was given to a sample of respondents (James & McIntyre, 2000). However, the more recent CRT for Creative Personality (CRT-CP; Schoen et al., 2018) went through a comparatively modest item and development process that only involved the authors. The current study for the CRT-WP proved to be somewhere in between the two as myself and the project supervisor revised the items multiple times up to this point, but in this study the initial list of 60 CRT-WP items went through a significant SME review.

Fifteen SMEs participated in this initial review study. The SMEs consisted of 14 graduate students in the I/O (9) and forensic (5) psychology programs, and one psychology professional with clinical psychology background. No demographic details were collected from the SMEs due to potential identifiability. Having a diverse group of experience and cross-field expertise is important to this current review of the CRT-WP items. Student SMEs with a background in psychology were theoretically more familiar with the corporate type of psychopath and the items which are worded in work contexts, while the clinician SME was likely more knowledgeable and more experienced with the construct overall.

The group of fifteen SMEs were given a paper and pencil document that contained the list of 60 initially developed CRT-WP items in a randomized order, and instructions for how to complete the task (see Appendix C). First, the SMEs were asked to identify which JM that each item is measuring by circling one of the six JMs which were all listed beside each item. Second, SMEs were also asked to identify which of the four response options represents a high degree of that JM and which response option represents a low degree of that JM. For example, respondents might circle “Ruthless Self-Interest” as the JM which they believe is being measured by a particular item, and then they might circle response option b) as the item which represents a high degree of Ruthless Self-Interest and then d) as the low Ruthless Self-Interest option.

In the instructions given to the SMEs they were given the names and descriptions of each JM, identical to Table 2, and they were also given a concise description of the implicit nature of the CRT-WP so that they were well informed about how the measure and items truly function. SMEs were also given an example item which demonstrated to them how they were to complete the task. The SMEs were also told to make comments about confusing wording or difficulties in understanding items and response options, which many of them did. SMEs were given a week to complete the task and were told to not discuss the task with other SMEs.

The responses of the SMEs were analyzed using percent agreement, and ability to identify the correct response options for each item. Originally, it was decided that any item that did not reach 70% agreement on both which JM is being measured, and the identification of the high and low response options, will be removed from the item list (as recommended by Hinkin, 1998). However, if there was better agreement than anticipated, the cut-off score would be raised. This caveat was suggested so that any item which functions worse than the others on the list was eliminated and only the best items were kept for additional studies, which is typical for any scale development process.

Additionally, whether all of the items below the cut-off score were deleted immediately or kept for re-wording or re-conceptualizing was dependent on how many items met the cut-off score. For example, if only 20 items met the 70% agreement then another 5 or 10 would be kept for re-wording, whereas if 50 items met 70% agreement then no more items would be kept for re-wording. This guideline was set beforehand so that there would be an adequate number of items following Study 1 which could result in the ideal number of items for the final version.

There were no restrictions as to how many items could be removed for each JM. Therefore, it was possible that after Study 1 there were only 5/10 items for Carefree Impulsivity remaining while there were still 9/10 items for Insensitivity, for example. As mentioned earlier, both the CRT-A and CRT-CP include items which measure more than one JM and the design to focus on only one JM per item in the current study was a purposeful choice. Although the previous implicit CRTs have demonstrated acceptable factor structures, the one-JM-per-item strategy of the CRT-WP may lead to a more defined factor structure.

Results

All of the 15 SMEs completed the task within the given time frame. However, the responses of one SME were removed from analyses due to what was deemed as inattentive responding. This one SME was only accurate in less than 50% of their selections, which is much lower than any of the other SMEs (as evidenced below). As well, this SME seemed to have reversed the high and low JM response options for many (almost half) of the items, which may indicate a lack of understanding the instructions. To ensure that this one SME being different from the rest was not attributed to the difference between the one professional clinician SME and the 14 student SMEs, it was confirmed that the removed SME was not the professional clinician. However, the individual student who was removed as an SME was never identified. SMEs were not given any compensation for participating.

The remaining 14 SMEs used for analyses were all assumed to be attentive and engaged in the task. The percentage agreement between the 14 SMEs was higher than anticipated, so three cut-off percentages (70%, 75%, and 80%) were explored to see how many items would be removed using each. For each of the cut-off percentages, the items were divided into three groups based on their percent agreement. The first group consisted of items which scored above the cut-off percentage on all three aspects the SMEs had to assess (which JM the item measured, identifying the low JM response, and identifying the high JM response). These items were deemed to be functioning as anticipated. The second group consisted of items which scored above the cut-off percentage on the identification of the low and high response options, but below the cut-off score on the identification of the JM. Recall that the identification of the correct JM is less of an issue, as items of previous implicit CRTs are designed to measure more than one JM (James et al., 2004; Schoen et al., 2018). The items in this second group were reassessed for confusing wording or whether they were best suited to measure a JM other than the one intended. Finally, the third group of items were those that did not reach the percentage agreement cut-off for at least one of the low/high response identifications. These items would be removed following this study.

At the original 70% agreement cut-off, 35 items were in the “pass” category, 19 items were in the reassess category, and only six items were in the removal category. At the 75% agreement cut-off, 26 items were in the “pass” category, 20 items were in the reassess category, and 14 items were in the removal category. At the 80% cut-off, only 18 items were in the “pass” category, 25 items were in the reassess category, and 17 items were in the removal category.

After considering the different cut-off scores, it was deemed that moving forward with the 75% agreement would be best considering that the identification of the JM is less important than the low/high response options, and the number of items in the reassess category. The breakdown of

percent agreement for all 60 initial items at the 75% cut-off level is given in Table 4. With 14 SMEs, this increase in cut-off from 70% to 75% means that 11/14 SMEs were required to have agreed instead of 10/14 SMEs. Thus, the 26 items from the “pass” category were only examined if there were comments or re-phrasing suggestions given by the SME, which resulted in no meaningful changes beyond very minor re-phrasing. The 14 items in the removal category were eliminated and never used again following this study. Out of the 20 items in the reassess category, myself and the project supervisor came up with meaningful changes or re-evaluated the JM being measured for 17 of them, which were kept after those revisions (the other three were deemed unsalvageable). In total, this process resulted in 43 items (6 Externalization, 6 Carefree Impulsivity, 6 Social Superiority, 8 Fearlessness, 9 Ruthless Self-Interest, 8 Insensitivity) passing this study of development. This revised list of items (including the changes to the 17 modified items) can be found in Appendix D.

Discussion

The purpose of this initial study was item removal and revision. On both accounts, the results were satisfactory and produced a good set of 43 items to carry into Study 2. Given that the intent through Studies 1 and 2 was to eliminate 50-60% of the initial item list and leave only the best items remaining, removing 17 items (28.33%) in this study was right on course. As Table 4 indicates, seven of the 10 items intended to measure fearlessness required re-wording or being moved to another JM (carefree impulsivity). According to SME comments and critical review of the JMs, this likely resulted from the fearlessness and carefree impulsivity JMs having similarly worded definitions. While the two concepts do have slight overlap, carefree impulsivity at its’ core is intended to describe the spontaneity, lack of long-term planning, and tendency to make decisions on a momentary impulse. On the other hand, fearlessness is intended to describe the preference (with planning and consideration involved) for high-risk behaviours and the lack of anxiety or fear that

Table 4

Agreement among SMEs for the Original List of 60 CRT-WP Items

<i>Item</i>	<i>JM</i>	<i>Low</i>	<i>High</i>	<i>Result</i>	<i>Item</i>	<i>JM</i>	<i>Low</i>	<i>High</i>	<i>Result</i>
EXT 1	14/14	14/14	14/14	Pass	FLN 1	11/14	13/14	13/14	Pass
EXT 2	10/14	11/14	11/14	Reword	FLN 2	13/14	13/14	13/14	Pass
EXT 3	14/14	13/14	13/14	Pass	FLN 3	7/14	13/14	13/14	Reword
EXT 4	3/14	10/14	10/14	Remove	FLN 4	6/14	12/14	12/14	Reword
EXT 5	10/14	10/14	10/14	Remove	FLN 5	6/14	14/14	14/14	Reword
EXT 6	11/13	10/13	8/13	Remove	FLN 6	4/14	14/14	14/14	Now CI
EXT 7	13/14	11/14	11/14	Pass	FLN 7	6/14	9/14	9/14	Remove
EXT 8	11/14	10/14	10/14	Remove	FLN 8	10/14	11/14	11/14	Reword
EXT 9	14/14	14/14	14/14	Pass	FLN 9	8/14	12/14	12/14	Reword
EXT 10	14/14	14/14	14/14	Pass	FLN 10	5/14	13/14	13/14	Reword
CI 1	12/14	14/14	14/14	Pass	RSI 1	13/14	14/14	14/14	Pass
CI 2	12/14	11/14	11/14	Pass	RSI 2	13/14	11/14	10/14	Remove
CI 3	1/14	12/14	12/14	Remove	RSI 3	7/14	6/14	6/14	Remove
CI 4	6/14	9/14	8/14	Remove	RSI 4	4/14	13/14	13/14	Reword
CI 5	14/14	14/14	14/14	Pass	RSI 5	12/14	10/14	10/14	Remove
CI 6	11/14	14/14	14/14	Pass	RSI 6	9/14	14/14	13/14	Reword
CI 7	13/14	10/14	10/14	Remove	RSI 7	10/14	14/14	14/14	Reword
CI 8	7/14	9/14	9/14	Remove	RSI 8	12/14	14/14	13/14	Pass
CI 9	7/14	6/12	6/12	Remove	RSI 9	13/14	11/14	11/14	Pass
CI 10	13/14	13/14	13/14	Pass	RSI 10	9/14	14/14	14/14	Reword
SS 1	13/14	14/14	14/14	Pass	INS 1	13/14	14/14	14/14	Pass
SS 2	12/14	14/14	14/14	Pass	INS 2	13/14	13/14	12/14	Pass
SS 3	6/14	13/14	13/14	Reword	INS 3	10/14	13/14	14/14	Reword
SS 4	6/14	14/14	14/14	Now RSI	INS 4	11/14	12/14	13/14	Pass
SS 5	12/14	12/14	12/14	Pass	INS 5	5/14	11/14	11/14	Remove
SS 6	7/13	10/13	10/13	Remove	INS 6	5/14	14/14	14/14	Now RSI
SS 7	1/14	13/14	13/14	Remove	INS 7	11/14	11/14	11/14	Pass
SS 8	5/13	10/13	10/13	Remove	INS 8	12/14	12/14	12/14	Pass
SS 9	13/14	14/14	14/14	Pass	INS 9	9/14	13/14	13/14	Reword
SS 10	13/14	13/14	13/14	Pass	INS 10	11/14	12/14	12/14	Pass

Note. Denominators of 13 or 12 results from one or two SMEs not making a selection. *JM* = Number of SMEs who identified the intended JM. *Low* = Number of SMEs who identified the low response option. *High* = Number of SMEs who identified the high response option. EXT = Externalization. CI = Carefree Impulsivity. SS = Social Superiority. FLN = Fearlessness. RSI = Ruthless Self-Interest. INS = Insensitivity.

most people would experience in those situations. However, the original definitions did not tease apart the two concepts enough according to the SMEs. Thus, the definition for carefree impulsivity was altered to the following: *“A predisposition for actions and decisions to be guided by impulsivity instead of reasoning, deliberation, or long-term planning. The unconscious excitement that spontaneity gives to the individual results in actions based on nothing other than a momentary impulse. Actions may often seem to have a disregard for socially accepted norms and behaviours, as most others give the time and thought to consider these things before acting.”* Additionally, the definition for fearlessness was reformed into the following: *“An inclination and preference for high-risk behaviours along with a high tolerance/resilience for the uncertainty in potential outcomes. This is accompanied by an abnormal disregard for, and lack of, fear or anxiety that most people experience in high-risk or high-stress situations. This is different from carefree impulsivity in that individuals high on fearlessness are making high-risk choices even after deliberation and planning, because they genuinely find them more attractive”*.

Four original ruthless self-interest items also needed to be re-worded. However, in three of these cases, the items only required to be tweaked so that they were perceived as more manipulative, more parasitic, and to have a more psychopathic response option. This was considered less of an issue but is still noted here given the number of items that it applied to. Nevertheless, the SMEs seemed to comprehend most of the original items and were more accurate in their agreement than expected (as evidenced by the slight increase in cut-off percentage). This was a promising outcome given that this was the first exposure SMEs had to the complex nature of implicit CRT items.

Study 2 – Exploratory Factor Analysis and Initial Reliabilities

Carrying forward the refined list of items from Study 1, Study 2 attempted to find a clear and interpretable factor structure within the CRT-WP, while also eliminating more items. To do this,

participants were gathered using the Mechanical Turk (MTurk) and TurkPrime online survey platforms. MTurk is a platform created by Amazon which allows users to create profiles and receive compensation (in USD) for completing surveys that are uploaded by researchers and businesses. TurkPrime is an additional paid service which allows for greater customization and control over the surveys uploaded (Litman, Robinson, & Abberbock, 2017). Some researchers take issue with MTurk samples, as they believe MTurk users are inattentive, professional test-takers, and more likely to deceive than more traditional samples (Hauser, Paolacci, & Chandler, 2018). However, recent reviews have demonstrated that MTurk samples are more diverse and at least as good as undergraduate student samples, if not better, and that most of these aforementioned concerns are disproven by research (Hauser et al., 2018; Landers & Behrend, 2015). Additionally, the recently published research developing the CRT-CP used MTurk participants in two of their five studies (Schoen et al., 2018).

Participants were given the revised list of CRT-WP items following Study 1 and exploratory factor analyses (EFAs) were conducted. EFAs are traditionally the first step taken when developing a new measure, and always precede confirmatory factor analyses (Hinkin, 1998). After an EFA is conducted and items are removed as a result, another EFA is then conducted with the remaining items to determine if the fit indices improved and if the structure is more interpretable. As mentioned earlier, previous CRT studies have found factor structures that fit their JMs even though they had items that measured more than one JM. However, since this is the first data collected with participants in the scale development process, EFAs will be largely exploratory.

Method

Sample. To aid in gathering only quality MTurk participants, only MTurk users who had at least a 90% approval rating and who had already completed at least 100 surveys prior to the

completion of this study were permitted (as suggested by Hauser et al., 2018). Any other users who did not meet these criteria were automatically unable to complete the study. Additionally, only users over the age of 18 and who have an IP address within the United States or Canada were allowed to complete the current study (though there was an issue with the IP Address restrictions for the first 100 participants; see discussion). This study required approximately 500 participants based on guidelines for use of exploratory factor analysis (EFA) in scale development research (Worthington & Whittaker, 2006). Although only 300-400 is satisfactory for the needs of an EFA, there were two hurdles in this data collection which resulted in participants being removed, therefore starting with 500 would most likely result in a usable sample of 400 for the analysis. After the survey closed on MTurk, the raw data consisted of 497 complete responses.

Three attention-check items were included in the online survey to identify participants who were responding carelessly (Huang, Curran, Keeney, Poposki, & DeShon, 2012). The idea behind these attention-check items is that they are framed so that there is only one very obvious response option, and any participant who does not select the correct option on all three items is assumed to be inattentive. These three items were “I eat concrete daily,” “I can travel through time and space,” and “I have never used a computer before.” Ordinarily, any MTurk user who did not reply with “strongly disagree” or “disagree” to all three of the items would be deleted. However, in this study there were four MTurk users who chose to email the researchers stating that they did not know how to respond to the item about travelling through time and space, because technically every person is doing that every day. In some of these cases, the MTurk users indicated that they selected “neither agree nor disagree” to that item. With this in mind, any participant who responded with “neither agree nor disagree” to that specific item, while still responding with “disagree” or “strongly disagree” to the other two items were permitted. This exception applied to eight respondents. Based on their

responses to attention check items, 76 respondents were assumed to be inattentive and were removed from the data, leaving 421 remaining.

Another common gauge of attention used in previous CRT research is the number of illogical/nonsensical response options selected by respondents. James et al. (2004) and James et al. (2005) suggested that any respondent who selects an illogical choice for five or more items should be removed as it is likely they are not paying attention or did not understand what they were asked to do. Schoen et al. (2018) notes that there is no magic number, and that five items is approximately one quarter of James' measure. Therefore, Schoen et al. states that any participant who selects illogical response options for at least 25% of the items should be removed. The current study followed this guideline, so that any respondent who selected 11 or more illogical response choices was eliminated before any statistical analyses were conducted. This additional cut-off resulted in 25 more respondents being removed from the dataset, leaving 396 for the final sample of participants in this study.

These 396 participants were 60.4% male, 39.4% female, and 0.3% indicated that they were "other." The sample was 78.5% Caucasian, 9.1% Black/African American, 5.1% Hispanic, 4.8% East Asian, 1.0% South Asian, and a total of 1.6% responded with Native/Aboriginal, Pacific Islander, or Other. Concerning their highest level of education attained, 66.9% reported having at least some college/university education, 32.1% indicated that they were high school graduates, and 1% either reported they had less than a high school education, or other. When asked to indicate their employment status, 87.6% of participants reported working at least part-time (62.6% full-time), 7.1% were unemployed, 2.5% were students, 1.5% was retired, and 1.3% indicated "other." The mean age was 37.36 ($SD = 10.89$).

Procedure and Analyses. Participants in Study 2 were first presented with an online consent form, which detailed their rights as a participant, that participation was voluntary, and that they could drop out of the survey at any time. However, the consent form did not reveal the true nature of the CRT-WP in that it is measuring psychopathic tendencies. Instead, the participants were told that they are being given the CRT-WP which stands for Conditional Reasoning Test for Workplace Problems. They were then told that the CRT-WP measures their problem-solving and reasoning ability to determine what is the correct response to different work-related word problems, and that this was being studied because it is an important element that leads to success in most workplaces. The use of passive deception here was necessary, as the implicit nature of the CRT-WP is its foundation.

Following the consent form, participants were asked to complete all of the CRT-WP items remaining after Study 1, and a demographic questionnaire. The demographic section asked for participants' age, gender, ethnicity, level of education, and current employment status (percentages are given in the previous section). The items of the CRT-WP were presented in a randomized order to each participant. The CRT-WP was scored using the same method as previous CRT measures, where selecting a "psychopathic" or "high JM" response option results in a +1, the selection of either of the two illogical response options results in a 0, and the selection of the "non-psychopathic" or "low JM" response option results in a -1. Accordingly, the scores for all items are summed together (instead of averaging, which is more common) so that there are total scores for the overall CRT-WP, and/or each JM individually. Previous CRT research has stated that this was the most usable and psychometrically sound scoring system, in comparison to others which were tried (James et al., 2004, James et al., 2005). As the CRT-WP was the only measure being analysed in this study, another short measure was included (IPIP-20; Goldberg, 1999) simply so that the three attention check items could be embedded within it. The IPIP-20 is a short personality measure based on the big-five but it was not analysed in the current study. The IPIP-20 was presented between the CRT-WP and

demographics sections. The mean completion time was 28.78 minutes ($SD = 9.73$) and after responding to demographics, participants were presented with the debriefing form which detailed the true nature of the study and psychopathy in general.

Table 5

Percentages of Response Selections for 43 Items in Study 2 and their Result After the EFA Process

<i>Item</i>	<i>High</i>	<i>Low</i>	<i>Illog.</i>	<i>Loading</i>	<i>Item</i>	<i>High</i>	<i>Low</i>	<i>Illog.</i>	<i>Loading</i>
EXT 1	72.5%	26.3%	1%	Removed	FLN 5	13.1%	81.8%	5%	Factor 1
EXT 2	32.3%	63.9%	3.3%	Factor 2	FLN 8	56.3%	42.7%	1.1%	Removed
EXT 3	53.5%	45.5%	1%	Removed	FLN 9	40.2%	59.3%	0.5%	Removed
EXT 7	87.1%	12.4%	0.5%	Removed	FLN 10	22.5%	76.5%	0.8%	Factor 1
EXT 9	17.7%	81.6%	0.8%	Factor 1	RSI 1	37.1%	62.6%	0.3%	Factor 2
EXT 10	8.6%	89.9%	1.5%	Factor 1	RSI 4	75.8%	20.7%	3.6%	Removed
CI 1	7.8%	91.4%	0.8%	Factor 1	RSI 6	27%	71.5%	1.3%	Factor 2
CI 2	83.3%	16.4%	0.3%	Removed	RSI 7	41.7%	57.1%	1.3%	Factor 2
CI 5	43.4%	51.3%	5.3%	Removed	RSI 8	36.9%	60.9%	2.3%	Factor 2
CI 6	21.7%	76.3%	2.1%	Factor 1	RSI 9	87.1%	11.6%	1.3%	Removed
CI 10	43.7%	40.9%	15.4%	Factor 1	RSI 10	33.3%	65.9%	0.8%	Factor 2
CI 11	33.6%	62.9%	3.3%	Factor 1	RSI 11	10.6%	84.6%	4.8%	Removed
SS 1	40.7%	58.8%	0.5%	Factor 1	RSI 12	37.6%	59.8%	2.5%	Factor 2
SS 2	29%	70.5%	0.5%	Factor 2	INS 1	8.8%	87.9%	3.3%	Removed
SS 3	28.8%	69.9%	1%	Factor 2	INS 2	58.8%	40.4%	0.8%	Factor 2
SS 5	63.9%	34.8%	1.3%	Removed	INS 3	43.2%	54.5%	2.3%	Factor 2
SS 9	24.7%	73.5%	1.8%	Factor 1	INS 4	54%	42.9%	3%	Factor 2
SS 10	57.8%	41.2%	1.1%	Removed	INS 7	77.5%	21%	1.3%	Factor 2
FLN 1	16.2%	81.3%	2.3%	Factor 1	INS 8	15.7%	83.1%	1.3%	Removed
FLN 2	21.5%	77.8%	0.8%	Factor 1	INS 9	51%	47.5%	1.3%	Factor 2
FLN 3	13.4%	85.9%	0.8%	Removed	INS 10	9.1%	90.4%	0.5%	Removed
FLN 4	39.6%	52.5%	7.6%	Factor 1					

Note. $N = 396$. *High* = Percentage of participants who selected the high JM (psychopathic) response option. *Low* = Percentage of participants who selected the low JM (non-psychopathic) response option. *Illog.* = Percentage of participants who selected either of the illogical response options. EXT = Externalization. CI = Carefree Impulsivity. SS = Social Superiority. FLN = Fearlessness. RSI = Ruthless Self-Interest. INS = Insensitivity.

Results

These response frequencies for all 43 items in Study 2 is found in Table 5. In theory, since high levels of psychopathy are less common in the general population, it makes sense that the psychopathic response option should be selected less than the non-psychopathic response option. However, note that any one item does not assess whether someone scores high on psychopathy or not. Therefore, some items having the psychopathic response option chosen more frequently is likely not a major issue and did not directly lead to the removal of any items. Additionally, in previous research it is recommended that the two illogical response options combined should not exceed 5% selection for any item, if possible (James, 1998; James & McIntyre, 2000). A high percentage of participants choosing the illogical response options is indicative of poor item structure.

The responses on the CRT-WP were analyzed using Mplus version 7.4 (Muthén & Muthén, 2017). This EFA used procedures 4.2 “EFA with categorical factor indicators” from the Mplus user guide. This choice was necessary due to the +1, 0, -1 scoring of the CRT-WP. The Geomin rotation method was used with an oblique rotation due to expected intercorrelations between factors. Since multiple EFAs were being conducted to determine the best factor structure, various indicators were compared including interpretability, eigenvalues, χ^2 , CFI, RMSEA, SRMR, and reliability coefficients. Given the categorical scoring format, the Kuder-Richardson-20 coefficient (KR-20; Kuder & Richardson, 1937) is the appropriate measure of reliability (James et al., 2005, Schoen et al., 2018). It should be noted that unless all items are of identical difficulty to participants (highly unlikely in most cases), the KR-20 is an estimate of the lower bound of internal consistency. Therefore, it is acceptable for values to be slightly lower than what would be expected when using Cronbach’s alpha.

The first EFA conducted included all 43 items brought into the current study following Study 1. All EFAs in this study were asked to calculate factor structures for 1-6 factor models. The resulting factor structures of the first EFA were assessed, and nine evidently problematic items were removed as a result. Items were considered problematic if they did not load strongly on any factor or if they forced an additional factor beyond an already theoretically meaningful structure (e.g., if there was a three-factor structure where two of the factors were interpretable, and the third factor consisted of two or three seemingly random items). The second EFA using the remaining 34 items generally provided better fits for all 1-6 factor models. However, an additional three items were identified as problematic at this stage, and they were removed. Thus, the third EFA conducted which contained 31 items. Again, the statistical fits of these models were significantly better than those of the previous EFA, indicating that removing problematic items is improving the measure structurally. After examining the factor loadings across the various models, another three items were removed. The fourth EFA, containing 28 items, provided better fit indices across all factor structures than the previous EFA, but the differences started to become noticeably smaller. Upon inspection of factor loadings, only one item was removed. The fifth EFA was conducted with the remaining 27 items and again the fit indices improved slightly across all models, however, most were only small changes. Based on the levelling off of fit indices, no items having problematic factor loadings, and being within the target range of 24-30 items, this was considered the final EFA and no more items were to be removed in this study.

The resulting fit indices for 1- to 5- factor models from this final 27-item EFA are presented in Table 6. Both interpretability and fit indices reaching acceptable thresholds support the 2-factor model as the best simple statistical structure. According to χ^2 difference tests, the 3-factor model is superior to the 2-factor model, however it is lacking any theoretical meaningfulness. The rotated loadings for the final 2-factor structure with 27 items is presented in Table 7.

Table 6

Fit indices for 1-5 Factor Models in Final EFA of Study 2 (27 items)

	χ^2	χ^2/df	CFI	RMSEA	90% CI		SRMR
					Lower	Upper	
1-Factor	413.75	1.28	0.82	0.03	0.02	0.03	0.09
2-Factor	317.42*	1.07	0.96	0.01	0.00	0.02	0.08
3-Factor	275.24*	1.01	0.99	0.01	0.00	0.02	0.07
4-Factor	239.67	0.96	0.99	0.00	0.00	0.01	0.06
5-Factor	206.84	0.92	1.00	0.00	0.00	0.01	0.06

Note. $N = 396$. * = χ^2 difference tests show this model is significantly superior to the previous model.

Table 7

Factor Loadings for 27-Item, 2-Factor Model Resulting from Study 2

<i>Item</i>	<i>Factor 1</i>	<i>Factor 2</i>	<i>Item</i>	<i>Factor 1</i>	<i>Factor 2</i>
EXT 9	.36	.05	EXT 2	.18	.29
EXT 10	.35	.10	SS 2	.07	.37
CI 1	.39	.12	SS 3	-.01	.58
CI 6	.29	.09	RSI 1	.08	.29
CI 10	.30	-.04	RSI 6	-.03	.53
CI 11	.43	.02	RSI 7	.05	.46
SS 1	.28	.07	RSI 8	.26	.44
SS 9	.21	.03	RSI 10	.28	.35
FLN 1	.63	-.18	RSI 12	.03	.32
FLN 2	.49	-.03	INS 2	.01	.50
FLN 4	.32	.31	INS 3	-.03	.52
FLN 5	.28	.02	INS 4	-.10	.23
FLN 10	.40	-.02	INS 7	-.06	.37
			INS 9	-.21	.36

Note. $N = 396$. Loadings are Geomin Rotated. Loadings in boldface type are significant at the .05 level. EXT = Externalization. CI = Carefree Impulsivity. SS = Social Superiority. FLN = Fearlessness. RSI = Ruthless Self-Interest. INS = Insensitivity.

Table 5 identifies all items which were removed through the entire EFA process, and whether the remaining 27 items loaded on the first or second factor. Finally, using the above factor structure,

total scores were summed for Factor 1, Factor 2, and the overall 27-item scale. These total scores were then correlated with each other and dichotomously scored demographic variables. These correlations, along with means, standard deviations, and reliability coefficients are given in Table 8.

Table 8

Descriptive Statistics, Reliability Coefficients, and Correlations for the 27-Item CRT-WP in Study 2

	<i>Scale</i>	<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.	7.
1. Factor 1	-13 to 13	-6.34	3.97	(.49)						
2. Factor 2	-14 to 14	-2.00	5.35	.24**	(.61)					
3. CRT-WP	-27 to 27	-8.34	7.38	.71**	.85**	(.64)				
4. Gender	1-2	1.39	0.49	-.05	-.09	-.09	-			
5. Ethnicity	1-2	1.21	0.41	.14**	-.06	.03	.03	-		
6. Education	1-2	1.53	0.50	-.01	.04	.03	-.01	-.04	-	
7. Employment	1-2	1.12	0.33	-.06	-.01	-.03	.06	-.01	.14**	-
8. Age	-	37.36	10.89	-.11*	-.14**	-.16**	.13**	-.14**	-.01	.05

Note. $N = 395$. Gender (1 = male, 2 = female). Ethnicity (1 = caucasian, 2 = minority). Education (1 = university, 2 = non-university). Employment (1 = employed, 2 = unemployed, student, and retired). KR-20 reliability coefficients are presented in parentheses. * = $p < .05$. ** = $p < .01$.

Discussion

Following the EFA process in this study, there are still a few remaining items which may have concerning properties. For example, the item CI 10 had 15.4% of respondents choosing the illogical response options, even though it fit well in the resulting factor structure. Item INS 7 had 77.5% of respondents choose the high JM (psychopathic) response which outnumbered the low JM response by more than a 3:1 ratio. Additionally, FLN 4 and INS 9 had significant loadings on both factors, which may be less of an issue considering that participants are given an overall score with the CRT-WP. While on the topic of factor loadings, it is evident that most of them in this study are much lower than the normal .50 - .70 threshold range (Conway & Huffcutt, 2003). However, this is the result of using a measure with categorical response options such as the CRT-WP. Item loadings,

scale reliabilities, and other parameters will likely be lower given the lack of range that a Likert-type scale has. That being said, the KR-20 reliabilities for Factor 2 and the overall CRT-WP are already acceptable, with only Factor 1 being considerably lower than the commonly used .70 cut-off for Cronbach's alpha.

After interpreting the 2 factors which resulted naturally from the EFA process, it appears that Factor 1 covers individual-oriented psychopathy (one's own thoughts and behaviours) and Factor 2 covers other-oriented psychopathy (thoughts and behaviours directed at others). The individual-oriented psychopathy factor includes all of the items designed to measure carefree impulsivity and fearlessness, along with items from the externalization and social superiority JMs that have one hypothetical character to focus on in the question. The other-oriented psychopathy factor includes all items designed to measure ruthless-self interest and insensitivity, along with items from externalization and social superiority that do not have one hypothetical character to focus on, but instead provide scenarios of how people interact with each other. For examples of this difference, the two social superiority items in the individual-oriented factor describe Stephen (SS 1) and Raheem (SS 9) and participants are asked a question about the individual's thoughts, motivations, or behaviours. Alternatively, the two social superiority items in the other-oriented factor describe situations that include how "some employees" (SS 2) and "some people in leadership positions" (SS 3) think about or behave toward others, and the participants are asked to respond based on that.

With the above interpretation in mind, the 2-factor structure is clear, meaningful, and makes intuitive sense in comparison to other psychopathy measures (see Table 3). For example, the "other-oriented psychopathy" factor covers the exact JMs covered by the primary factor of the LSRP, and the "individual-oriented psychopathy" factor covers almost exclusively the elements of the secondary factor of the LSRP. Thus, the CRT-WP maps onto the LSRP conceptualization very well.

The same can be said for the two-factor structure of the PCL-R/SRP. The “other-oriented psychopathy” factor of the CRT-WP appears to measure the same content of factor 1 of the PCL-R/SRP, which describes psychopaths as remorseless, manipulative, and having certain attitudes and behavioural tendencies. Similarly, the “individual-oriented psychopathy factor” of the CRT-WP seems to cover factor two of the PCL-R/SRP which consists of items that detail chronic instability, social deviance, and impulsive decision making. In comparison to the Triarchic conceptualization, the “individual-oriented psychopathy” factor of the CRT-WP seems to represent both disinhibition and boldness, while the “other-oriented psychopathy” factor consists of items which all measure meanness with the few exceptions noted in the paragraph above. Thus, the two factors of the CRT-WP also make sense within the Triarchic model. Finally, the two factors of the CRT-WP do not map onto the factors of the PPI as neatly, however it appears that they cover the factors of the PPI that they were intended to measure (e.g., “individual-oriented psychopathy” includes PPI factors such as carefree non-planfulness, impulsive nonconformity, and stress immunity; see Table 3).

Finally, the negative values for the means of the CRT-WP factors indicated that most people were choosing non-psychopathic options most often (evidenced by them scoring more -1's than 1's). As mentioned earlier, this was expected under the assumption that most people are relatively non-psychopathic and only few people are highly psychopathic. Factor 1 being significantly correlated with ethnicity (in that minorities scored higher on individual-oriented psychopathy) could be a concerning finding. However, when considering the overall CRT-WP, it was no longer related to ethnicity. The relationship between CRT-WP factors and ethnicity was scrutinised moving forward.

Study 3 – Two-Wave Study with Confirmatory Factor Analysis and Validity Assessment

The goal of Study 3 was to further demonstrate the reliability and validity of this refined version of the CRT-WP. In wave one of this study, participants were first presented with a consent

form, and were then be asked to complete the revised CRT-WP following Study 2, again with the item order being randomized for each participant. Equivalent to Study 2, the same three attention-check items were used in this study and were embedded within the mini-IPIP for wave one. Along with the same demographic questionnaire from Study 2, this is all that participants completed in this first wave of data collection.

One week later, participants were invited back to participate in a second wave of data collection via their MTurk identification numbers. In this wave, again participants were first presented with the CRT-WP. However, during this second wave of data collection they also completed two popular self-report measures psychopathy (SRP-III and TriPM), a self-report measure of counterproductive work behaviours (CWBs), and a measure of socially desirable responding, followed by a debriefing form. Having two waves of CRT-WP data along with the other measures of psychopathy and CWBs allows for multiple analyses in Study 3. First, confirmatory factor analyses (CFA) were conducted on wave one of CRT-WP data using Mplus version 7.4. It was hypothesized that the remaining items will fit best to the same meaningful factor structure that resulted from the EFA in Study 2, which was the 2-factor model of individual-focused psychopathy and other-oriented psychopathy detailed above. Rotated factor loadings and fit indices including chi-square, RMSEA, and CFI statistics were used to assess fit between competing models, consistent with Study 2.

Hypothesis 1: The same factor structure resulting from Study 2 will produce the best fit to the data, indicating a consistent structure within the 27-item version of the CRT-WP.

Second, the correlation between CRT-WP scores from wave one and wave two provides an estimate of test-retest reliability for the CRT-WP. The previous implicit CRTs in the literature have demonstrated good test-retest reliability in multiple studies (James et al., 2004, James et al., 2005,

Schoen et al., 2018). Therefore, participants' scores on the CRT-WP from wave one and wave two should be strongly correlated.

Hypothesis 2: Participants' CRT-WP scores from wave one and wave two will result in a strong positive correlation, demonstrating good test-retest reliability.

Additionally, the existing literature shows that scores on a covert CRT and overt self-report measures of the same construct do not strongly correlate (James et al., 2004, James et al., 2005, Schoen et al., 2018). More precisely, while most overt measures of the same construct typically correlate around .70 or higher, scores on an implicit CRT and overt measures for the same construct typically correlate around .30 or lower (James et al., 2005, Schoen et al., 2018). In theory, this is due to the difference in what information participants will give on overt and covert measures. To reduce concerns of common method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), CRT-WP scores from wave one will always be used in correlations with wave two variables.

Hypothesis 3: Participants' CRT-WP scores from wave one and SRP-III scores from wave two will produce significant positive, but small, correlations.

Hypothesis 4: Participants' CRT-WP scores from wave one and TriPM scores from wave two will produce significant positive, but small, correlations.

Existing CRTs have demonstrated the ability to predict objective outcomes of the construct they are measuring better than typical self-report measures (James et al., 2004, James et al., 2005, Schoen et al., 2018). Although the current study does not have an objective measure of psychopathic behaviours at work, self-reported perpetration of counterproductive work behaviours was used as a proxy. The relationship between psychopathy and increased CWBs was discussed earlier in this thesis, and it is well-documented in the literature (Boddy, 2014; Scherer et al., 2013). Therefore, it is expected that participants' scores on the CRT-WP will significantly predict their self-reported level

of CWBs. It is also expected that the CRT-WP will be able to predict CWB levels above and beyond the two overt self-report psychopathy measures in the current study. These findings would establish the ability of the CRT-WP to predict an outcome variable that is measured at a later time, and incremental validity of the CRT-WP. These hypotheses will be tested using regression analyses in SPSS version 24.

Hypothesis 5: CRT-WP scores will be a significant predictor of self-reported CWBs.

Hypothesis 6: CRT-WP scores will be able to predict self-reported CWBs above and beyond the SRP-III or TriPM.

Finally, the hallmark of the CRT-WP is that it is covert, which should prevent respondents from engaging in impression management or faking to artificially reduce their scores. However, to prove that the CRT-WP is contributing something to psychopathy measurement beyond what the existing overt self-report measures can already do, CRT-WP scores must be **unrelated** to socially desirable responding when the existing overt measures **are significantly related** to it under the same conditions and context. Therefore, a measure of socially desirable responding and impression management is included in the current study along with the following hypothesis.

Hypothesis 7: CRT-WP scores will not be significantly correlated with socially desirable responding, while the SRP-III and TriPM will be.

Methods

Sample. For wave one, data was collected from 301 MTurk users who did not complete Study 2. The same guidelines were followed, such as 90% or above approval rating and participation in at least 100 MTurk studies being required (as recommended by Hauser et al., 2018). Again, only MTurk users who are above the age of 18 and who have an IP address located in the United States or

Canada were included. It was necessary to collect data from at least 300 participants based on recommendations for CFA analyses that will be conducted in Study 3 and the expected removal of participants (Worthington & Whittaker, 2006). The same checks from Study 2 for respondent attentiveness were enforced here. Based on responses to the three attention check items (“strongly disagree” or “disagree” required for all three), 37 respondents were removed from the data. Due to the CRT-WP now containing 27 items, the new 25% cut-off for illogical responses was set at seven. Ten more participants were removed from the data based on having selected seven or more illogical responses on the CRT-WP. This resulted in a final sample of 254 participants for wave one.

These 254 participants had a mean age of 36.61 ($SD = 10.45$). The sample was 58.7% male, 39.8% female, and 1.6% identifying as “other.” Most participants were Caucasian (78%), while 8.3% were Black/African American, 4.7% were Hispanic, 3.9% were East Asian, 1.2% were South Asian, and 4% combined were Middle Eastern, Native/Aboriginal, Pacific Islander, or “other.” Regarding education, 68.9% reported having at least some college/university education, while 31.1% reported being a high school graduate. The sample was 88.6% employed (68.9% full-time), with 5.1% reporting that they were unemployed, 3.9% being students, 1.6% being retired, and 0.8% “other.” The mean completion time for the wave one survey package was 19.4 minutes ($SD = 7.41$).

As mentioned earlier, these 254 participants were invited back to complete wave two of data collection one week following wave one. The mean completion time for wave two was 25.85 minutes ($SD = 10.46$). The survey remained open for three days and 208 (81.89%) participants completed wave two. Using the same criteria as before, four participants were removed based on their responses to attention check items, and five participants were removed due to selecting seven or more illogical responses on the CRT-WP. Thus, the final sample for wave two included 199 participants that returned from wave one.

Measures.

Conditional Reasoning Test for Workplace Psychopathy. The 27-item version of the CRT-WP following Study 2 was used here, with no changes to the remaining items. The measure continued to be scored in the same +1, 0, and -1 format. Along with the overall 27-item score, factor scores were totaled for the individual-oriented and other-oriented psychopathy factors. The CRT-WP was included in both waves of data collection, but was the only measure of interest in wave one. For both waves, item order was randomized for each participant.

Self-reported psychopathy. Self-reported psychopathy was measured using two of the measures detailed earlier. First, participants were asked to complete an adapted version of the Self-Report Psychopathy scale third edition (SRP-III; Paulhus, Hemphill, & Hare, 2012). The regular version of the SRP-III is 64 items and measures four sub-scales. However, a shorter 34-item version produced by Mahmut, Menictas, Stevenson, and Homewood (2011) was used for the current study (Appendix E). This version was chosen because it is shorter, accessible online, and demonstrated powerful reliability and validity (Mahmut et al., 2011). This version also uses the original 5-point Likert-scale response format ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). Also, items load on to the same four sub-scales as the complete SRP-III, which are: callous affect (CA), erratic lifestyle (ELS), interpersonal manipulation (IPM), and criminal tendencies (CT). An example item is “I sometimes enjoy hurting the people who care for me” which measures callous affect. Note that the criminal tendencies sub-scale covers an aspect of psychopathy that was purposefully chosen not to be covered by the CRT-WP. As a result, participants were given two composite scores for the SRP-III, one that included all four sub-scales and one that included only the other three (CA, ELS, and IPM).

The second measure of self-report psychopathy was the brief version of the Triarchic Psychopathy Measure (TriPM; Patrick, 2010; Appendix F). The brief TriPM was chosen because it is

shorter than other measures, is accessible online, and covers what may be the most widely accepted model of psychopathy in the current literature (as discussed earlier). The brief TriPM consists of 58 items which measure the three factors of boldness, meanness, and disinhibition. Participants are asked to indicate their level of agreement with each item on a scale of 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). An example item is “I sometimes insult people on purpose to get a reaction from them” which measures meanness. Through multiple studies, the reliability and validity of the brief TriPM has been supported (Patrick, 2010). For the TriPM, both an overall score and individual factor scores were calculated for each participant.

Counterproductive work behaviours. CWBs were measured by the Counterproductive Work Behaviour Checklist (CWB-C; Spector et al., 2006; Appendix G). The CWB-C is a 33-item measure which consists of five subscales: abuse, production deviance, sabotage, theft, and withdrawal. Participants are asked to indicate how often they engage in each CWB via response choices ranging from 1 (*Never*) to 5 (*Daily or Almost Daily*). An example item is “Blamed someone at work for an error you made” which measures the abuse sub-scale. Respondents are typically instructed to answer with respect to their present job, but given that some MTurk users may not have another job in which they interact with people, they were asked to answer with respect to their last job if they do not currently have one. Spector et al. reported a reliability coefficient of .87 for the total scale, so total scores are given to participants.

Socially desirable responding. Behaviours related to socially desirable responding were measured using the 16-item version of the Balanced Inventory for Desirable Responding (BIDR-16; Hart, Ritchie, Hepper, & Gebauer, 2015). The original BIDR contains 40 items and measures two-factors; self-deceptive positivity and impression management (Paulhus, 1988). The 16-item version used in the current study contains 8-item sub-scales for both factors. Self-deceptive positivity is

intended to assess thoughts that an individual truly believes but are presented in an overly positive manner, while impression management measures deliberate false presentation (Paulhus, 1988). Thus, the impression management factor is of particular importance to the current study. The 16-item BIDR was chosen for the current study due to it being shorter while having scale properties that are similar to the original version (Hart et al., 2015). Respondents are asked to rate items on a 7-point Likert-scale from 1 (*Not True*) to 7 (*Very True*) based on the extent to which they agree with the statement. An example is “When I hear people talking privately, I avoid listening” which measures impression management. Only factor scores are calculated for the BIDR, not an overall score.

Results

Wave one. The means, standard deviations, reliability coefficients, and inter-correlations for all wave one variables (CRT-WP and demographics) are presented in Table 9. The frequency of response selections for each item was similar to that of Study 2, with no item having a meaningfully different distribution. To assess whether the 2-factor model found in Study 2 was the best-fitting

Table 9

Descriptive Statistics, Reliability Coefficients, and Correlations for Study 3 – Wave 1

	<i>Scale</i>	<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.	7.
1. CRT-WP Self	-13 to 13	-6.57	3.84	(.45)						
2. CRT-WP Other	-14 to 14	-3.06	5.51	.34**	(.65)					
3. CRT-WP Total	-27 to 27	-9.63	7.70	.74**	.88**	(.68)				
4. Gender	1-2	1.40	0.49	-.04	.03	.00	-			
5. Ethnicity	1-2	1.22	0.42	.09	-.03	.02	.02	-		
6. Education	1-2	1.46	0.50	-.02	.01	.00	-.07	-.09	-	
7. Employment	1-2	1.11	0.32	-.10	-.08	-.11	.04	.05	.09	-
8. Age	-	36.61	10.45	-.20**	-.13*	-.20**	.09	-.14**	.06	-.05

Note. $N = 254$. Gender (1 = male, 2 = female). Ethnicity (1 = caucasian, 2 = minority). Education (1 = university, 2 = non-university). Employment (1 = employed, 2 = unemployed, student, and retired). KR-20 reliability coefficients are presented in parentheses. * = $p < .05$. ** = $p < .01$.

factor structure, four CFAs were conducted with four different competing models. The analyses were conducted using Mplus version 7.4 (Muthén & Muthén, 2017) and section 5.2 “CFA with categorical factor indicators” from the Mplus user guide was followed. The first model tested was the 2-factor model which resulted from Study 2. The second was a 3-factor model which attempted to mimic the Triarchic conceptualization of psychopathy based on the CRT-WP content coverage found earlier in Table 3. The third was a 6-factor model which allocated one factor for the items of each JM. Finally, the fourth model had a 2nd Order structure where there was one factor for each JM at the first level, with each of those factors loading onto one overarching “psychopathy” factor. The fit indices for these four competing models are presented in Table 10, which appear to indicate that the 2-factor model was best, or at least tied with the 6-factor model, supporting hypothesis 1.

Table 10

Fit indices for Four Competing CFA Models in Study 3 – Wave One (27 items)

	χ^2	χ^2/df	CFI	RMSEA	90% CI		WRMR
					Lower	Upper	
2-Factor	326.75	1.01	0.99	0.01	0.00	0.03	0.85
3-Factor	330.66	1.03	0.97	0.01	0.00	0.03	0.85
6-Factor	313.57	1.01	0.99	0.01	0.00	0.03	0.83
2 nd Order	333.44	1.05	0.96	0.01	0.00	0.03	0.86

Note. $N = 254$. Only slight differences were observed between all four models, likely due to the categorical nature of response options on the CRT-WP.

Wave two. The means, standard deviations, reliability coefficients, and inter-correlations for all wave two variables are presented in Table 11. The frequency of response selections for each item were similar to that of Study 2 and Study 3 – wave one. KR-20 reliability coefficients for the CRT-WP and its’ two factors were larger in wave two than they were in wave one or Study 2, indicating good internal consistency. Supporting hypothesis 2, CRT-WP scores from wave one and wave two

Table 11
Descriptive Statistics, Reliability Coefficients, and Correlations for Study 3 – Wave 2

	<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
1. W1 CRT-WP Self	-6.74	3.74	(.42)																		
2. W1 CRT-WP Other	-2.82	5.47	.38**	(.64)																	
3. W1 CRT-WP Total	-9.56	7.72	.76**	.89**	(.68)																
4. W2 CRT-WP Self	-6.95	4.24	.62**	.32**	.53**	(.57)															
5. W2 CRT-WP Other	-3.41	5.65	.39**	.72**	.70**	.47**	(.68)														
6. W2 CRT-WP Total	-10.36	8.52	.57**	.64**	.73**	.81**	.90**	(.75)													
7. SRP-III Overall	1.96	0.53	.09	.24**	.21**	.09	.21**	.18*	(.92)												
8. SRP-III (No CT)	2.01	0.53	.16*	.27**	.27**	.12	.25**	.23**	.94**	(.90)											
9. TriPM Overall	2.25	0.40	.05	.18*	.15*	.05	.18	.14*	.86**	.85**	(.91)										
10. TriPM Bold.	3.02	0.69	.08	.04	.07	.04	.10	.09	.33**	.37**	.59**	(.90)									
11. TriPM Mean.	1.83	0.54	.11	.23**	.22**	.05	.24**	.19**	.77**	.83**	.81**	.25**	(.89)								
12. TriPM Disin.	1.91	0.59	-.09	.11	.03	.00	.04	.02	.66**	.54**	.61**	-.17*	.44**	(.90)							
13. BIDR SDE	4.45	1.06	.04	-.11	-.06	.02	.00	.01	-.07	.01	.11	.65**	.01	-.51**	(.78)						
14. BIDR IM	4.54	1.16	.06	-.16*	-.09	-.02	-.11	-.08	-.46**	-.39**	-.34**	.20**	-.38**	-.55**	.51**	(.81)					
15. CWB Overall	1.30	0.37	.06	.19**	.16*	.11	.13	.14*	.57**	.50**	.52**	.03	.46**	.59**	-.25**	-.46**	(.95)				
16. Gender	1.40	0.49	-.09	.04	-.02	-.04	.02	-.01	-.23**	-.27**	-.22**	-.21**	-.25**	.02	-.17*	.03	-.14	-			
17. Ethnicity	1.21	0.40	.06	.01	.04	.06	.01	.04	-.05	-.03	-.04	-.01	-.05	-.02	-.07	.01	-.02	.00	-		
18. Education	1.49	0.50	.00	.01	.01	.01	.02	.02	.16*	.10	.13	.01	.08	.18*	.02	-.02	.07	-.03	-.12	-	
19. Employment	1.10	0.30	-.11	-.10	-.12	-.02	-.06	-.05	-.18*	-.19**	-.17*	-.26**	-.08	.02	-.08	-.04	-.07	.05	.05	.06	-
20. Age	37.62	10.87	-.23**	-.17*	-.23**	-.16*	-.16*	-.19**	-.24**	-.27**	-.19**	-.07	-.16*	-.15*	.04	.20**	-.11	.12	-.13	.02	.00

Note. Listwise *N* = 195. Gender (1 = male, 2 = female). Ethnicity (1 = caucasian, 2 = minority). Education (1 = university, 2 = non-university). Employment (1 = employed, 2 = unemployed, student, and retired). W1 = Wave one. W2 = Wave two. SDE = Self-deceptive enhancement. IM = Impression management. “No CT” refers to the calculated SRP-III scores using the other three factors and not the criminal tendencies factor. KR-20 and Cronbach’s α reliability coefficients are presented in parentheses. * = $p < .05$. ** = $p < .01$.

were strongly correlated ($r = .73, p = .001$) demonstrating sufficient test-retest reliability.

Additionally, scores from both factors in wave one and wave two were strongly correlated as well.

As expected CRT-WP scores from wave one had significant weak positive correlations with both the SRP-III (with CT factor $r = .21, p = .001$, and without CT factor $r = .27, p = .001$) and the TriPM ($r = .15, p = .04$), supporting both hypotheses 3 and 4. The correlation with the overall TriPM is slightly lower than expected. However, the correlation between the CRT-WP and the meanness factor is stronger, and the meanness factor appears to be the one that is most central to the TriPM given other correlations. Finally, overall CRT-WP scores from wave one or wave two were not significantly correlated with either self-deceptive enhancement or impression management, as intended. Correspondingly, the SRP-III, the SRP-III without the criminal tendencies factor, the TriPM overall, and all of the TriPM factors were all moderately correlated (r 's between .20 and -.55) with impression management in the expected directions. The boldness and disinhibition factors of the TriPM were also moderately correlated with self-deceptive enhancement as well. Altogether, both overt measures of psychopathy being significantly related to socially desirable responding behaviours while the CRT-WP was unrelated, supports hypothesis 7.

The final analyses determined whether CRT-WP scores from wave one could significantly predict CWB scores from wave two. The correlation between CRT-WP scores from wave one and CWBs was significant ($r = .16, p = .02$), supporting hypothesis 5. However, two hierarchical regressions revealed that CRT-WP scores from wave one did not significantly predict CWB scores beyond the variance predicted by either the SRP-III or TriPM overall scores. Thus, hypothesis 6 was not supported. The results of these two hierarchical regressions can be found in Table 12.

Additional non-planned analyses were performed to test whether adding the interaction between each measure and the CRT-WP explained significantly more variance in a third step for

Table 12

Hierarchical Regressions Testing CRT-WP Scores Predicting CWBs Beyond the SRP-III and TriPM

Step	Predictor	Unstandardized Coefficients		Standardized Coefficients		Adj. R^2	ΔR^2	ΔF	p
		B	SE	β	p				
1						.32	.32	92.88	.001
	SRP-III Overall	.40	.04	.57	.001				
2						.32	.00	.45	.503
	SRP-III Overall	.39	.04	.56	.001				
	CRT-WP (Wave 1)	.01	.00	.04	.503				
1						.27	.27	71.54	.001
	TriPM Overall	.48	.06	.52	.001				
2						.27	.01	1.88	.172
	TriPM Overall	.47	.06	.51	.001				
	CRT-WP (Wave 1)	.01	.00	.09	.172				

Note. ($N = 195$) SE = Standard Error of B . First p -value indicates significance of predictor to that step of the model. Second p -value indicates significant change in F for that step of the model. A hierarchical analysis was performed for both the SRP-III and the SRP-III without the criminal tendencies factor, but there were no differences in significant results so only one is presented here.

each of the above hierarchical regressions. According to the “channeling hypothesis” overt and covert measures assess different parts of the same construct, and that the combination of both is necessary to assess the “full” construct (Bing et al., 2007). In both cases, the new third step was significant. Thus, the interactions between the SRP-III and CRT-WP ($\Delta F = 10.17$, $p < .01$, $\Delta R^2 = .034$, $\beta = .68$), and the TriPM and the CRT-WP ($\Delta F = 3.90$, $p = .05$, $\Delta R^2 = .01$, $\beta = .73$), were both helpful in predicting CWBs beyond the two pairs of measures alone.

Discussion

Overall the results of Study 3 are encouraging. The goal of this study was to provide further evidence for the reliability and validity of the CRT-WP, and having six of seven hypotheses

supported met that goal. The CFA indicating that the same 2-factor model fit the data well again leads one to believe that there is a consistent structure underlying the current 27-item version of the CRT-WP. The finding that the estimates of internal consistency were generally higher in this study is also encouraging. In wave two the overall CRT-WP had a KR-20 coefficient of .75, which is only a lower-bound estimate. The last point supporting the reliability of the CRT-WP is that wave one and wave two scores were highly correlated as hypothesized, indicating good test-retest reliability. For the CRT-WP to be useful in practice, it had to be demonstrated that an individual's psychopathy score remained relatively consistent at two measurement points, and this finding supported that.

Hypotheses 3 through 7 aimed to provide evidence for construct, convergent, discriminant, and criterion-related validity. In line with previous literature, the CRT-WP was weakly, yet significantly, correlated with both the SRP-III and the TriPM (James et al., 2004; Schoen et al., 2018). Although one of the correlations was slightly lower than expected, the notion that these correlation values fell within a hypothesized narrow range is strong evidence that the CRT-WP measures psychopathy at least as well as existing CRTs measure their constructs of interest. These low correlations were expected because although self-report measures of the same construct should highly correlate (as shown by the correlation between the SRP-III and TriPM), covert and overt measures assess the construct differently. In theory, overt measures only assess levels of a construct that participants are willing to report, whereas covert measures such as the CRT-WP get an indication of the construct that is uninhibited by the participants' willingness to report (also mentioned by the "channeling hypothesis" referred to in the additional analyses).

Regardless of any other findings, if the CRT-WP was found to be related to socially desirable responding behaviours, it would not have much promise. However, CRT-WP scores being unrelated to self-deceptive enhancement or impression management is significant evidence that the measure is working covertly as intended. This study also showed that the existing overt self-report measures of

psychopathy are highly susceptible to impression management, despite their widespread usage and popularity. These two findings in combination are strong theoretical support for the CRT-WP contributing something significant to both the literature and practice of psychopathy measurement.

CRT-WP scores from wave one alone were significantly related to CWB scores from wave two, although it was weak in magnitude. It would have been a more impressive result if the relationship was stronger, but it can still be viewed as a success that this version of a covert measure which is still being developed was able to predict the outcome variable which was measured at a different time point. Additionally, given the methodological choice of not including anti-social/criminal behaviour in the conceptualization of the CRT-WP, it was unlikely that its' correlation with CWBs was going to be as strong as that of other measures. For example, the SRP-III including the criminal tendencies factor had a much stronger correlation with CWBs because there was overlap between some factors of the measure (criminal tendencies) and the outcome (theft). However, it is difficult to argue that the CRT-WP not being a significant predictor of CWBs beyond the SRP-III or the TriPM is a discouraging finding. However, the likely explanation is that CWBs were also self-reported, and at the same time point as the SRP-III and TriPM. Being measured at the same time point with the same method may artificially inflate the relationship through common method bias (Podsakoff et al., 2003). The fact that CWBs were also self-reported in an overt manner means that it is also susceptible to self-deceptive enhancement and impression management similar to the issues with the SRP-III and TriPM. This troublesome confound is indicated by the significant correlations of CWBs with both SDE ($r = -.25, p = .001$) and IM ($r = -.46, p = .001$). Thus, it would appear that impression management particularly is a significant concern with the measure of CWBs included in this study, as many people may not have responded accurately. If this were the case, it would make intuitive sense that the self-report measures of psychopathy would be much more related to self-reported CWBs than the covert CRT-WP would be. In future studies, it would be better to

include an objective measure of CWBs to determine whether the CRT-WP would be able to get a more accurate reading than what people are willing to admit.

Lastly, there is a consistent negative correlation between all three psychopathy measures and age. These correlations are likely evidence of the steady decline of “negative” personality traits with age, including psychopathy (Olver & Wong, 2015). If there is one underlying explanation to these relationships, this can be seen as further evidence that the three measures are evaluating the same construct, tendencies, and behaviours. Also notice that the CRT-WP is not correlated with any other demographic variables whereas the SRP-III and TriPM are both significantly correlated with gender, with males scoring higher in psychopathy. One explanation is that males truly do score higher in “true” psychopathy, and there may be a fault with the CRT-WP. On the other hand, it could be that the existing conceptualizations of psychopathy (especially the PCL-R and SRP-III) could be biased toward measuring traits and tendencies that are more common among male psychopaths since they were primarily developed with only male samples. Indeed, past research has suggested that the PCL-R and SRP framework of psychopathy simply does not cover the female-specific psychopathic traits and tendencies (Forouzan & Cooke, 2005). Perhaps the CRT-WP may not have this criticism.

Study 4 – Additional Validation with a Student Sample

Study 4 aimed to provide additional validation using a student sample so that there is some diversity in participants across studies, as in Schoen et al. (2018). This study was very similar to the second wave of Study 3, in that many measures were included in a survey and the goal was to examine the correlations between measures. However, there is still important information to be learned with regard to how scores on the CRT-WP correlate with measures other than those in Study 3. The survey package included in this study first included a similar consent form and the same version of the CRT-WP from Study 3. Following the CRT-WP, this sample was also asked to

complete a measure of general personality and two measures of academic dishonesty/cheating, in that order, followed by a demographics section and a debriefing form. Again, the same three attention-check items were embedded within other measures.

Up to this point, there has not been a standard measure of personality included in these studies. Relating the CRT-WP to the “big five” personality traits for example, would both add to the nomological network of the CRT-WP and lend insight into whether the CRT-WP has the same relationships that other measures of psychopathy have with these personality factors. Given that psychopaths are described as manipulative and lying, it is likely that scores on the CRT-WP will negatively correlate with the honesty-humility scale of the HEXACO, which has been supported by recent research (Lee & Ashton, 2014). The HEXACO is a measure of the usual “big five” personality traits, plus the honesty-humility factor. Additionally, the insensitive and impulsive psychopath may also be less agreeable than others, and in fact a negative relationship between the two constructs has also been demonstrated by previous research (Lee & Ashton, 2005). Otherwise, it is unclear whether CRT-WP scores will significantly correlate with any of the other factors of the HEXACO or in which direction. Given the above points, the following hypotheses are presented.

Hypothesis 8: CRT-WP scores will significantly negatively correlate with honesty-humility.

Hypothesis 9: CRT-WP scores will significantly negatively correlate with agreeableness.

Finding a link between scores on the CRT-WP and academic dishonesty/cheating would further increase the criterion-related validity of the CRT-WP. There is considerable research which suggests that students with higher levels of psychopathic traits tend to engage in more cheating behaviours (Coyne & Thomas, 2008; Nathanson, Paulhus, & Williams, 2006; Williams, Nathanson, & Paulhus, 2010; Ternes, Babin, Woodworth, & Stephens, 2019). Therefore, it is expected that scores on the CRT-WP will be significantly positively correlated with self-reported academic

cheating and condoning cheating behaviours, which are both measured in this study. This result, along with the relationship between the CRT-WP and counterproductive work behaviours in Study 3, would increase construct validity and show practical significance.

Hypothesis 10: CRT-WP scores will be significantly positively correlated with self-reported academic cheating.

Hypothesis 11: CRT-WP scores will be significantly positively correlated with condoning academic cheating behaviours.

Method

Sample. Student participants from Saint Mary's University were recruited through the online research participation system (SONA) and a research assistant who gave paper copies to students who were sitting in campus common areas. The latter participants were given a paper and pencil version of the survey package and they were asked to complete it while they remained seated in the common area that they were found. The research assistant then remained in the area while the student completed the survey package, and if they did not have time to complete the whole package, they were still asked to return it to the research assistant. This recruitment process was structured this way so that the environment was somewhat monitored, and so that students could not leave with the survey package which has items and data that should be kept confidential. Students who completed the paper version were compensated with a \$5 gift card, and those who completed the online version received extra credit in a psychology course. For the paper version, the research assistant handed physical copies of the consent and debriefing forms to the students on separate pieces of paper.

There were 146 completed survey packages combined between the two versions. Once the paper copy data was entered into the data file, the responses went through the same checks for data quality as in the previous two studies. First, 33 respondents were removed from the data file based on

their responses to the attention check items being anything other than “strongly disagree” or “disagree” in all three cases. Eight more participants were then deleted for having selected seven or more illogical response options on the CRT-WP, leaving a final sample of 105 participants.

Of the 105 participants in the sample, 77 completed the online version and 28 completed the paper version of the survey package. The mean age was 21.83 ($SD = 3.98$) which is much lower than previous studies due to the sample consisting of university students. The sample was 22.9% male and 77.1% female, which is also noticeably different from the samples in previous studies. The majority of the students were Caucasian (66.7%), 19% were Black, 3.8% were Middle Eastern, 1.9% were East Asian, 1.9% were South Asian, 1% was Native/Aboriginal, and 5.7% responded with “other.” All of the participants except for one (99%) were undergraduate students. Regarding their employment status, only 34.3% reported being employed (5.7% full-time), and 65.7% reported either that they were unemployed, just a student, or “other.” Participants reported their major in a text field, but combining their responses into categories indicates that 43.2% were psychology majors (or double-majors with one being psychology), 15.8% were business majors, 7.7% were biology majors, 7.7% were criminology majors, 3.9% were environmental science majors, 2.9% were anthropology majors, 2.9% were computer science majors, and the remaining 15.9% indicated other majors or were undeclared. The average time to complete the survey package cannot be accurately determined for this study because participants were permitted to keep the online version open on their web browser for as long as they liked, which led to some very high completion times.

Measures.

Personality. The HEXACO-60 (Ashton & Lee, 2009; Appendix H) was used to measure personality. The HEXACO model includes the traditional big five personality traits (emotional stability, extraversion, agreeableness, conscientiousness, and openness) and a sixth important factor

labelled honesty-humility. The HEXACO-60 was chosen over other big five measures of personality because the unique honesty-humility factor is of particular relevance to psychopathy, and because it is accessible online. The 60 items measure the six factors listed above, and participants respond to each item based on the extent to which they agree it describes them, on a scale of 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). An example item is “I would never accept a bribe, even if it were large” which measures the honesty-humility factor. The HEXACO-60 has consistently demonstrated good validity and scale reliability (Ashton & Lee, 2009).

Academic Dishonesty/Cheating. The first measure of academic dishonesty and cheating was an adapted version of the 12-item scale used in McCabe and Trevino (1993; Appendix I). This measure was chosen because it includes more common academic cheating behaviours, and it has good internal consistency (McCabe & Trevino, 1993). Participants are asked to indicate how often they have engaged in each behaviour on a scale of 1 (*Never*) to 4 (*Many Times*). Two example items are: “copying from another student during a test” and “copying material and turning it in as your own work.” The current study adapted this measure by dropping two items which seem to overlap with others (i.e., “copying a few sentences from a published source” overlaps with “copying material and turning it in as your own work”), and re-wording some of the items so that they reflect modern language (i.e., changing “using crib notes on a test” to “using unauthorized notes during a test”). The adapted version used by this study contains 10 items.

The second measure of academic dishonesty in the current study was a four-item measure from Pulfrey and Butera (2013) which assesses the degree to which students condone cheating behaviours. This measure was chosen because it does not ask about the student’s specific history with cheating, and instead measures their attitudes toward others cheating which may lead to more honest responding. The four items are: “I can imagine that some students might copy off the Internet

without citing,” “I can understand it if some students copy off others,” “Some students probably get external help for their coursework,” and “Getting outside help to do coursework is no big deal.”

Participants indicate the degree to which they agree with each item on a scale from 1 (*Totally Disagree*) to 7 (*Totally Agree*). Pulfrey and Butera found acceptable internal consistency ($\alpha = .72$).

Results

Means, standard deviations, reliability coefficients, and correlations for all Study 4 variables is presented in Table 13. Reliability coefficients for the CRT-WP and its' two factors were considerably lower in this study. Despite the consistent negative correlations between the CRT-WP and age in previous studies, this sample of university students had lower mean CRT-WP scores than previous samples. However, overall CRT-WP scores were significantly negatively correlated with the honesty-humility factor of the HEXACO ($r = -.31, p = .002$), supporting hypothesis 8. On the other hand, overall CRT-WP scores were not significantly correlated with the agreeableness factor ($r = -.11, p = .28$), failing to support hypothesis 9.

Neither cheating behaviour ($r = -.01, p = .89$) nor attitudes toward cheating ($r = .05, p = .64$) were significantly correlated with CRT-WP scores, failing to support hypotheses 10 and 11. On the other hand, honesty-humility was significantly correlated to both cheating behaviour and attitudes toward cheating in the expected direction. With honesty-humility being related to academic misbehaviour, and the CRT-WP being related to honesty-humility both in the expected directions, it is curious that the CRT-WP was not related to academic misbehaviour. For consistency, the same additional analyses conducted in Study 3 to test the channeling hypothesis were also conducted here. However, for both outcome variables, the interaction between honesty-humility and the CRT-WP did not add significant prediction beyond the two measures alone.

Table 13

Descriptive Statistics, Reliability Coefficients, and Correlations for Study 4 Variables

	<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. CRT-WP Self	-6.88	3.37	(.34)														
2. CRT-WP Other	-5.64	4.21	.18	(.46)													
3. CRT-WP Total	-12.52	5.85	.71**	.82**	(.48)												
4. Honesty-Humility	3.28	0.57	-.12	-.32**	-.31**	(.72)											
5. Emotionality	3.50	0.67	-.09	.05	-.02	.05	(.84)										
6. Extraversion	3.37	0.64	.01	-.12	-.08	-.08	-.20*	(.82)									
7. Agreeableness	3.18	0.55	-.07	-.09	-.11	.02	-.13	.07	(.75)								
8. Conscientiousness	3.69	0.53	-.22*	.07	-.08	.17	.14	.20*	-.01	(.77)							
9. Openness	3.25	0.59	-.00	-.01	-.01	-.01	-.09	.12	.17	-.08	(.69)						
10. Cheat Behaviour	1.43	0.50	.10	-.10	-.01	-.28**	-.02	.06	-.12	-.11	-.27**	(.88)					
11. Cheat Attitudes	5.32	0.83	-.04	.10	.05	-.38**	.00	-.01	-.02	.00	-.05	.28**	(.44)				
12. Gender	1.77	0.42	-.05	-.06	-.08	.15	.27**	-.13	-.11	.16	.02	-.04	-.01	-			
13. Ethnicity	1.34	0.47	.05	.12	.12	-.28**	-.04	.07	.01	-.04	.02	.25**	.14	-.19	-		
14. Employment	1.65	0.48	.05	.09	.09	.06	.07	.01	.10	.18	-.02	.10	.02	-.11	.05	-	
15. Age	21.87	3.98	.09	.21*	.21*	.03	-.13	.11	-.15	-.14	.19	-.14	-.18	-.07	.09	-.20*	-

Note. Listwise $N = 104$. Education not included due to all but one participant being an undergraduate. Gender (1 = male, 2 = female). Ethnicity (1 = Caucasian, 2 = minority). Employment (1 = employed, 2 = unemployed, student, and other). KR-20 and Cronbach's α reliability coefficients are presented in parentheses. * = $p < .05$. ** = $p < .01$.

Discussion

The results of Study 4 were much less supportive of the CRT-WP than those of previous studies. Reliability coefficients were much lower, indicating that the measure may not have functioned the same way for students as it did for MTurk samples. However, upon splitting the data file by online and paper participants, the KR-20 coefficient for the CRT-WP was noticeably higher for paper ($KR-20 = .57$) than online ($KR-20 = .45$). This may indicate that online participants had more difficulty with some items more than others or that they responded less consistently, but this is still a strange result considering that the MTurk samples also completed the survey packages online. The difference could be that the student participants in Study 4 were permitted to keep the survey window open as long as they wanted (even overnight), whereas MTurk participants had a time limit. Additionally, CRT-WP scores being unrelated to academic cheating and attitudes toward condoning cheating does not support its' criterion-related validity in the student setting. However, another element to consider is that (similar to CWBs in Study 3) both academic cheating and attitudes toward cheating were overt self-report measures. Thus, many participants may have lowered their scores on these two measures for purposes of social desirability, which is supported by a very low mean for cheating behaviour. However, other research has found that students will self-report academic cheating to high degrees (McCabe, Butterfield, & Trevino, 2012). Regardless, to explore this possibility it would have been beneficial for this study to also include the BIDR or another measure of socially desirable responding, as Study 3 did. Alternatively, future research could use a game-based task to measure cheating behaviour more covertly, instead of a self-report measure.

In line with previous research, psychopathy scores as measured by the CRT-WP were negatively correlated with honesty-humility (Lee & Ashton, 2014). This is the one result of the current study that seems to support the CRT-WP is measuring what it is intended to. CRT-WP Other

(other-oriented psychopathy) was related to honesty-humility, while CRT-WP Self (individual-oriented psychopathy) was not. This finding suggests that JMs such as ruthless self-interest and insensitivity are more related to honesty-humility beliefs and behaviours than fearlessness, for example. The relationship between the CRT-WP and an individual's level of honesty-humility is meaningful, as honesty-humility is related to self-reported outcomes such as academic misconduct in this study.

Contrary to previous research (e.g., Lee & Ashton, 2005), agreeableness was unrelated to psychopathy scores as measured by the CRT-WP. This finding may indicate that the CRT-WP measures something different than other measures of psychopathy which have found this significant relationship. No other factors were significantly related to overall CRT-WP scores, but conscientiousness was significantly negatively correlated to CRT-WP Self. This correlation suggests that more conscientious people are less impulsive and less fearless (more fearful), which makes sense considering that highly conscientious people are often described as working toward their goals, striving for perfection, and deliberating carefully before making decisions (Ashton & Lee, 2009). Finally, age had the inverse relationship with the CRT-WP than found in previous studies, though the age range was much more restricted given the student sample.

General Discussion

As other researchers in the field have noted, a faking-resistant self-report measure of psychopathy would be of tremendous use in the employment selection process (Smith & Lilienfeld, 2013; James et al., 2004; Mathieu & Babiak, 2016b; Mathieu et al., 2013; Wiita et al., 2017). Wu and LeBreton (2011) have gone even further in saying that implicit measures of dark personality traits are needed for practical use. Specifically, CRTs have been suggested as the most promising future direction for measuring dark personality traits at work (Spain, Harms, & LeBreton, 2014). Catano et

al. (2016) generally suggest that the outlook for personality measures in employment selection is good, however, most of the research has focused on the Big Five model. Catano et al. cite three main flaws surrounding the use of personality measures in employment selection. The first is that “personality” or the target construct is sometimes poorly defined, however, the CRT-WP aims to dodge this pitfall since the justification mechanisms were developed using an abundance of previous literature and taking all conceptualizations into consideration. Second, personality measures are often used without any direct connection to job performance. As mentioned above, the content of the JMs and items being contextualized in relevant work situations should help with this issue. The third flaw is that self-report personality inventories are subject to faking and socially desirable responding since there is a clear incentive for applicants to present themselves in the best way possible. However, the foundation of the CRT-WP is that it attempts to avoid this criticism. Therefore, it appears that the CRT-WP has an answer to all of these issues, while also responding to the calls of many researchers.

The results of Studies 1 through 4 intended to provide the main basis for the CRT-WP, which would be a measure that would have an answer to all the points made in the previous paragraph. Overall, the results are largely positive, despite some unsupported hypotheses which were likely due to methodological issues. A summary of all hypotheses is presented in Table 14.

Study 1 provided the item revision process from SMEs, which gave valuable information about the structure of some items, the definitions of the JMs, and which items would not function well when given to participants. The confusion and overlap between the “carefree impulsivity” and “fearlessness” JMs was addressed, and now the definitions are clearer in what they intend to measure for future research. Although the factor structure resulting from the EFA in Study 2 was not the 6-factor model (one per JM) that was proposed, the 2-factor model produced was interpretable enough to be valuable. As noted earlier, previous CRTs which have been created contain items which assess

Table 14

Summary of Hypotheses

<i>Hyp.</i>	<i>Study</i>	<i>Brief Summary</i>	<i>Result</i>
1	3-1	Same factor structure from Study 2 would provide best fit to the data	Supported
2	3-2	CRT-WP scores from W1 and W2 would strongly positively correlate	Supported
3	3-2	CRT-WP would weakly, yet significantly, correlate with SRP-III	Supported
4	3-2	CRT-WP would weakly, yet significantly, correlate with TriPM	Supported
5	3-2	CRT-WP scores from W1 would sig. predict CWB scores from W2	Supported
6	3-2	CRT-WP (W1) would predict CWBs above and beyond SRP-III/TriPM	Not Supported
7	3-2	CRT-WP unrelated to SDR while SRP-III/TriPM both related to SDR	Supported
8	4	CRT-WP negatively correlated with honesty-humility	Supported
9	4	CRT-WP negatively correlated with agreeableness	Not Supported
10	4	CRT-WP positively correlated with academic cheating behaviours	Not Supported
11	4	CRT-WP positively correlated with attitudes toward academic cheating	Not Supported

Note. Hyp. = Hypothesis. W1 = wave one. W2 = wave two. SDR = socially desirable responding.

more than one JM and still have meaningful factor structures (James et al., 2004; Schoen et al., 2018). The 2-factor structure appears to be somewhat consistent as well, considering that it was the model which provided the best fit (or arguably tied with the 6-factor model) in wave one of Study 3. As mentioned earlier, this 2-factor structure is similar to existing models of psychopathy which have “primary and “secondary” factors such as the LSRP. Regardless, the CRT-WP is intended to be used similar to existing CRTs where one overall score measures the construct.

Reliability coefficients across Study 2 and 3 were also sufficient for the two factors and the overall CRT-WP, though those in Study 4 were noticeably lower. CRT-WP Self (individual-oriented psychopathy) had a consistently lower reliability coefficient than CRT-WP Other (other-oriented psychopathy) or the overall CRT-WP in every sample. This could result from the few items remaining in the CRT-WP after Study 2 which have some concerns. For example, item CI-10 had over 15% illogical response selections in Study 2 which is much higher than recommended (James & McIntyre, 2000) and remains in the CRT-WP Self factor (which persisted with 17.3% in Study 3

wave one, 15.6% in Study 3 wave two). Items FLN-4 and INS-9 cross-loaded between both factors. Lastly, item INS-7 had between 73-79% choose the psychopathic response option throughout Studies 2-4 (abnormal relative to other items) and also remains in the CRT-WP Other factor. Altogether, there may still be a small group of items which could be removed to improve some psychometric properties. However, these 27 items were kept based on factor loadings and structural fit in Study 2.

The labels given to the two factors, individual-oriented and other-oriented psychopathy respectively, were based on the theoretical grouping of JMs and items that are found within each factor. Some of the correlations in Study 3 and 4 would support this conceptualization of the two factors, however, in most cases CRT-WP Self simply has weaker correlations than CRT-WP Other with the same variables. Moving forward, it is not certain whether the individual- and other-oriented labels will be supported, or if there will prove to be a clearer description of what separates these two factors. For example, with additional research it may be argued that the factors should be labelled “Primary” and “Secondary” in order to create more links with existing measures of psychopathy.

Test-retest reliability was assessed in Study 3 and appears to be strong based on this one result. The correlation between wave one and wave two CRT-WP scores was comparable to test-retest reliability values for previous CRT measures (James et al., 2004; Schoen et al., 2018). Additionally, the weak, yet significant, positive correlations between the CRT-WP and other measures of psychopathy were in line with previous CRT research (James et al., 2004; Schoen et al., 2018), supporting construct validity. As mentioned earlier, overt and covert measures do not function the same, and participants do not respond the same way, which is the theoretical reason for why they do not correlate strongly. If the CRT-WP were to have been unrelated to the other measures of psychopathy at all, it would have reflected very poorly on the construct validity of the CRT-WP. However, according to the channeling hypothesis mentioned earlier, the combination between overt

and covert measures assess the construct more completely, which was supported by the significant interactions in Study 3 but not Study 4 (Bing et al., 2007). The fact that all three measures of psychopathy in Study 3 had similar relationships with age is also evidence that they are measuring a similar construct, albeit in different ways. Finally, the significant negative correlation between the CRT-WP and the honesty-humility factor of the HEXACO was expected, based on previous research using other psychopathy measures (Lee & Ashton, 2014).

Evidence for criterion-related validity is weaker. CRT-WP scores from wave one did significantly predict self-reported CWBs from wave two, which is some noteworthy support even though the prediction was not large in magnitude. However, previous CRTs have demonstrated incremental validity in that they can in fact predict meaningful outcomes more than what overt measures can predict (James et al., 2004; Schoen et al., 2018). This was not found in the current research, as the CRT-WP did not predict CWBs above and beyond the SRP-III or TriPM, failing to provide additional evidence for criterion-related or incremental validity. One possible explanation is that if one's goal was to predict CWBs, it would be better to use the SRP-III or TriPM even though people are engaging in impression management and similar behaviours. Another explanation that could explain a lack of criterion-related validity is that the outcome variable (CWBs) was also measured by an overt self-report measure. As mentioned in the discussion of Study 3, CWBs and both overt psychopathy measures were all significantly moderately correlated with impression management, while the CRT-WP was not. Thus, it is highly possible that participants adopted a response style throughout completion of the survey package, that could have been honest or dishonest. For example, one participant may have gone through all of the measures wishing to present themselves more favourably than they may score in reality, which would consist of lowering CWBs and lowering psychopathy together. If this were the case, they would not have been able to "present" that same image in their CRT-WP score due to their uncertainty in what it is measuring.

The CRT-WP was also unrelated to both measures of academic dishonesty in Study 4. Although the literature indicates that students who score higher in psychopathy engage in more cheating behaviours (e.g., Ternes et al., 2019) the CRT-WP could not capture this relationship. Similar to the previous study, one explanation is that if one's goal was to predict academic misconduct, using the honesty-humility factor of the HEXACO-60 would work better than the CRT-WP since it was significantly related to both measures of academic dishonesty. However, since both measures of academic misconduct were also overt self-report measures, the same alternate explanation can be presented here. It may be possible that many students engaged in socially desirable responding to some degree, and thus artificially lowered their scores on the two outcome measures (especially past cheating behaviours which had a very low mean). Altogether, if an objective outcome measure was used (academic demerits for students and performance appraisals for employees, for example) there would be much more insight into which measures are most predictive of the outcome of interest. The recent CRT for creative personality did use an objective measure in their study which found that the CRT was more predictive than overt measures (Schoen et al., 2018).

Other than one significant correlation between ethnicity and the CRT-WP Self (individual-oriented psychopathy) factor, the CRT-WP was not significantly related to gender, ethnicity, education, or employment, which is preliminary evidence that it would not contribute to adverse impact, though more research is needed to be certain¹. This is significant considering that many other measures of psychopathy have been linked to significant scoring differences in visible minorities (Gatner, Blanchard, Douglas, Lilienfeld, & Edens, 2018).

¹ MANOVAs comparing White, Black, Asian, and Hispanic participants were conducted for Study 2, 3-1, and 3-2 (Group sizes were not large enough for Study 4). In all cases, there were no significant differences for the overall CRT-WP or either of the two factors. The samples from Study 2, 3-2, and 4 were then combined into one data set of 700 participants. Using this large combined sample, A MANOVA with much higher statistical power was conducted comparing White, Black, Asian, and Hispanic participants which again resulted in no significant differences.

With lower reliability coefficients and less hypothesized relationships being confirmed, it is not certain why the CRT-WP functioned worse in Study 4. However, one explanation is the difference between samples. Study 2 and Study 3 both used MTurk participants and in both of these studies the CRT-WP had similar scale reliabilities, factor structures, and more. The student participants of Study 4 may not have responded to items using the same thought processes as MTurkers, they may not have had enough work experience to apply to the items (as in their mind it is about workplace situations), or there could have been significant noise in the data from having two means of data collection. It is possible that many of the MTurkers in the Study 2 and Study 3 samples complete many surveys every day and respond in more “consistent” response patterns than students. The explanation which touches on work experience is an intuitive one because the CRT-WP is presenting them with different workplace situations and participants believe they are supposed to select the most logical response. Therefore, people with more work experience should feel more comfortable and have a wealth of past experiences to draw on when responding to these questions, whereas many university students may have never had a job yet. In theory, this lack of experience would put them at a disadvantage or have them believing they needed to go through additional thought processes when responding to each item, even though this is not the case. The previous research developing the CRT for creative personality (Schoen et al., 2018) also used both MTurk and student samples and that CRT functioned well enough in both cases.

Implications

From the results of these studies, it is fair to say that a preliminary version of the CRT-WP has been developed and gone through partial validation. There is more work to be done, but the current research has taken many steps forward in providing the first faking-preventative and implicit measure of psychopathy in the workplace. This research is answering the call of the previous

researchers who have recognized the need for implicit measures of dark personality traits, specifically psychopathy, for use in the selection process (Smith & Lilienfeld, 2013; James et al., 2004; Mathieu & Babiak, 2016b; Mathieu et al., 2013; Spain et al., 2013; Wu & LeBreton, 2011).

The CRT-WP would likely not see any challenges in meeting professional guidelines for selection instruments considering the CRT for aggression has already seen use for over a decade. The issue of whether HR managers or hiring professionals are qualified enough to give this test is also not relevant as it is self-report. The interpretation of scoring by these individuals could potentially be an issue, however, the CRT-WP would eventually come with a guide for its usage that would detail exactly how to score respondents and what those scores mean. Informed consent, access to results, and confidentiality concerns (three critical notions for selection measures by Catano et al., 2016) should all be of minimal consequence, as applicants could still receive instructions and a debriefing of their general performance (good or bad, with percentiles) without actually revealing the implicit nature of the test, if desired. After completion of the test, applicants could be informed about the true nature of the CRT-WP for full transparency. The CRT-WP is also less legally invasive than other more clinical measures of psychopathy such as the PCL-R, as it cannot be argued as a “medical examination” of mental health, disorders, or the like. The CRT-WP measures psychopathic personality through attitudes, tendencies, and biases, but does not intend to make a clinical judgement as respondents are self-reporting and there are no professional raters.

Ultimately, the main implication is that the current research has laid the groundwork for a measure that would be able to screen out corporate psychopaths in the personnel selection process. All of the negatives surrounding corporate psychopaths in the workplace were detailed in the introduction, but let it be said that having a measure like this one would result in many positive outcomes including much healthier workplaces, less risky environmental practices, and many

employees not getting bullied or abused, just to name a few. However, there are still more studies which need to be conducted to further validate the use of the CRT-WP in such a situation.

Limitations and Future Research

The current research has both strengths and limitations. One strength is the use of a multi-stage development process which saw the initial item list go through four stages of revisions before being sent to SMEs. The study also followed strict scale development standards according to Hinkin (1998), including achieving large samples for both the EFA and CFA, as well as conducting the proper analyses at each of those stages. Also, strict standards for developing an implicit CRT following James (1998), James and McIntyre (2000), James et al. (2004), and Schoen et al. (2018) were adhered to including a lengthy review of all existing conceptualizations of the construct, as well as developing justification mechanisms and providing a content coverage map using existing measures. Another strength is the use of different samples, even if it may have resulted in some unexpected inefficiencies (e.g., much lower reliability coefficients). Finally, having two waves of data collection for Study 3 allowed for a proper assessment of test-retest reliability while also reducing common method bias in the analyses.

The first potential limitation is perhaps relying on factor structure and loadings more than some other standards set for CRT items during the EFA stage. As noted, there are still three or four items which *could* be functioning differently from the others, and the measure may benefit overall from these items being removed to reduce noise. Additionally, even though it follows Hinkin's (1998) recommendations for scale development, conducting EFAs with only one sample means that the use of a different sample could have resulted in an entirely different set of items passing the EFA stage. Another limitation was the reliance on self-report measures for all outcome variables. As mentioned before, the key component of the CRT-WP is that it is covert and measures what it

intends to without the participant being aware. Following that logic, and seeing some of the correlations with the BIDR, it would have provided more insight into the predictive validity of the CRT-WP if objective outcome measures were used. Therefore, future research should use objective measures of outcome variables, if possible. Overt self-report measures are limited in their application and an accurate, uninhibited, assessment of “dark” or antisocial behaviours is especially needed in this case. Studies conducted with employees or job applicants could use performance appraisals, or documented complaints, for example. Lastly, perhaps the most lacking piece of preliminary validation is an indication of external validity. The CRT-WP is intended to be eventually used in the application or promotion process where the respondents will be applying for a job. The MTurkers and students included in the current research were not completing the CRT-WP as part of an application process. Thus, to determine whether the CRT-WP should be used in the situation it is designed to, a similar study (or set of studies) would have to be conducted with participants who are going through the personnel selection process (Catano et al., 2016). If the CRT-WP is to be legally defensible as a selection tool, these studies must demonstrate the same results found in the current research also apply to samples of participants who are currently applying for jobs or promotions. This is the only way to ascertain that the CRT-WP is reliable, valid, and still covert when used for the main purpose it was designed for, high-stakes situations.

Other than the specific limitations mentioned above, future research should seek to replicate the current findings to provide more support (or lack of support) for the CRT-WP as a measure of psychopathy in the workplace. Another issue that remains uncertain is the designation of cut-off scores. Although the previous CRT for aggression has simply assumed that higher scores mean more extreme aggressiveness, there is no definitive answer as to how much aggressiveness is too much or how much would be acceptable in different jobs (i.e., hospital manager vs. professional football coach). This is another key point that would have to be addressed before the CRT-WP is completely

legally defensible. It is likely that the most accepted option would be to follow the CRT-A and not provide strict cut-off scores for use in practice. Therefore, organizations could simply compare the scores of multiple applicants and make decisions on a case-by-case basis, rather than any normative cut-off points to be applied in all cases. This would also help with the issue of providing feedback, as the organization could say they believe the applicants behavioural and attitudinal tendencies are not in line with the organization's ideals, based on their score on the CRT-WP. Thus, organizations would also avoid the issue of labelling applicants as psychopaths and creating false positive situations due to them scoring above a critical point. In addition, applicants have the rights to both accommodation and additional language options in the selection process (Catano et al., 2016). Thus, it is no doubt that alternative versions of the CRT-WP would have to be validated with other languages and administrative procedures.

Recall that even when given an incentive, such as being told that results would influence chances of being hired, CRTs have still proved resistant to faking as respondents cannot distort their scores if they are oblivious to what is truly being measured (LeBreton et al., 2007; Wiita et al., 2017). However, this is all conditional upon the fact that the implicit nature remains intact. Multiple studies have shown that participants *can* fake their way to lower scores on CRTs if the researchers inform them how the test *truly* works beforehand (Bowler et al., 2013; LeBreton et al., 2007; Lee, 2014; Rasmussen, 2016; Wiita et al., 2017). Although it may seem that this is a significant concern surrounding further use of CRT-WP, this can be addressed with future research. Recently, Wiita et al. added a faking detection scale to the CRT-A. This faking detection scale embedded within the CRT-A uses the same item format, but alters the response options so that these new items on the measure *only have a correct answer that is aggressive*. The researchers found that respondents who were unaware that aggression was actually being measured tended to select the aggressive options to these specific items because now it is the *only* "correct" option from a conditional reasoning

standpoint. However, respondents who are privy to the secret of the CRT and choose to fake to lower their scores end up getting caught because they consistently avoid the aggressive options on these specific items.

Wiita et al. (2017) added 11 of these “honeypot” or trap items to the CRT-A, so that there is a clear indicator of faking based on how many of these honeypots are selected. Wiita et al. gave participants the new CRT-A with embedded faking items, informed them of how the test truly works, and then told them to fake to look good because it would increase their likelihood of being hired. They found that, consistent with previous research, participants were able to significantly lower their aggression scores since they were informed that the test truly measure aggression. However, these fakers got caught as they selected 7/11 honeypots on average. In comparison, an average of 1/11 honeypots were selected in other samples that were not informed of the CRT-A’s true purpose as they continued to respond based on logic and conditional reasoning. In a follow-up study, participants were given full information and they were asked to try and pick out which items were the ones from the faking detection scale. The participants could not differentiate between the normal items and the trap items any better than chance levels (Wiita et al., 2017). It is important to acknowledge that this is a true *faking* detection system, and not a measure of social desirability or impression management. This can be said because participants are actively recognizing the “correct” logical answers are *aggressive*, and are then choosing other options based on their desire to appear *non-aggressive* specifically, which is different from simply going through the whole test wanting to present oneself positively. Future research would follow this same procedure to develop a faking-detection scale for the CRT-WP so that the measure is both proactively faking-preventative and reactively faking-detecting, so even the most skilled fakers are unable to get past this assessment.

Conclusion

The development of the CRT-WP would help fill multiple gaps in the literature, and a demonstrated gap in the employment selection practice. Most importantly, the CRT-WP may allow organizations to avoid hiring or promoting corporate psychopaths to positions of power. There is more research which needs to be done, but the current studies provide support for an initial list of items that retains a consistent structure across samples. Additionally, many of the relationships between the CRT-WP and other measures included in this research seem to support the notion that what the CRT-WP covers has a considerable degree of overlap with what existing measures cover. As it stands, there is only partial evidence for the CRT-WP being able to predict important outcomes.

In theory, the development and further validation of the CRT-WP will improve organizational performance overall, and the individual performance of employees who would have been otherwise tormented by corporate psychopaths. Depending on how successful the CRT-WP would be, it could be used as a basis for other implicit measures of psychopathy in different contexts, or a screening tool could be designed to prevent corporate psychopaths from even making it to selection decisions.

References

- Anderson, J. L., Sellbom, M., Wygant, D. B., & Edens, J. F. (2013). Examining the necessity for and utility of the Psychopathic Personality Inventory—Revised (PPI–R) validity scales. *Law and Human Behavior, 37*, 312-320. doi: 10.1037/lhb0000018
- Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340-345. doi: 10.1080/00223890902935878
- Babiak, P. (1995). When psychopaths go to work: A case study of an industrial psychopath. *Applied Psychology: An International Review, 44*, 171-188. doi: 10.1111/j.1464-0597.1995.tb01073.x
- Babiak, P. (2007). From Darkness Into the Light: Psychopathy in Industrial and Organizational Psychology. In Hervé, H., & Yuille, J. C. (Eds.), *The psychopath: Theory, research, and practice* (pp. 411-428). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Babiak, P. (2010). Psychopathic manipulation at work. In Gacono, C. B. (Ed.), *The clinical and forensic assessment of psychopathy: A practitioner's guide* (pp. 287-311). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Babiak, P., & Hare, R. D. (2012). The B-Scan 360 Manual. Manuscript in preparation.
- Babiak, P., Neumann, C. S., & Hare, R. D. (2010). Corporate psychopathy: Talking the walk. *Behavioural Sciences & the Law, 28*, 174-193. doi: 10.1102/bsl.925
- Berry, C. M., Sackett, P. R., & Tobares, V. (2010). A meta-analysis of conditional reasoning tests of aggression. *Personnel Psychology, 63*, 361-384. doi: 10.1111/j.1744-6570.2010.01173.x

- Bing, M. N., Stewart, S. M., Davison, H. K., Green, P. D., McIntyre, M. D., & James, L. R. (2007). An integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology, 92*, 722-744. doi: 10.1037/0021-9010.92.3.722
- Blickle, G., & Schütte, N. (2017). Trait psychopathy, task performance, and counterproductive work behavior directed toward the organization. *Personality and Individual Differences, 109*, 225-231. doi: 10.1016/j.paid.2017.01.006
- Boddy, C. R. (2011). Corporate psychopaths, bullying and unfair supervision in the workplace. *Journal of Business Ethics, 100*, 367-379. doi: 10.1007/s10551-010-0689-5
- Boddy, C. R., Ladyshevsky, R. K., & Galvin, P. (2010). The influence of corporate psychopaths on corporate social responsibility and organizational commitment to employees. *Journal of Business Ethics, 97*, 1-19. doi: 10.1007/s10551-010-0492-3
- Boddy, C., & Taplin, R. (2017). A note on workplace psychopathic bullying - Measuring its frequency and severity. *Aggression and Violent Behavior, 34*, 117-119. doi: 10.1016/j.avb.2017.02.001
- Boduszek, D., & Debowska, A. (2016). Critical evaluation of psychopathy measurement (PCL-R and SRP-III/SF) and recommendations for future research. *Journal of Criminal Justice, 44*, 1-12. doi: 10.1016/j.jcrimjus.2015.11.004
- Boduszek, D., Debowska, A., Dhingra, K., & DeLisi, M. (2016). Introduction and validation of Psychopathic Personality Traits Scale (PPTS) in a large prison sample. *Journal of Criminal Justice, 46*, 9-17. doi: 10.1016/j.jcrimjus.2016.02.004

- Book, A. S., Holden, R. R., Starzyk, K. B., Wasylkiw, L., & Edwards, M. J. (2006). Psychopathic traits and experimentally induced deception in self-report assessment. *Personality and Individual Differences, 41*, 601-608. doi: 10.1016/j.paid.2006.02.011
- Bowler, J. L., & Bowler, M. C. (2014). Evaluating the fakability of a conditional reasoning test of addiction proneness. *International Journal of Psychology, 49*, 415-419. doi: 10.1002/ijop.12030
- Bowler, J. L., Bowler, M. C., & Cope, J. G. (2013). Measurement issues associated with conditional reasoning tests: An examination of faking. *Personality and Individual Differences, 55*, 459-464. doi: 10.1016/j.paid.2013.04.011
- Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance, 24*, 358-372. doi: 10.1080/08959285.2011.597473
- Caponecchia, C., Sun, A. Y. Z., & Wyatt, A. (2012). 'Psychopaths' at work? Implications of lay persons' use of labels and behavioural criteria for psychopathy. *Journal of Business Ethics, 107*, 399-407. doi: 10.1007/s10551-011-1049-9
- Catano, V. M., Wiesner, W. H., & Hackett, R. D. (2016). *Recruitment and selection in Canada*, Nelson Edu. Ltd., USA.
- Christian, E., & Sellbom, M. (2016). Development and validation of an expanded version of the three-factor Levenson Self-Report Psychopathy Scale. *Journal of Personality Assessment, 98*, 155-168. doi: 10.1080/08959285.2011.597473
- Clark, T. D. (2017). Creation of a conditional reasoning task for hostile sexism. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 78*.

- Cleckley, H. (1941). *The mask of sanity; an attempt to reinterpret the so-called psychopathic personality*. Oxford, England: Mosby.
- Cohen, A. (2015). Are they among us? A conceptual framework of the relationship between the dark triad personality and counterproductive work behaviours (CWBs). *Human Resource Management Review*, 26, 69-85. doi: 10.1016/j.hrmmr.2015.07.003
- Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6, 147-168. doi: 10.1177/1094428103251541
- Cooke, D. J., Hart, S. D., Logan, C., & Michie, C. (2012). Explicating the construct of psychopathy: Development and validation of a conceptual model, the Comprehensive Assessment of Psychopathic Personality (CAPP). *International Journal of Forensic Mental Health*, 11, 242-252. doi: 10.1080/14999013.2012.746759
- Coyne, S. M., & Thomas, T. J. (2008). Psychopathy, aggression, and cheating behavior: A test of the Cheater–Hawk hypothesis. *Personality and Individual Differences*, 44, 1105-1115. doi: 10.1016/j.paid.2007.11.002
- DeLisi, M., Angton, A., Vaughn, M. G., Trulson, C. R., Caudill, J. W., & Beaver, K. M. (2014). Not my fault: Blame externalization is the psychopathic feature most associated with pathological delinquency among confined delinquents. *International Journal of Offender Therapy and Comparative Criminology*, 58, 1415-1430. doi: 10.1177/0306624X13496543
- Derefinko, K. J., & Lynam, D. R. (2006). Convergence and divergence among self-report psychopathy measures: A personality-based approach. *Journal of Personality Disorders*, 20, 261-280. doi: 10.1521/pedi.2006.20.3.261

- DeSimone, J. A., & James, L. R. (2015). An item analysis of the Conditional Reasoning Test of Aggression. *Journal of Applied Psychology, 100*, 1872-1886. doi: 10.1037/apl0000026
- Drislane, L. E., Patrick, C. J., & Arsal, G. (2014). Clarifying the content coverage of differing psychopathy inventories through reference to the Triarchic Psychopathy Measure. *Psychological Assessment, 26*, 350-362. doi: 10.1037/a0035152
- Edens, J. F. (2004). Effect of response distortion on the assessment of divergent facets of psychopathy. *Assessment, 11*, 109-112. doi: 10.1177/1073191103261474
- Edens, J. F., Buffington, J. K., Tomicic, T. L., & Riley, B. D. (2001). Effects of positive impression management on the Psychopathic Personality Inventory. *Law and Human Behavior, 25*, 235-256. doi: 10.1023/A:1010793810896
- Edens, J. F., Poythress, N. G., & Lilienfeld, S. O. (1999). Identifying inmates at risk for disciplinary infractions: A comparison of two measures of psychopathy. *Behavioral Sciences & the Law, 17*, 435-443. doi: 10.1002/(SICI)1099-0798(199910/12)17:4<435::AID-BSL356>3.0.CO;2-Z
- Evans, L., & Tully, R. J. (2016). The Triarchic Psychopathy Measure (TriPM): Alternative to the PCL-R?. *Aggression and Violent Behavior, 27*, 79-86. doi: 10.1016/j.avb.2016.03.004
- Falkenbach, D., Poythress, N., Falki, M., & Manchak, S. (2007). Reliability and validity of two self-report measures of psychopathy. *Assessment, 14*, 341-350. doi: 10.1177/1073191107305612
- Fine, S., & Gottlieb-Litvin, Y. (2013). Justifying counterproductive work behaviors and an integrity-based Conditional Reasoning Test: Back to the drawing board?. *International Journal of Selection and Assessment, 21*, 328-333. doi: 10.1111/ijsa.12042

- Fisher, P. A., Robie, C., Christiansen, N. D., & Komar, S. (2018). The impact of psychopathy and warnings on faking behavior: A multisaturation perspective. *Personality and Individual Differences, 127*, 39-43. doi: 10.1016/j.paid.2018.01.033
- Forouzan, E., & Cooke, D. J. (2005). Figuring out la femme fatale: Conceptual and assessment issues concerning psychopathy in females. *Behavioral Sciences & the Law, 23*, 765-778. doi: 10.1002/bsl.669
- Galić, Z. (2016). Conditional Reasoning Test for Aggression: Further evidence about incremental validity. *International Journal of Selection and Assessment, 24*, 24-33. doi: 10.1002/bsl.669
- Galić, Z., Ružojčić, M., Trojak, N., & Zeljko, L. (2018, June). Implicit power motive assessment: Fakeability and predictive validity of the Conditional Reasoning Test for Power. In ENESER 2018: EAWOP Small Group Meeting on Recent Developments in Recruitment and Selection.
- Galić, Z., Scherer, K. T., & LeBreton, J. M. (2014). Validity evidence for a Croatian version of the Conditional Reasoning Test for Aggression. *International Journal of Selection and Assessment, 22*, 343-354. doi: 10.1111/ijsa.12082
- Gao, Y., & Raine, A. (2010). Successful and unsuccessful psychopaths: A neurobiological model. *Behavioral Sciences & the Law, 28*, 194-210. doi: 10.1111/ijsa.12082
- Gatner, D. T., Blanchard, A. J. E., Douglas, K. S., Lilienfeld, S. O., & Edens, J. F. (2018). Psychopathy in a multiethnic world: Investigating multiple measures of psychopathy in Hispanic, African American, and Caucasian offenders. *Assessment, 25*, 206-221. doi: 10.1177/1073191116639374
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe, 7*, 7-28.

Retrieved from: <http://admin.umd.edu.pk/Media/Site/STD/FileManager/OsamaArticle/26august2015/A%20broad-bandwidth%20inventory.pdf>

Gordts, S., Uzieblo, K., Neumann, C., Van den Bussche, E., & Rossi, G. (2017). Validity of the Self-Report Psychopathy Scales (SRP-III full and short versions) in a community sample. *Assessment, 24*, 308-325. doi: 10.1177/1073191115606205

Gustafson, S. B. (2000). *Personality and organizational destructiveness: Fact, fiction, and fable*. In Bergman, L. R., Cairns, R. B., Nilsson, L-G., & Nystedt, L. (Eds.), *Developmental science and the holistic approach*, 299-313.

Grieve, R. (2012). The role of personality, psychopathy, and previous experience with assessment in intentions to fake in psychological testing. *Current Psychology, 31*, 414-422. doi: 10.1007/s12144-012-9158-x

Hall, J. R., Benning, S. D., & Patrick, C. J. (2004). Criterion-related validity of the three-factor model of psychopathy: Personality, behavior, and adaptive functioning. *Assessment, 11*, 4-16. doi: 10.1177/1073191103261466

Hall, J. R., Drislane, L. E., Patrick, C. J., Morano, M., Lilienfeld, S. O., & Poythress, N. G. (2014). Development and validation of triarchic construct scales from the psychopathic personality inventory. *Psychological Assessment, 26*, 447-461. doi: 10.1037/a0035665

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences, 1*, 111-119. doi: 10.1016/0191-8869(80)90028-8

Hare, R. D. (1985). Comparison of procedures for the assessment of psychopathy. *Journal of Consulting and Clinical Psychology, 53*, 7-16. doi: 10.1037/0022-006X.53.1.7

- Harris, P. B., Boccaccini, M. T., & Murrie, D. C. (2015). Rater differences in psychopathy measure scoring and predictive validity. *Law and Human Behavior, 39*, 321-331. doi: 10.1037/lhb0000115
- Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *Sage Open, 5*, 1-9. doi: 10.1177/2158244015621113
- Hart, S. D., Cox, D. N., & Hare, R. D. (1995). Hare psychopathy checklist: Screening version (PCL: SV). Toronto, Canada: Multi-Heath Systems.
- Hauser, D., Paolacci, G., & Chandler, J. J. (2018). Common Concerns with Mturk as a Participant Pool: Evidence and Solutions. *PsyArXiv*. September, 1, 2018.
- Henley, A. G. (2002). Psychopathy and career interest in a noncriminal population. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*.
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*, 104-121. doi: 10.1177/109442819800100106
- Howe, J., Falkenbach, D., & Massey, C. (2014). The relationship among psychopathy, emotional intelligence, and professional success in finance. *International Journal of Forensic Mental Health, 13*, 337-347. doi: 10.1080/14999013.2014.951103
- Huang, J., Curran, P., Keeney, J., Paposki, E., & DeShon, R. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*, 99-114. doi: 10.1007/s10869-011-9231-8

Ireland, J. L., Ireland, C. A., Lewis, M., Jones, C., & Keeley, S. (2016). Initial development of an implicit and explicit approach to assess psychopathy. *International Journal of Law and Psychiatry, 46*, 117-128. doi: 10.1016/j.ijlp.2016.02.008

James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational research methods, 1*, 131-163. doi: 10.1016/j.ijlp.2016.02.008

James, L. R., & LeBreton, J. M. (2010). Assessing aggression using conditional reasoning. *Current Directions in Psychological Science, 19*, 30-35. doi: 10.1177/0963721409359279

James, L. R., & McIntyre, M. D. (2000). Conditional reasoning test of aggression: Test manual. Innovative Assessment Technology.

James, L. R., McIntyre, M. D., Glisson, C. A., Bowler, J. L., & Mitchell, T. R. (2004). The conditional reasoning measurement system for aggression: An overview. *Human Performance, 17*, 271-295. doi: 10.1207/s15327043hup1703

Kelley, S. E., Edens, J. F., Donnellan, M. B., Mowle, E. N., & Sörman, K. (2018). Self-and informant perceptions of psychopathic traits in relation to the triarchic model. *Journal of personality, 86*, 738-751. doi: 10.1111/jopy.12354

Kelley, S. E., Edens, J. F., Donnellan, M. B., Ruchensky, J. R., Witt, E. A., & McDermott, B. E. (2016). Development and validation of an inconsistent responding scale for an abbreviated version of the Psychopathic Personality Inventory—Revised. *Personality and Individual Differences, 91*, 58-62. doi: 10.1016/j.paid.2015.11.033

Kelsey, K. R. (2016). The vulnerability of self-report measures of psychopathy to positive impression management: A simulation study with inmates. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 77*.

- Kelsey, K. R., Rogers, R., & Robinson, E. V. (2015). Self-report measures of psychopathy: What is their role in forensic assessments?. *Journal of Psychopathology and Behavioral Assessment*, 37, 380-391. doi: 10.1007/s10862-014-9475-5
- Kreis, M. K., Cooke, D. J., Michie, C., Hoff, H. A., & Logan, C. (2012). The Comprehensive Assessment of Psychopathic Personality (CAPP): Content validation using prototypical analysis. *Journal of Personality Disorders*, 26, 402-413. doi: 10.1521/pedi.2012.26.3.402
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2, 151–160. doi: 10.1007/BF02288391
- Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 8, 142-164. doi:10.1017/iop.2015.13
- LeBreton, J. M., Barksdale, C. D., Robin, J., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *Journal of Applied Psychology*, 92, 1-16. doi: 10.1037/0021-9010.92.1.1
- Lee, H. J. (2014). Equivalence and faking issues of the aggression questionnaire and the conditional reasoning test for aggression in Korean and American samples. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 74.
- Lee, K., & Ashton, M. C. (2005). Psychopathy, machiavellianism, and narcissism in the five-factor model and the HEXACO model of personality structure. *Personality and Individual Differences*, 38, 1571-1582. doi: 10.1016/j.paid.2004.09.016

- Lee, K., & Ashton, M. C. (2014). The dark triad, the big five, and the HEXACO model. *Personality and Individual Differences, 67*, 2-5. doi: 10.1016/j.paid.2014.01.048
- Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. *Journal of Applied Psychology, 92*, 1638-1656. doi: 10.1037/0021-9010.92.6.1638
- Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*, 151-158. doi: 10.1037/2F0022-3514.68.1.151
- Lewis, M., Ireland, J. L., Abbott, J., & Ireland, C. A. (2017). Initial development of the Psychopathic Processing and Personality Assessment (PAPA) across populations. *International Journal of Law and Psychiatry, 54*, 118-132. doi: 10.1016/j.ijlp.2017.06.006
- Lilienfeld, S. O. (1990). Development and preliminary validation of a self-report measure of psychopathic personality (Vol. 1). University of Minnesota.
- Lilienfeld, S. O. (1994). Conceptual problems in the assessment of psychopathy. *Clinical Psychology Review, 14*, 17-38. doi: 10.1016/0272-7358(94)90046-9
- Lilienfeld, S. O. (1998). Methodological advances and developments in the assessment of psychopathy. *Behaviour Research and Therapy, 36*, 99-125. doi: 10.1016/S0005-7967(97)10021-3
- Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal population. *Journal of Personality Assessment, 66*, 488-524. doi: 10.1207/s15327752jpa6603_3

- Lilienfeld, S. O., Smith, S. F., Sauvigné, K. C., Patrick, C. J., Drislane, L. E., Litzman, R. D., & Krueger, R. F. (2016). Is boldness relevant to psychopathic personality? Meta-analytic relations with non-Psychopathy Checklist-based measures of psychopathy. *Psychological Assessment, 28*, 1172-1185. doi: 10.1037/pas0000244
- Lilienfeld, S. O., Waldman, I. D., Landfield, K., Watts, A. L., Rubenzer, S., & Faschingbauer, T. R. (2012). Fearless dominance and the U.S. presidency: Implications of psychopathic personality traits for successful and unsuccessful political leadership. *Journal of Personality and Social Psychology, 103*, 489-505. doi: 10.1037/a0029392
- Lilienfeld, S. O., Watts, A. L., & Smith, S. F. (2015). Successful psychopathy: A scientific status report. *Current Directions in Psychological Science, 24*, 298-303. doi: 10.1177/0963721415580297
- Lilienfeld, S. O., Widows, M. R., & Staff, P. A. R. (2005). Psychopathic Personality Inventory™-Revised. *Social Influence (SOI), 61*, 97.
- Litman, L., Robinson, J., & Abberbock, (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*, 433. doi: 10.3758/s13428-016-0727-z
- Lynam, D. R., Gaughan, E. T., Miller, J. D., Miller, D. J., Mullins-Sweatt, S., & Widiger, T. A. (2011). Assessing the basic traits associated with psychopathy: Development and validation of the Elemental Psychopathy Assessment. *Psychological Assessment, 23*, 108-124. doi: 10.1037/a0021146

Lynam, D. R., Sherman, E. D., Samuel, D., Miller, J. D., Few, L. R., & Widiger, T. A. (2013).

Development of a short form of the Elemental Psychopathy Assessment. *Assessment, 20*, 659-669. doi: 10.1177/1073191113502072

Mahmut, M. K., Menictas, C., Stevenson, R. J., & Homewood, J. (2011). Validating the factor

structure of the Self-Report Psychopathy Scale in a community sample. *Psychological Assessment, 23*, 670-678. doi: 10.1037/a0023090

Marcus, D. K., Church, A. S., O'Connell, D., & Lilienfeld, S. O. (2018). Identifying careless

responding with the Psychopathic Personality Inventory–Revised validity scales. *Assessment, 25*, 31-39. doi: 10.1177/1073191116641507

Mathieu, C., & Babiak, P. (2015). Tell me who you are, I'll tell you how you lead: Beyond the Full-

Range Leadership Model, the role of corporate psychopathy on employee attitudes. *Personality and Individual Differences, 87*, 8-12. doi: 10.1016/j.paid.2015.07.016

Mathieu, C., & Babiak, P. (2016a). Corporate psychopathy and abusive supervision: Their influence

on employees' job satisfaction and turnover intentions. *Personality and Individual Differences, 91*, 102-106. doi: 10.1016/j.paid.2015.12.002

Mathieu, C., & Babiak, P. (2016b). Validating the B-Scan Self: A self-report measure of

psychopathy in the workplace. *International Journal of Selection and Assessment, 24*, 272-284. doi: 10.1111/ijsa.12146

Mathieu, C., Hare, R. D., Jones, D. N., Babiak, P., & Neumann, C. S. (2013). Factor structure of the

B-Scan 360: A measure of corporate psychopathy. *Psychological Assessment, 25*, 288-293. doi: 10.1037/2Fa0029262

- Mathieu, C., Neumann, C. S., Hare, R. D., & Babiak, P. (2014). A dark side of leadership: Corporate psychopathy and its influence on employee well-being and job satisfaction. *Personality and Individual Differences, 59*, 83-88. doi: 10.1016/j.paid.2013.11.010
- McCabe, D. L., Butterfield, K. D., & Trevino, L. K. (2012). *Cheating in college: Why students do it and what educators can do about it*. JHU Press.
- McCabe, D. L., & Trevino, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. *The Journal of Higher Education, 64*, 522-538. doi: 10.1080/00221546.1993.11778446
- McCoy, W. K., & Edens, J. F. (2006). Do black and white youths differ in levels of psychopathic traits? A meta-analysis of the psychopathy checklist measures. *Journal of Consulting and Clinical Psychology, 74*, 386-392. doi: 10.1037/0022-006X.74.2.386
- McHoskey, J. W., Worzel, W., & Szyarto, C. (1998). Machiavellianism and psychopathy. *Journal of Personality and Social Psychology, 74*, 192-210. doi: 10.1037/0022-3514.74.1.192
- Miller, J. D., Gaughan, E. T., Maples, J., Gentile, B., Lynam, D. R., & Widiger, T. A. (2011). Examining the construct validity of the Elemental Psychopathy Assessment. *Assessment, 18*, 106-114. doi: 10.1177/1073191110393139
- Miller, J. D., Hyatt, C. S., Rausher, S., Maples, J. L., & Zeichner, A. (2014). A test of the construct validity of the Elemental Psychopathy Assessment scores in a community sample of adults. *Psychological Assessment, 26*, 555-562. doi: 10.1037/a0035952
- Muthén, L. K., & Muthén, B. (2017). *Mplus User's Guide: Statistical Analysis with Latent Variables, User's Guide*. Muthén & Muthén.

- Nathanson, C., Paulhus, D. L., & Williams, K. M. (2006). Predictors of a behavioral measure of scholastic cheating: Personality and competence but not demographics. *Contemporary Educational Psychology, 31*, 97-122. doi: 10.1016/j.cedpsych.2005.03.001
- Neal, T. M., & Sellbom, M. (2012). Examining the factor structure of the hare self-report psychopathy scale. *Journal of Personality Assessment, 94*, 244-253. doi: 10.1080/00223891.2011.648294
- Olver, M. E., & Wong, S. C. (2015). Short-and long-term recidivism prediction of the PCL-R and the effects of age: A 24-year follow-up. *Personality Disorders: Theory, Research, and Treatment, 6*, 97-105. doi: 10.1037/2Fper0000095
- Pardue, A. D., Robinson, M. B., & Arrigo, B. A. (2013). Psychopathy and corporate crime: A preliminary examination, part 2. *Journal of Forensic Psychology Practice, 13*, 145-169. doi: 10.1080/15228932.2013.765746
- Patrick, C. J. (2010). Triarchic Psychopathy Measure (TriPM). PhenX Toolkit online assessment catalog, 813-814.
- Patton, T. W. (2000). Measuring personal reliability via conditional reasoning: Identifying people who will work reliably. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 60*.
- Paulhus, D. L. (1988). Balanced inventory of desirable responding (BIDR). Acceptance and Commitment Therapy. *Measures Package*, 41-43. Retrieved from: <http://www.mindfulness-extended.nl/content3/wp-content/uploads/2013/07/Measures-Package.pdf#page=41>
- Paulhus, D. L., Hemphill, J. F., & Hare, R. D. (2012). Self-report psychopathy scale (SRP-III). Toronto: Multi-Health Systems.

- Paulhus, D., Neumann, C. S., & Hare, R. D. (2015). The SRP-4. Toronto, Canada: Multi-Health Systems.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903. doi: 10.1037/0021-9010.88.5.879
- Pulfrey, C., & Butera, F. (2013). Why Neoliberal values of self-enhancement lead to cheating in higher education: A motivational account. *Psychological Science, 24*, 2153-2162. doi: 10.1177/0956797613487221
- Raine, A. (1986). Psychopathy, schizoid personality and borderline/schizotypal personality disorders. *Personality and Individual Differences, 7*, 493-501. doi: 10.1016/0191-8869(86)90128-5
- Rasmussen, J. L. (2016). Implicit measurement of extraversion and agreeableness using conditional reasoning tests: The impact of faking. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 77*.
- Ray, J. V., & Jones, S. (2011). Self-reported psychopathic traits and their relation to intentions to engage in environmental offending. *International Journal of Offender Therapy, 55*, 370-391. doi: 10.1177/0306624X10361582
- Robinson, E. V., & Rogers, R. (2015). Empathy faking in psychopathic offenders: The vulnerability of empathy measures. *Journal of Psychopathology and Behavioral Assessment, 37*, 545-552. doi: 10.1007/s10862-015-9479-9
- Rogers, R., Vitacco, M. J., Jackson, R. L., Martin, M., Collins, M., & Sewell, K. W. (2002). Faking psychopathy? An examination of response styles with antisocial youth. *Journal of Personality Assessment, 78*, 31-46. doi: 10.1207/S15327752JPA7801_03

- Roulin, N., & Krings, F. (2016). When winning is everything: the relationship between competitive worldviews and job applicant faking. *Applied Psychology, 65*, 643-670. doi: 10.1111/apps.12072
- Salekin, R. T., Chen, D. R., Sellbom, M., Lester, W. S., & MacDougall, E. (2014). Examining the factor structure and convergent and discriminant validity of the Levenson Self-Report Psychopathy Scale: Is the two-factor model the best fitting model?. *Personality Disorders: Theory, Research, and Treatment, 5*, 289-304. doi: 10.1037/per0000073
- Scherer, K. T., Baysinger, M., Zolynsky, D., & LeBreton, J. M. (2013). Predicting counterproductive work behaviors with sub-clinical psychopathy: Beyond the Five Factor Model of personality. *Personality and Individual Differences, 55*, 300-305. doi: 10.1016/j.paid.2013.03.007
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274. doi: 0033-2909/98/S3.00
- Schoen, J. L., Bowler, J. L., & Schilpzand, M. C. (2018). Conditional reasoning test for creative personality: Rationale, theoretical development, and validation. *Journal of Management, 44*, 1651-1677. doi: 10.1177/0149206315618012
- Schroeder, M. L., Schroeder, K. G., & Hare, R. D. (1983). Generalizability of a checklist for assessment of psychopathy. *Journal of Consulting and Clinical Psychology, 51*, 511-516. doi: 10.1037/0022-006X.51.4.511
- Sellbom, M., Wygant, D. B., & Drislane, L. E. (2015). Elucidating the construct validity of the psychopathic personality inventory triarchic scales. *Journal of Personality Assessment, 97*, 374-381. doi: 10.1080/00223891.2014.962654

- Shou, Y., Sellbom, M., Xu, J., Chen, T., & Sui, A. (2017). Elaborating on the construct validity of Triarchic Psychopathy Measure in Chinese clinical and nonclinical samples. *Psychological Assessment, 29*, 1071-1081. doi: 10.1037/pas0000398
- Sica, C., Drislane, L., Caudek, C., Angrilli, A., Bottesi, G., Cerea, S., & Ghisi, M. (2015). A test of the construct validity of the Triarchic Psychopathy Measure in an Italian community sample. *Personality and Individual Differences, 82*, 163-168. doi: 10.1016/j.paid.2015.03.015
- Skeem, J. L., & Cooke, D. J. (2010). One measure does not a construct make: Directions toward reinvigorating psychopathy research—reply to Hare and Neumann (2010). *Psychological Assessment, 22*, 455-459. doi: 10.1037/a0014862
- Smith, S. F., & Lilienfeld, S. O. (2013). Psychopathy in the workplace: The knowns and unknowns. *Aggression and Violent Behavior, 18*, 204-218. doi: 10.1016/j.avb.2012.11.007
- Smith, S. F., Watts, A. L., & Lilienfeld, S. O. (2014). On the trail of the elusive successful psychopath. *The Psychologist, 27*, 506-510. Retrieved from: <http://sungraffix.net/02-2ndSite/school/Psychopaths/On%20the%20trail%20of%20the%20elusive%20successful%20psychopath.pdf>
- Somma, A., Fossati, A., Patrick, C., Maffei, C., & Borroni, S. (2014). The three-factor structure of the Levenson Self-Report Psychopathy Scale: Fool's gold or true gold? A study in a sample of Italian adult non-clinical participants. *Personality and Mental Health, 8*, 337-347. doi: 10.1002/pmh.1267
- Spain, S. M., Harms, P., & LeBreton, J. M. (2014). The dark side of personality at work. *Journal of Organizational Behavior, 35*, 41-60. doi: 10.1002/job.1894

- Spector, P. E., Fox, S., Penney, L. M., Bruursema, K., Goh, A., & Kessler, S. (2006). The dimensionality of counterproductivity: Are all counterproductive behaviours created equal? *Journal of Vocational Behaviour, 68*, 446-460. doi: 10.1016/j.jvb.2005.10.005
- Steinert, S. W., Lishner, D. A., Vitacco, M. J., & Hong, P. Y. (2017). Conceptualizing successful psychopathy: An elaboration of the moderated-expression model. *Aggression and Violent Behavior, 36*, 44-51. doi: 10.1016/j.avb.2017.07.005
- Stevens, G. W., Deuling, J. K., & Armenakis, A. A. (2012). Successful psychopaths: Are they unethical decision-makers and why?. *Journal of Business Ethics, 105*, 139-149. doi: 10.1007/s10551-011-0963-1
- Tapscott, J. L., Vernon, P. A., & Veselka, L. (2012). A comparison of the construct validity of two alternative approaches to the assessment of psychopathy in the community. *Journal of Personality Assessment, 94*, 541-554. doi: 10.1080/00223891.2012.668595
- ten Brinke, L., Kish, A., & Keltner, D. (2018). Hedge fund managers with psychopathic tendencies make for worse investors. *Personality and Social Psychology Bulletin, 44*, 214-223. doi: 10.1177/0146167217733080
- Tsang, S., Salekin, R. T., Coffey, C. A., & Cox, J. (2017). A comparison of self-report measures of psychopathy among nonforensic samples using item response theory analyses. *Psychological Assessment, 30*, 311-327. doi: 10.1037/pas0000481
- van Dongen, J. D., Drislane, L. E., Nijman, H., Soe-Agnie, S. E., & van Marle, H. J. (2017). Further evidence for reliability and validity of the Triarchic Psychopathy Measure in a forensic sample and a community sample. *Journal of Psychopathology and Behavioral Assessment, 39*, 58-66. doi: 10.1007/s10862-016-9567-5

- Vergauwe, J., Wille, B., Hofmans, J., & De Fruyt, F. (2019). Successful psychopathic leaders: Evidence for a differential configuration model. Presented at the 19th Congress of the EAWOP.
- Verschuere, B., Uzieblo, K., De Schryver, M., Douma, H., Onraedt, T., & Crombez, G. (2014). The inverse relation between psychopathy and faking good: Not response bias, but true variance in psychopathic personality. *The Journal of Forensic Psychiatry & Psychology*, *25*, 705-713. doi: 10.1080/14789949.2014.952767
- Walters, G. D., Brinkley, C. A., Magaletta, P. R., & Diamond, P. M. (2008). Taxometric analysis of the Levenson self-report psychopathy scale. *Journal of Personality Assessment*, *90*, 491-498. doi: 10.1080/00223890802248828
- Wiita, N. E., Meyer, R. D., Kelly, E. D., & Collins, B. J. (2017). Not aggressive or just faking it? Examining faking and faking detection on the Conditional Reasoning Test of Aggression. *Organizational Research Methods*, 1-28. doi: 10.1177/1094428117703685.
- Williams, K. M., Nathanson, C., & Paulhus, D. L. (2010). Identifying and profiling scholastic cheaters: Their personality, cognitive ability, and motivation. *Journal of Experimental Psychology: Applied*, *16*, 293-307. doi: 10.1037/a0020773
- Williams, K. M., & Paulhus, D. L. (2004). Factor structure of the Self-Report Psychopathy scale (SRP-II) in non-forensic samples. *Personality and Individual Differences*, *37*, 765-778. doi: /10.1016/j.paid.2003.11.004
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*, 806-838. doi: 10.1177/0011000006288127

Wright, M. A. (2012). Investigating the validity of the Conditional Reasoning Test for Leadership.

Dissertation Abstracts International: Section B: The Sciences and Engineering, 73.

Wu, J., & LeBreton, J. M. (2011). Reconsidering the dispositional basis of counterproductive work

behavior: The role of aberrant personality. *Personnel Psychology*, 64, 593-626. doi:

10.1111/j.1744-6570.2011.01220.x

Appendix A

Expanded Details of Psychopathy Measures

Psychopathy Checklist-Revised and Self-Report Psychopathy scale. Hare (1980) thought that none of the measures to date truly covered psychopathy as Cleckley had conceptualized it many years before. Thus, a “Psychopathy Checklist” (PCL) was created based on Cleckley’s original 16 criteria of a psychopath. Hare and an assistant created a list of over 100 psychopathic behaviours and he and his colleague rated a criminal sample on all of these behaviours, from zero to two. A factor analysis of this data found five factors and reduced the list to 22 essential criteria. The final Psychopathy Checklist kept these 22 factors with the same rating scale (total scores could be from 0 to 44). There was immediate support for the PCL’s factor structure and clinical usefulness (Raine, 1986; Schroeder, Schroeder, & Hare, 1983). Acknowledging the popularity of self-report measures, Hare (1985) decided to translate the exact PCL factors and items into the *Self-Report Psychopathy scale* (SRP). The SRP gave the most similar scores to actual clinician assessments than any of the other self-report measures that existed at the time. It was later found that there is a clear two-factor structure in the PCL. The first factor contains items that describe psychopaths as being selfish, remorseless, manipulative, and having certain attitudes and behavioural tendencies. The second factor consists of items that detail the psychopath’s history, such as chronic instability, social deviance, and others for which there has to be some objective evidence (i.e., crime). Further, each factor consists of two facets, giving the PCL-R four facets in total. The PCL later became the *Psychopathy Checklist – Revised* (PCL-R) which dropped two items, leaving a total of 20 and a maximum score of 40. The recommended cut-off score for labelling someone a psychopath is 30/40. Additionally, the SRP has been revised multiple times over the years with the current adaptation being called the SRP-4 which has a long and short version (Paulhus et al., 2015).

Since their origin, the PCL-R and SRP have been the most frequently used measures of psychopathy (Evans & Tully, 2016). However, they are not without many criticisms. The first critical issue is that although the PCL-R and SRP are based on Cleckley's conceptualization of psychopathy, there are some inconsistencies in how this is measured (Lilienfeld & Andrews, 1996; Skeem & Cooke, 2010). For example, Cleckley (1941) states that psychopaths are calculating and systematic, not impulsive, yet one of the PCL-R items is impulsivity. The second issue is that there are clear ways in which the PCL-R and SRP can give false positives and false negatives. Consider a successful psychopath who may score high on factor 1 but with no documented criminal/deviant history they would get a zero on factor 2 and go undetected by the PCL-R (Lilienfeld, 1994). As a response to this criticism, a version of the PCL-R was adapted for non-criminal use and was called the Psychopathy Checklist – Screening Version (PCL:SV; Hart et al., 1995). The PCL:SV takes 12 of the 20 PCL-R items (eliminating the ones about deviant history) but still uses the zero to two rating scale and requires a professional to conduct the assessment.

Although all of this information may seem outdated, issues surrounding the PCL-R based measures have been continuously published to this day. Various researchers have questioned the factor structure of the PCL-R-based measures, with some consistently demonstrating completely different factors (Boduszek & Debowska, 2016; Boduszek et al., 2016; Drislane et al., 2014; Hall et al., 2004; Lilienfeld et al., 2016; Skeem & Cooke, 2010; Williams & Paulhus, 2004). However, other researchers have supported the original factor structure (Gordts, Uzieblo, Neumann, Van den Bussche, & Rossi, 2017; Mahmut et al., 2011; Neal & Sellbom, 2012; Tsang et al., 2017). Additionally, some researchers are concerned about whom the PCL-R-based measures can accurately assess since all of them except the PCL:SV were developed using male criminal samples. Forouzan and Cooke (2005) argue that the PCL-R does not accurately capture the female psychopath, while others give evidence for possible discrimination against minorities (McCoy & Edens, 2006; Gatner et

al., 2018). Finally, Harris, Boccaccini, and Murrie (2015) have demonstrated that some PCL:SV raters consistently give high/low ratings which theoretically should not happen. It is for these reasons above why the PCL-R and SRP do not function properly in a selection context. Most importantly, the PCL-R and SRP contain items focused on criminal/antisocial history which are likely unable to capture the successful psychopath. The PCL:SV, designed for non-criminal populations, still requires hiring a professional rater which is likely beyond what most applicants and organizations are willing to do during the hiring process.

Levenson Self-Report Psychopathy scale. The *Levenson Self-Report Psychopathy scale* (LSRP; Levenson et al., 1995) was created around the same time as Hare's SRP. The LSRP is modelled after the PCL-R as well, and contains 26 items covering the two factors of "primary" and "secondary" psychopathy. Similar to the PCL-R and SRP, the "primary" factor consists of the more fundamental psychopathy traits such as manipulation and selfishness, while the "secondary" factor measures antisocial behaviours. The LSRP was also developed with mainly male criminals (Levenson et al., 1995). Shortly after publication, the LSRP was criticized as measuring anti-social personality more than it measures psychopathy, though this claim is based on a very PCL-R mindset where psychopathy requires some sort of criminal behaviour (McHoskey, Worzel, & Szyarto, 1998). As a result of this difference, it is likely that the LSRP is more applicable to measuring *successful* psychopathy than the PCL-R or the SRP (Smith & Lilienfeld, 2013). The LSRP has remained almost completely unchanged from the original version, although it is still the focus of recent research which most seems to indicate that the psychometric properties have remained valid (Falkenbach et al., 2007; Walters et al., 2008). However, other researchers have questioned the factor structure of the LSRP and whether the measure is still relevant in its original, potentially outdated, form (Salekin, Chen, Sellbom, Lester, & MacDougall, 2014; Somma, Fossati, Patrick, Maffei, & Borroni, 2014; Tsang et al., 2017). Recently, Christian and Sellbom (2016) acknowledged the criticisms of the

LSRP and attempted to expand and revise the scale. Their new version contains 36 items and a three-factor structure of egocentric, callous, and antisocial factors. There are also concerns about the LSRP being potentially discriminatory to minorities (Gatner et al., 2018).

Psychopathic Personality Inventory – Revised. The *Psychopathic Personality Inventory* (PPI; Lilienfeld, 1990) was the first self-report psychopathy measure, as it pre-dated the SRP and LSRP. Lilienfeld was the first to criticize the PCL-R's dependence on deviant/criminal behaviour as part of the assessment of psychopathy, as he argued that although the two were highly related, the former was not required by the latter. Lilienfeld also challenged the PCL-R's conceptualization and factor structure in various studies (Lilienfeld, 1990; Lilienfeld, 1994, Lilienfeld & Andrews, 1996; Lilienfeld, 1998). Since he believed that psychopaths exist in all contexts, the PPI measures the "core" psychopathy traits such as a lack of empathy, superficial charm, and others, without considering the illegal behaviours which are only inherent to the forensic version of psychopathy. Thus, the PPI was designed specifically to measure psychopathy in non-criminal contexts, though it has been used only second to the PCL-R in clinical contexts as well. The PPI later became the PPI-R (Lilienfeld & Widows, 2005) which consists of 154 items that are answered on a 4-point Likert-scale. The PPI-R has two factors, the first (fearless dominance) consists of more adaptive traits such as fearlessness and stress immunity, while the second (self-centered impulsivity) represents the dishonourable traits such as cold-heartedness and Machiavellian egocentricity. Total scores are used to determine levels of psychopathy. The PPI-R also has built-in scales to detect virtuous responding (faking good), deviant responding (faking bad), and inconsistent responding (not paying attention). There is also a short version of the PPI-R which consists of 40 items and drops the inconsistent responding scale. Comparing the PPI-R to the SRP, studies have noted that the two scales had strong convergence at the overall level, but factor correlations between the two were very weak, as expected (Derefinko & Lynam, 2006; Walters et al., 2008).

Similar to the other measures mentioned so far, the PPI-R has some experts who favour it over other measures (i.e., Tapscott et al., 2012) and some who do not (i.e., Tsang et al., 2017). When compared to the PCL-R, the PPI was more predictive of non-aggressive and verbal disciplinary reports for criminals than the PCL-R (Edens, Poythress, & Lilienfeld, 1999). Researchers have found support for the psychometric properties of the PPI-R (Edens, 2004; Falkenback et al., 2007; Tapscott et al., 2012). However, there have been others who have raised concerns about the structure of the PPI-R, its application, and its built-in detection scales (Kelley et al., 2016; Marcus et al., 2018; Tsang et al., 2017). Specifically, Hall et al. (2014) recently fit the PPI-R into a three-factor structure which was consistent across both criminal and student samples and has been validated further by Sellbom and colleagues (2015). This new three-factor structure has been generally approved of by the main researchers behind the PPI-R (Lilienfeld et al., 2016). The PPI-R has also been implicated as having potential discrimination toward minorities, Hispanics in particular (Gatner et al., 2018). Despite some of these issues, the PPI-R was designed to specifically measure the non-criminal psychopath, has been validated and improved in recent years, and contains response distortion scales. For these reasons, it was recently classified as one of the most potentially useful tools to measure corporate psychopathy (Smith & Lilienfeld, 2013).

Elemental Psychopathy Assessment. Lynam et al. (2011) advocated for measure of psychopathy that was more connected to the rest of personality research. Their answer to this was the self-report *Elemental Psychopathy Assessment* (EPA) which is based on the Five-Factor Model (FFM) of personality. The EPA consists of re-worded items for 18 of the FFM sub-factors which are most applicable to psychopathy. These 178 items are worded to reflect the negative manifestations of these sub-factors (i.e., assertiveness) and they are responded to on a Likert-scale similar to scales measuring the FFM. Lynam et al. validated the EPA in three large student samples, and one small sample of criminals. They found that the EPA had a consistent structure, was related to other self-

report measures of psychopathy, and could predict psychopathy scores beyond the FFM. Other researchers have supported the construct validity of the EPA, citing that it brings psychopathy into the nomological network of personality, and that it is a promising tool that can be used flexibly (Miller et al., 2011; Miller, Hyatt, Rausher, Maples, & Zeichner, 2014). A 72-item short form of the EPA was later developed which is supposed to be equally stable and predictive (Lynam et al., 2013). The EPA is relatively new, and it has seen little-to-no use in measuring workplace psychopathy. Considering that this measure was developed using a non-criminal sample and has only been supported so far, the EPA may prove to be one of the more promising measures of psychopathy for use in employment selection.

Triarchic Psychopathy Measure. As mentioned in previous sections, newer three-factor structures were supported consistently across the PCL-R, SRP, and PPI-R (Hall et al., 2004; Hall et al., 2014). Based on this underlying three-factor structure, the *Triarchic Psychopathy Measure* was created (TriPM; Patrick, 2010). The TriPM is a 58-item self-report inventory where each item is responded to using a 4-point Likert-scale. The three factors inherent to the Triarchic model are Boldness (associated with more of the adaptive traits such as fearlessness), Disinhibition (associated with indices of deviant behaviours that are not necessarily criminal), and Meanness (associated with the more typical traits such as Machiavellianism, cold-heartedness, etc.). The TriPM is credited with integrating the different conceptualizations of psychopathy into one measure, with good construct validity (Drislane et al., 2014). Another strength of the TriPM is that it appears to remain consistent and independent of age, education, and culture (Sica et al., 2015). Though, other researchers have found inconsistencies across cultures (Shou, Sellbom, Xu, Chen, & Sui, 2017). The TriPM appears to be the most universally favoured measure of psychopathy in the current literature (Evans & Tully, 2016; Lilienfeld, 2016; van Dongen et al., 2017). It appears that the TriPM may be one of the best at measuring corporate psychopathy in employment selection, given its widespread support and

usefulness in non-criminal samples. Despite this, the TriPM has seen relatively low use in work contexts thus far, though this may be changing (van Dongen et al., 2017).

Newer Measures. There are other new psychopathy measures that have not yet received much research. One of these is the *Comprehensive Assessment of Psychopathic Personality* (CAPP; Cooke et al., 2012), which is an assessment consisting of both personality measurement and professional judgement from a clinician. It remains largely a research tool that has seen little use, though it has received content validation (Kreis, Cooke, Michie, Hoff, & Logan, 2012).

The *Affective, Cognitive, Lifestyle assessment* (ACL; Ireland, Ireland, Lewis, Jones, & Keeley, 2016) is another new tool that takes a comprehensive approach. The ACL is made up of self-report items, an interview, and task completions. The original research demonstrated reliability and validity in both student and criminal samples, yet no more use of the ACL has been published since.

The *Psychopathic Personality Traits Scale* (PPTS; Boduszek et al., 2016) is a new, 20-item self-report measure that is designed for both criminal and non-criminal samples, though it was validated with only Polish criminals. Respondents answer in a binary agree or disagree format, as Boduszek et al. that multiple-point Likert-scale were confusing for some of the uneducated criminals in the sample.

Finally, the *Psychopathic Processing and Personality Assessment* (PAPA; Lewis et al., 2017) was developed using an expert Delphi, which constructed the scale from item generation to the final version. This is another self-report measure consisting of 32 items which are answered on a five-point Likert-scale. The PAPA was developed with both criminal and student samples, and factor analyses generated a four-factor structure (dissocial tendencies, emotional detachment, disregard for others, and lack of sensitivity to emotion). PAPA scores also demonstrated convergent validity with the LSRP (Lewis et al., 2017).

Appendix B
CRT-WP Initial Item List

BOLD = High JM / “Psychopathic” option

ITALICS = Low JM / “Anti-Psychopathic” option

Example Item. Some people in leadership positions consider their subordinates as pawns that are used to get things done for more important people, similar to the pawns in a game of chess. This means that these leaders think that it is best to use, control, and manipulate all of their subordinates to achieve the goals of the organization in any way that they see fit. This leadership style can be a very effective one.

However, what is the biggest issue with comparing subordinate employees to pawns?

- A) **Unlike chess pieces, subordinate employees do not always do what you tell them to**
- B) It bridges the gap between fellow organizations
- C) It is not a viable strategy in workplaces with no internet connection
- D) *All employees should be treated with respect and consideration*

[All 60 items were originally presented here (10 items per JM). Due to the implicit nature of the CRT-WP, the items are removed from this version to keep them private. However, please contact the author if you wish to see the entire list of items – ryan.cook@smu.ca]

Appendix C

Subject-Matter-Expert Task

James (1998) argued that differences in one’s personality are reflected in attitudinal tendencies and biases to favour, associate, assume, and choose certain options in every day life. By designing conditional reasoning problems with response options that are associated with different attitudinal tendencies and biases, we should be able to implicitly measure personality without the respondent even knowing. The current study is designing an implicit measure of workplace psychopathy by using this method.

At this time, please detach the last page of this booklet. That page gives names and definitions of the six justification mechanisms (i.e., biases, attitudes, belief systems) that are inherent to psychopathic ways of thinking. Please familiarize yourself with these six definitions. Similar to “factors” in a regular personality measure, each item in the rest of this document was designed to measure **one** of these six justification mechanisms. To measure them implicitly, each of the following items actually have **two logically correct response options**, and **two nonsensical response options**. One of the logically correct response options is worded in a way that **someone high in that justification mechanism would choose**, and the other correct response option is worded in a way that **someone low in that justification mechanism would choose**. Using the detached page of definitions, your task is to **identify the justification mechanism measured by each item, and the response options which represent high and low**.

An example is provided below this paragraph for you to follow along. It may be easiest to identify the nonsensical response options first. In this case, option A) is clearly not logical, and neither is option C) since you chose to hire Applicant B to avoid hiring someone with criminal offense. We eliminate these two options with an “X” in both boxes. Now, with only two remaining options, B) and D), it may be easier for you to identify which justification mechanism is being targeted by this item, since one indicates a high degree of it and the other indicates a low degree of it. In this case, option B) would signify that your motivation to do this questionable action is likely guided by your own beliefs, while option D) would mean deflecting the blame to someone else as the likely explanation. Based on this information, and consulting the justification mechanism page provided, we can say that this item likely measures externalization (EXT) and that B) would be low externalization, and D) would be high externalization. Thus, we write “low” in the box for B), “high” in the box for D), and circle “EXT” before moving to the next item.

0. You are in charge with hiring a new employee, and there are two applicants remaining for you to choose from. Applicant A has better test scores than Applicant B, however, Applicant A also has a criminal record. Based on your knowledge of the criminal offense, you choose to hire Applicant B.

Which is most likely true about your motivations in choosing Applicant B?

- | | |
|------|--|
| X | A) You may have chosen Applicant A if it were later in the week |
| low | B) You have a personal bias against hiring criminals |
| X | C) You believe Applicant B also has a criminal record |
| high | D) You believe that your superiors would disapprove of hiring someone with a criminal record |

EXT	CI	SS	FLN	RSI	INS
-----	----	----	-----	-----	-----

Follow the instructions given on the previous page and please write any other comments you have about wording, item difficulty, etc. beside that item, or on the back of the final page.

1. Some people in leadership positions consider their subordinates pawns that are used to get things done for more important people, similar to the pawns in a game of chess. This means that these leaders think that it is best to use, control, and manipulate all of their subordinates to achieve the goals of the organization in any way that they see fit. This leadership style can be a very effective one.

However, what is the biggest issue with comparing subordinate employees to pawns?

- | | |
|--------------------------|--|
| <input type="checkbox"/> | A) Unlike chess pieces, subordinate employees do not always do what you tell them to |
| <input type="checkbox"/> | B) It bridges the gap between fellow organizations |
| <input type="checkbox"/> | C) It is not a viable strategy in workplaces with no internet connection |
| <input type="checkbox"/> | D) All employees should be treated with respect and consideration |

EXT	CI	SS	FLN	RSI	INS
-----	----	----	-----	-----	-----

[All 60 items were originally presented here in randomized order. Due to the implicit nature of the CRT-WP, the items are removed from this version to keep them private. However, please contact the author if you wish to see the entire list of items – ryan.cook@smu.ca]

Justification Mechanisms

Externalization (EXT)	A propensity to blame other people or external factors for negative occurrences. This bias appears as a “global irresponsibility” for actions and outcomes that clearly resulted from choices under the control of the individual. Individuals with this bias will deflect blame and absolve themselves of any wrongdoings.
Carefree Impulsivity (CI)	A predisposition for actions and decisions to be guided by impulsivity instead of reasoning, deliberation, or long-term planning. Actions will often seem to have a disregard for socially accepted norms and behaviours. The excitement of spontaneity combined with a lack of consideration for potentially harmful outcomes results in this predisposition.
Social Superiority (SS)	A persisting belief that one’s social status and social skills are superior to generally everyone around them. The individual will believe that they can charm and persuade others in any situation. The individual also believes that he/she is a dominant, alpha social personality that should be considered above others.
Fearlessness (FLN)	A proclivity toward risk-taking behaviours along with a high tolerance/resilience for the uncertainty in the outcomes. This is accompanied by an abnormal disregard for, and lack of, the fear or anxiety that most people experience in high stress situations. The individual has an aversion to the status-quo and is not satisfied with just being content.
Ruthless Self-Interest (RSI)	The tendency to actively seek out opportunities for self-promotion with complete disregard for anyone or anything other than the self. The individual strives to achieve their own goals and advancement at any cost, and will find a way to justify exploitation and other behaviours that negatively effect others as a result. There is a survival-of-the-fittest mentality.
Insensitivity (INS)	A disinclination to feel concern, guilt, remorse, or give any consideration to the feelings of others. This is a complete lack of empathy in any situation. This differs from ruthless personal gain in that this insensitivity is present even in situations where there is nothing to gain for the individual.

Appendix D**Revised Item List Following Study 1 SME Revision**

BOLD = High JM / “Psychopathic” option

ITALICS = Low JM / “Anti-Psychopathic” option

Example Item. Some people in leadership positions consider their subordinates as pawns that are used to get things done for more important people, similar to the pawns in a game of chess. This means that these leaders think that it is best to use, control, and manipulate all of their subordinates to achieve the goals of the organization in any way that they see fit. This leadership style can be a very effective one.

However, what is the biggest issue with comparing subordinate employees to pawns?

- A) Unlike chess pieces, subordinate employees do not always do what you tell them to do**
- B) It bridges the gap between fellow organizations
- C) It is not a viable strategy in workplaces with no internet connection
- D) All employees should be treated with respect and consideration*

[The revised list of 43 items was originally presented here. Due to the implicit nature of the CRT-WP, the items are removed from this version to keep them private. However, please contact the author if you wish to see the entire list of items – ryan.cook@smu.ca]

Appendix E**Items for Mahmut et al.'s (2011) adapted version of the SRP-III**

Uses a 5-point Likert-scale format ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Numbers after each item's factor correspond to the order of the original SRP-III

1. It bothers me to hurt other peoples' feelings. (Callous Affect 1) [R]
2. I am careful about what I say to people. (CA2) [R]
3. I am often rude to people. (CA3)
4. I get in trouble for the same things time after time. (CA4)
5. I sometimes enjoy hurting the people who care for me. (CA6)
6. On average, my friends would probably say I am a kind person. (CA7) [R]
7. I'm not afraid to step on others to get what I want. (CA8)
8. I'm a soft-hearted person. (CA9) [R]
9. I'm a rebellious person. (Erratic LifeStyle 1)
10. I like to change jobs often. (ELS2)
11. I've often done something dangerous just for the thrill of it. (ELS3)
12. I enjoy taking risks. (ELS4)
13. I'd be good at a dangerous job because I make fast decisions. (ELS5)
14. I hate high speed driving. (ELS7) [R]
15. I enjoy drinking and doing wild things. (ELS8)
16. Rules are made to be broken. (ELS9)
17. I think I could "beat" a lie detector. (Interpersonal Manipulation 1)
18. It's amusing to see other people get tricked. (IPM2)
19. I don't think of myself as tricky or sly. (IPM3) [R]
20. I would get a "kick" out of scamming someone. (IPM4)
21. It's fun to see how far you can push people before they get upset. (IPM6)
22. I find it easy to manipulate people. (IPM8)
23. Conning people makes me nervous. (IPM9) [R]
24. People can usually tell if I am lying. (IPM10) [R]
25. I have stolen money from my parents. (Criminal Tendencies 1)
26. I have avoided paying for things, such as movies, bus or train rides and food. (CT2)
27. I have cheated on school tests. (CT3)
28. I have been arrested. (CT4)
29. I have handed in a school essay that I copied at least partly from someone else. (CT5)

30. I have been involved in delinquent gang activity. (CT6)
31. I have broken into a building or vehicle in order to steal something or to vandalize. (CT7)
32. I have yelled at a teacher. (CT8)
33. I have tried a drug that could have been dangerous. (CT9)
34. I have shoplifted. (CT10)

Appendix F

Items for the Brief Triarchic Psychopathy Measure (TriPM; Patrick, 2010)

Participants are asked to indicate their level of agreement with each item on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree).

Boldness factor

1. I'm optimistic more often than not.
2. I have no strong desire to parachute out of an airplane. [R]
3. I am well-equipped to deal with stress.
4. I get scared easily. [R]
5. I'm a born leader.
6. I have a hard time making things turn out the way I want. [R]
7. I have a knack for influencing people.
8. I function well in new situations, even when unprepared.
9. I don't think of myself as talented. [R]
10. I'm afraid of far fewer things than most people.
11. I can get over things that would traumatize others.
12. It worries me to go into an unfamiliar situation without knowing all the details. [R]
13. I can convince people to do what I want.
14. I don't like to take the lead in groups. [R]
15. It's easy to embarrass me. [R]
16. I stay away from physical danger as much as I can. [R]
17. I don't stack up well against most others. [R]
18. I never worry about making a fool of myself with others.
19. I'm not very good at influencing people. [R]

Meanness factor

20. How other people feel is important to me. [R]
21. I would enjoy being in a high-speed chase.
22. I don't mind if someone I dislike gets hurt.
23. I sympathize with others' problems. [R]
24. I enjoy a good physical fight.
25. I return insults.
26. It doesn't bother me to see someone else in pain.

27. I enjoy pushing people around sometimes.
28. I taunt people just to stir things up.
29. I don't see any point in worrying if what I do hurts someone else.
30. I am sensitive to the feelings of others. [R]
31. I don't have much sympathy for people.
32. For me, honesty really is the best policy. [R]
33. I've injured people to see them in pain.
34. I sometimes insult people on purpose to get a reaction from them.
35. Things are more fun if a little danger is involved.
36. I don't care much if what I do hurts others.
37. It's easy for me to relate to other people's emotions. [R]
38. It doesn't bother me when people around me are hurting.

Disinhibition factor

39. I often act on immediate needs.
40. I've often missed things I promised to attend.
41. My impulsive decisions have caused problems with loved ones.
42. I have missed work without bothering to call in.
43. I jump into things without thinking.
44. I've gotten in trouble because I missed too much school.
45. I have good control over myself. [R]
46. I have taken money from someone's purse or wallet without asking.
47. People often abuse my trust.
48. I keep appointments I make. [R]
49. I often get bored quickly and lose interest.
50. I have conned people to get money from them.
51. I get in trouble for not considering the consequences of my actions.
52. I have taken items from a store without paying for them.
53. I have a hard time waiting patiently for things I want.
54. I have lost a friend because of irresponsible things I've done.
55. Others have told me they are concerned about my lack of self-control.
56. I have robbed someone.
57. I have had problems at work because I was irresponsible.
58. I have stolen something out of a vehicle.

Appendix G

Items for the Counterproductive Work Behavior Checklist (CWB-C; Spector et al., 2006)

Participants are asked to indicate how often they engage in each of the following CWB via response choices ranging from 1 (Never) to 5 (Daily or Almost Daily).

1. Purposely wasted your employer's materials/supplies (Sabotage)
2. Purposely damaged a piece of equipment or property (Sabotage)
3. Purposely dirtied or littered your place of work (Sabotage)
4. Came to work late without permission (Withdrawal)
5. Stayed home from work and said you were sick when you were not (Withdrawal)
6. Taken a longer break than you were allowed to take (Withdrawal)
7. Left work earlier than you were allowed to (Withdrawal)
8. Purposely did your work incorrectly (Production Deviance)
9. Purposely worked slowly when things needed to get done (Production Deviance)
10. Purposely failed to follow instructions (Production Deviance)
11. Stolen something belonging to your employer (Theft)
12. Took supplies or tools home without permission (Theft)
13. Put in to be paid more hours than you worked (Theft)
14. Took money from your employer without permission (Theft)
15. Stole something belonging to someone at work (Theft)
16. Told people outside the job what a lousy place you work for (Abuse)
17. Started or continued a damaging or harmful rumor at work (Abuse)
18. Been nasty or rude to a client or customer (Abuse)
19. Insulted someone about their job performance (Abuse)
20. Made fun of someone's personal life (Abuse)
21. Ignored someone at work (Abuse)
22. Blamed someone at work for an error you made (Abuse)
23. Started an argument with someone at work (Abuse)
24. Verbally abused someone at work (Abuse)
25. Made an obscene gesture (i.e., the finger) to someone at work (Abuse)
26. Threatened someone at work with violence (Abuse)
27. Threatened someone at work, but not physically (Abuse)
28. Said something obscene to someone at work to make them feel bad (Abuse)
29. Did something to make someone at work look bad (Abuse)

- 30. Played a mean prank to embarrass someone at work (Abuse)
- 31. Looked at someone at work's private mail/property without permission (Abuse)
- 32. Hit or pushed someone at work (Abuse)
- 33. Insulted or made fun of someone at work (Abuse)

Appendix H**Items for the HEXACO-60 (Ashton & Lee, 2009)**

Participants respond to each item based on the extent to which they agree it describes them on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree).

1. I would be quite bored by a visit to an art gallery. (Openness) [R]
2. I plan ahead and organize things, to avoid scrambling at the last minute. (Conscientiousness)
3. I rarely hold a grudge, even against people who have badly wronged me. (Agreeableness)
4. I feel reasonably satisfied with myself overall. (Extraversion)
5. I would feel afraid if I had to travel in bad weather conditions. (Emotional Stability)
6. I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed. (Honesty-Humility)
7. I'm interested in learning about the history and politics of other countries. (O)
8. I often push myself very hard when trying to achieve a goal. (C)
9. People sometimes tell me that I am too critical of others. (A) [R]
10. I rarely express my opinions in group meetings. (Ex) [R]
11. I sometimes can't help worrying about little things. (E)
12. If I knew that I could never get caught, I would be willing to steal a million dollars. (H-H) [R]
13. I would enjoy creating a work of art, such as a novel, a song, or a painting. (O)
14. When working on something, I don't pay much attention to small details. (C) [R]
15. People sometimes tell me that I'm too stubborn. (A) [R]
16. I prefer jobs that involve active social interaction to those that involve working alone. (Ex)
17. When I suffer from a painful experience, I need someone to make me feel comfortable. (E)
18. Having a lot of money is not especially important to me. (H-H)
19. I think that paying attention to radical ideas is a waste of time. (O) [R]
20. I make decisions based on the feeling of the moment rather than on careful thought. (C) [R]
21. People think of me as someone who has a quick temper. (A) [R]
22. On most days, I feel cheerful and optimistic. (Ex)
23. I feel like crying when I see other people crying. (E)
24. I think that I am entitled to more respect than the average person is. (H-H) [R]
25. If I had the opportunity, I would like to attend a classical music concert. (O)
26. When working, I sometimes have difficulties due to being disorganized. (C) [R]
27. My attitude toward people who have treated me badly is "forgive and forget". (A)
28. I feel that I am an unpopular person. (Ex) [R]

29. When it comes to physical danger, I am very fearful. (E)
30. If I want something from someone, I will laugh at that person's worst jokes. (H-H) [R]
31. I've never really enjoyed looking through an encyclopedia. (O) [R]
32. I do only the minimum amount of work needed to get by. (C) [R]
33. I tend to be lenient in judging other people. (A)
34. In social situations, I'm usually the one who makes the first move. (Ex)
35. I worry a lot less than most people do. (E) [R]
36. I would never accept a bribe, even if it were very large. (H-H)
37. People have often told me that I have a good imagination. (O)
38. I always try to be accurate in my work, even at the expense of time. (C)
39. I am usually quite flexible in my opinions when people disagree with me. (A)
40. The first thing that I always do in a new place is to make friends. (Ex)
41. I can handle difficult situations without needing emotional support from anyone else. (E) [R]
42. I would get a lot of pleasure from owning expensive luxury goods. (H-H) [R]
43. I like people who have unconventional views. (O)
44. I make a lot of mistakes because I don't think before I act. (C) [R]
45. Most people tend to get angry more quickly than I do. (A)
46. Most people are more upbeat and dynamic than I generally am. (Ex) [R]
47. I feel strong emotions when someone close to me is going away for a long time. (E)
48. I want people to know that I am an important person of high status. (H-H) [R]
49. I don't think of myself as the artistic or creative type. (O) [R]
50. People often call me a perfectionist. (C)
51. Even when people make a lot of mistakes, I rarely say anything negative. (A)
52. I sometimes feel that I am a worthless person. (Ex) [R]
53. Even in an emergency I wouldn't feel like panicking. (E) [R]
54. I wouldn't pretend to like someone just to get that person to do favors for me. (H-H)
55. I find it boring to discuss philosophy. (O) [R]
56. I prefer to do whatever comes to mind, rather than stick to a plan. (C) [R]
57. When people tell me that I'm wrong, my first reaction is to argue with them. (A) [R]
58. When I'm in a group of people, I'm often the one who speaks on behalf of the group. (Ex)
59. I remain unemotional even in situations where most people get very sentimental. (E) [R]
60. I'd be tempted to use counterfeit money, if I were sure I could get away with it. (H-H) [R]

Appendix I

Adapted Items for Academic Dishonesty/Cheating based on McCabe and Trevino (1993)

Respondents identify the frequency with which they had engaged in each of these behaviors on a four-point Likert scale from Never (1) to Many Times (4).

1. using unauthorized notes during a test
2. communicating with another student during a test
3. using unfair methods to learn what was on a test before it was given
4. copying from another student during a test without their knowledge
5. helping someone else to cheat on a test
6. copying material from another student and turning it in as your own work
7. handing in work which was completed entirely by someone else
8. receiving substantial unpermitted help on an assignment or paper
9. collaborating on an assignment or paper when the instructor asked for individual work
10. copying a few sentences of material from a published source without citing it