Psychological Injury and Law
https://doi.org/10.1007/s12207-023-09486-0

Over-Reporting Detection on the Psychological Inventory of Criminal Thinking Styles (PICTS) Confusion (Cf-r) Scale in Justice-Involved Individuals

Sarah Hirsch[1] · Paul B. Ingram[1] · Kymmalett A. Ross[1] · Jessica Mattera[2] · Robert D. Morgan[1,2]

Received: 11 May 2023 / Accepted: 27 September 2023

Abstract
Evaluation of criminal thinking is important in correctional assessments because of its salience to recidivism, or relapse of criminal behavior. The Psychological Inventory of Criminal Thinking Styles (PICTS) is a common instrument which assesses criminal thinking, one of the most salient risk factors of recidivism. However, little is known about the accuracy of the validity scales of this instrument. This study examines the effectiveness of the PICTS' over-reporting validity indicator, the Confusion-revised (Cf-r) scale, using the Minnesota Multiphasic Personality Inventory-2-Restructured Form's (MMPI-2-RF) over-reporting validity scales as criterion measures. The sample is composed of 165 mostly white (66%), mid-to-late 20 s ($M = 27.4$ years old, SD = 8.3), males who underwent a mental health evaluation in a court-ordered residential treatment facility. All participants had a prior substance-use-disorder diagnosis and moderate to high criminal risk. Mean comparisons across MMPI-2-RF recommended over-reporting scales as well as classification accuracy analyses were conducted. The results support the use of the PICTS recommended cut-score for Cf-r, as it produced high specificity (> 0.90) and strong area under the curve (AUC) classification accuracy (AUC = 0.76, 95% CI = [0.67, 0.84]). The implications for this study are the discussions in terms of criminal-thinking evaluation.

Keywords Over-reporting · SVT · Validity · PICTS · MMPI-2-RF

Introduction

The USA has one of the highest populations of persons incarcerated in the entire world with 1.53 million people (0.5%) in state or federal correctional facilities (Carson & Anderson, 2016). The USA also has a higher-than-average recidivism rate, at a staggering 60% (Yukhnenko et al., 2019). Housing justice-involved persons in correctional facilities is expensive, costing the USA $391.18 per justice-involved person per day (approximately $143,000 per person each year; Morgan, 2018). This net cost is nine times the cost of utilizing community correctional programs (e.g., probation, parole, work-release; Morgan, 2018). Accordingly, reducing recidivism is a major emphasis of correctional settings, and the accurate prediction of future criminal behavior is paramount to that effort. Assessing criminal risk factors is the most common and effective way to predict future criminal behavior (Bonta, 1996; Bonta & Andrews, 2007).

Bonta and Andrews (2016) developed a well-established model identifying eight prominent factors (termed criminogenic risk factors) associated with future criminal behavior—these eight factors are referred to as the "central eight". This model includes both static and dynamic risk factors (i.e., those which are unchangeable due to their historical nature, such as criminal history, and those which reflect held attitudes or ongoing behavioral choices), all of which exist conjunctively in contributing to overall risk. The best predictors (called the "Big Four") of future criminal behavior include

✉ Sarah Hirsch
Sarah.hirsch@ttu.edu

Paul B. Ingram
Paul.ingram@ttu.edu

Kymmalett A. Ross
Kymmalett.ross@ttu.edu

Jessica Mattera
Jessica.mattera@siu.edu

Robert D. Morgan
Robert.d.morgan@siu.edu

[1] Department of Psychological Sciences, Texas Tech University, 2810 18Th Street, Box 42051, Lubbock, TX 79409, USA

[2] Southern Illinois University, 1263 Lincoln Dr, Carbondale, IL 62901, USA

Published online: 12 October 2023

both static (criminal history) and dynamic (antisocial personality pattern, antisocial thinking, and antisocial associates) dimensions of the central eight (Andrews & Bonta, 2010; Andrews et al., 2006; Gendreau, 1996). As the dynamic risk factors are those which are most amenable to change, they are the ones which offer the greatest promise for correctional intervention (e.g., Andrews & Bonta, 2010). Among the "big four," antisocial cognitions are especially important in predicting future criminogenic behavior (Schlager & Pacheco, 2011). One of the most extensively used measures to assess such cognitions is the Psychological Inventory of Criminal Thinking Styles (PICTS; Walters, 1995a). PICTS is a self-report measure that has clinical utility for predicting recidivism for females and males in both community probation and incarcerated (e.g., prison and jail) samples (Walters & Lowenkamp, 2016), lower engagement in treatment (Yang et al., 2013), longer sentence lengths, and increased denial of mental health services (Mandracchia & Morgan, 2010).

Because of its utility and use within correctional populations, PICTS contains validity scales which are designed to assist clinicians in determining if the respondent is answering questions openly and honestly. Little is known about the ability of PICTS to detect invalid responding, as few studies have examined its utility (e.g., Rosenman et al., 2011; Tonks & Stephenson, 2020). When it was first created, PICTS included two validity scales to assess exaggerated (confusion scale; Cf) or minimized (defensiveness scale; Df) psychopathology, respectively. However, these scales had unacceptably low rates of retest stability, internal consistency, and classification accuracy (Walters, 1995a, b) which led to their ultimate revision (Walters, 2001). In the revision process of PICTS' validity scales, Walters (2001) describes removing four negatively scored Cf items, dropping four positively scored Df items, replacing these eight items with items from the thinking styles scales pools, and ultimately retaining these eight items from the thinking styles scales pool as the final items in these scales (Walters, 2013). As a result of this change, the revised confusion (Cf-r) and defensiveness (Df-r) scales were most improved in their internal consistency (Cf/Cf-r $\Delta\alpha = 0.35$; Df/Df-r $\Delta\alpha = 0.36$) and test–retest reliability (Cf/Cf-r $\Delta r = 0.29$; Df/Df-r $\Delta r = -0.19$; see Walters, 2001). Although modest improvements in defensive response detection were made, over-reporting detection was relatively unchanged (Df/Df-r Cohen $d\Delta = 0.54$ [medium effect difference], Cf/Cf-r Cohen $d\Delta = 0.11$ [negligible effect difference]; see Walters, 2001). This is a noteworthy limitation given the propensity of justice-involved individuals to over- or under-report psychopathology. In fact, some estimates put the rate of dishonest responding as high as 32% in correctional evaluations (Pollock et al., 1997).

Limited research has examined the construct validity of these scales; Walters and Geyer (2005) examined the correlation between the PICTS validity scales to validity scales on a widely utilized personality assessment, the Personality Assessment Inventory (PAI; Morey, 1991). The authors found positive correlations (r [134] = 0.45, $p < 0.001$) between PICTS' Cf-r and negative impression (NIM; examining "faking bad") and negative correlations (r [134] = − 0.50, $p < 0.001$) between PICTS' Cf-r and the positive impression (PIM; examining "faking good").

Current Study

This study builds off the revision and existing literature on the PICTS validity scales by examining the construct validity of the Cf-r scale. Inclusive of both versions, and despite the impact of invalid responding on test interpretation, PICTS' validity scales remain largely understudied with a lingering criticism being that no independent investigations have been untaken for these scales in justice-involved populations (see Tonks & Stephenson, 2020). This study examines the effectiveness of PICTS' Cf-r scale to detect over-reporting. As the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) is a widely used measure in the assessment with justice-involved individuals (e.g., Burchett & Bagby, 2021), and is a gold standard for these assessments due to the highly validated over-reporting scales (Ingram & Ternes, 2016; Sellbom & Bagby, 2010; Sharf et al., 2017), we used this measure to form the basis of our PICTS examination. The over-reporting scales of the MMPI-2-RF are effective at detecting invalid over-reporting in community and criminal forensic settings (Aparcero et al., 2022; Ingram & Ternes, 2016; Rogers et al., 2011; Sellbom & Bagby, 2010; Sharf et al., 2017), and include scales that capture a variety of over-reporting concerns in various populations (i.e., infrequent response scale [F-r], infrequent psychopathology responses scale [Fp-r], infrequent somatic response scale [Fs-r], Symptom validity scale [RBS], and response bias scale [RBS]).

For this study, we hypothesized (1) that PICTS' Cf-r scale would demonstrate moderate-effect size differences between those with valid and invalid MMPI-2-RF over-reporting scales, with those who have invalid MMPI-2-RF profiles having higher Cf-r scores, and (2) that PICTS' Cf-r scale would produce better specificity than sensitivity at its suggested cut-score, consistent with typical over-reporting scale performance and design (see Slick et al., 1999; Sweet et al., 2021).

Method

Participants

Potential participants were 194 male residents with a substance-use disorder and moderate to high criminal risk

in a court-ordered residential treatment facility referred for a psychological evaluation. Following the exclusion for random and fixed responding patterns (VRIN-r and TRIN-r; see Procedure), the study sample included 165 male residents. In general, participants are mostly white (66.1%), not Hispanic/Latino[1] (73.3%), 27.3 years old (SD = 8.3), and have an average of 11.4 years of education (SD = 1.7). Most participants were in the facility for their first crime (81.8%). A total of 15% of the participants' current crimes were against another person and 15% of the participants' current crimes were violent. One-quarter of the participants' past crimes were against another person and one-quarter of participants' past crimes were violent. Participants reported various mental health diagnoses, with the most common diagnosis of a substance use disorder ($n = 97$, 61.0%). See Table 1 for complete demographic data.

Table 1 Sample demographic information

	M	*SD*	*n*	%
Age	27.3	8.3		
Education length	11.4	1.7		
Male			165	100
Race				
White/Caucasian			109	66.1
Mexican			25	15.2
Black/African American			18	10.9
Asian			1	0.6
Native American/Pacific islander			1	0.6
Ethnicity				
Not Hispanic/Latino			121	73.3
Hispanic/Latino			30	18.2
Mental health diagnoses (non-exclusive)				
Attention deficit hyperactivity disorder			5	3
Posttraumatic stress disorder			15	9.1
Bipolar disorder			27	16.4
Unipolar depressive disorder			50	30.3
Generalized anxiety disorder			40	24.2
Other anxiety disorder			11	6.7
Antisocial personality disorder			74	44.8
Psychotic disorder			11	6.7
Other personality disorder			17	10.3
Substance use disorder			100	60.6
Other disorder			10	4.8
Current crime type				
Violent crime			26	15.7
Against a person			25	15.1
Crime history type				
Violent crime			43	26
Against a person			44	26.6
Number of previous charges	3.9	5.6		

Demographic information is not fully representative of the sample due to the missing data. $n = 165$

Measures

Psychological Inventory of Criminal Thinking Styles

The Psychological Inventory of Criminal Thinking Styles (PICTS; Walters, 2002) is an 80-item self-report measure designed to measure a participant's thinking associated with criminal behavior using a four-point Likert-type scale (response options ranging from one to four) including *disagree, uncertain, agree,* and *strongly agree* with higher scores yielding higher levels of criminal cognitions. PICTS includes 15 scales, including two validity scales (confusion-revised [Cf-r] and defensiveness-revised [Df-r]); two content scales (current [CUR] and historical [HIS]); two composite scales (reactive [R] and proactive [P]); eight thinking pattern scales (cutoff [Co]; cognitive indolence [Ci]; discontinuity [Ds]; entitlement [En]; mollification [Mo]; power orientation [Po]; sentimentality [Sn]; and super-optimism [So]); and a measure of general criminal thinking (GCT).

Particularly relevant to the purpose of this study are the validity scales of PICTS. The Cf-r scale assesses the exaggeration of symptoms (i.e., "faking bad") by including items covering experiences not common in the general population (e.g., smelling odd odors or hearing voices). The Cf-r scale is utilized because the main scope of this paper is understanding the validity of overreporting scales. The confusion-revised scale has demonstrated acceptable levels of internal consistency for both men (coefficient $\alpha = 0.66$) and women (coefficient $\alpha = 0.77$), and a test–retest stability for men in a 10-week interval ($r = 0.64$) and for women in a 12-week interval ($r = 0.87$; Walters, 2001). In this study, Cf-r has high internal consistency ($\alpha = 0.82$), and cut-scores were based

[1] Standard wording provided by the United States Census Bureau was used to assess race and ethnicity: The U.S. Office of Management and Budget (OMB) requires federal agencies to use a minimum of two ethnicities in collecting and reporting data: Hispanic or Latino and Not Hispanic or Latino. OMB defines "Hispanic or Latino" as a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race. People who identify with the terms "Hispanic" or "Latino" are those who classify themselves in one of the specific Hispanic or Latino categories listed on the decennial census questionnaire and various Census Bureau survey questionnaires—"Mexican, Mexican Am., Chicano" or "Puerto Rican" or "Cuban"—as well as those who indicate that they are "another Hispanic, Latino, or Spanish origin" (U.S. Census Bureau, 2022).

on the manual defined invalid profile for Cf-r (see Walters, 2001).[2]

Minnesota Multiphasic Personality Inventory-2-Restructured Form

The Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Tellegen & Ben-Porath, 2008b/2011) is a 338-item self-report measure that assesses a broad array of psychopathology and related concerns. MMPI-2-RF contains 51 scales, including nine validity scales (2 non-content-based responding, 5 over-reporting, and 2 under-reporting), three higher-order (H–O) scales, nine restructured clinical (RC) scales, 23 specific problem (SP) scales which cover four symptom categories (i.e., somatic, internalizing, externalizing, and interpersonal scales), two vocational interest scales, and five personality psychopathology five (PSY-5) scales.

The over-reporting scales on MMPI-2-RF includes the infrequent response (F-r) scale designed to differentiate between genuine psychopathology and symptom over-reporting in the general population, and the infrequent psychopathology responses (Fp-r) scale was designed to differentiate between genuine psychopathology and symptom over-reporting in psychiatric populations. The infrequent somatic response (Fs-r) scale is designed to differentiate between genuine somatic complaints and over-reporting somatic complaints in medical patient populations, the symptom validity (RBS) scale is designed to differentiate between genuine and over-reported somatic and cognitive complaints, and the response bias (RBS) scale is designed to detect exaggerated memory complaints.

The over-reporting scales of MMPI-2-RF are effective at detecting invalid over-reporting in community and criminal forensic settings (Aparcero et al., 2022; Ingram & Ternes, 2016; Rogers et al., 2011; Sellbom & Bagby, 2010; Sharf et al., 2017), with meta-analytically derived large effect sizes ranging in magnitude from 1.08 (FBS-r) to 1.43 (Fp-r; Ingram & Ternes, 2016). For reference, the recommended cut scores for the MMPI-2-RF over-reporting scales are 120, 100, 100, 100, and 100 for F-r, Fp-r, Fs-r, FBS-r, and RBS, respectively. These scores were used as the MMPI-2-RF cut-scores within our analyses.

[2] There are other ways in which invalidity may be conceptualized, including at lower scores on the PICTS validity scales (see Walters, 2019) or by removing cases with cases scales that do not meet the T80 cut-score but demonstrate some signs of elevation (see McGee et al., 2016). We selected T80 for mean score comparisons because it is identified within the manual (Walters, 2001) and other suggested scores are not yet widely recommended. Other cut scores were considered as part of ROC analyses and their effectiveness is detailed in that set of analyses.

Procedure

Data from this forensic sample was collected from clinical files at the residential treatment facility and scanned into digital copies by a trained research assistant. Participants were included in this study if their psychological evaluation record included both PICTS and MMPI-2-RF. Due to the nature of data collection, we were unable to know if the administration on the measures were conducted in the same testing day or over the course of two testing sessions (or potentially more on rare occasions, based on correctional setting events which necessitate such modifications). Participants were screened for non-responding and non-content-based responding on MMPI-2-RF and PICTS. Participants were excluded because of random (VRIN-r $\geq$ 80; $n = 3$) or fixed (TRIN-r $\geq$ 80; $n = 20$) responding. No exclusions were made for under-reporting (K-r; L-r) and cannot say (CNS). Individuals were also excluded if 10 or more responses were missing on PICTS ($n = 6$), as suggested by PICTS' scoring manual (Walters, 2002).

Statistical Analyses

Descriptive statistics and correlations between PICTS Cf-r and MMPI-2-RF scales were conducted (see Tables 1 and 2). Next, independent *t*-tests (along with effect sizes) were conducted on the Cf-r scale, contrasting those individuals who passed and failed each MMPI-2-RF over-reporting scale based on their standard cut scores (F-r = 120; Fp-r $\geq$ 100; Fs-r $\geq$ 100; FBS-r $\geq$ 100; RBS $\geq$ 100; see Table 3).[3] In the current study, 22 participants invalidated the F-r scale, 70 participants invalidated the Fp-r scale, 18 participants invalidated the Fs-r scale, one participant invalidated the FBS-r scale, and 18 participants invalidated the RBS scale. Effect-sizes magnitudes were determined using Cohen's (1988) guidelines: $d = \geq 0.80, \geq 0.50$, and ≥ 0.20 for large, medium, and small, respectively.

Medium effects or larger were determined, a priori, to indicate clinically meaningful differences.

Because the MMPI-2-RF over-reporting scales assess a variety of content (e.g., infrequent psychopathology and infrequent somatic symptoms) and because these scales have utilized distinctive approaches to validation (e.g., infrequent item endorsement [F-r, Fp-r, or Fs] and items were identified using Performance Validity Test identified [RBS]), these comparisons provide a means through which to quantify the types of presentation PICTS' over-reporting scale is most likely to detect. Subsequently, a

[3] We opted to retain a parametric analysis because when the data was examined using a non-parametric substitute for t-test the results did not differ.

Table 2 Correlations between PICTS and MMPI-2-RF validity scales

Pearson r correlations						
	F-r	Fp-r	Fs	FBS-r	RBS	Cf-r
F-r	1					
Fp-r	0.90**	1				
Fs	0.85**	0.84**	1			
FBS-r	0.82**	0.85**	0.84**	1		
RBS	0.92**	0.89**	0.87**	0.88**	1	
PICTS Cf-r	0.39**	0.30**	0.27**	0.10	0.26**	1

r Pearson *r* correlations, *PICTS Cf-*r psychological inventory of criminal thinking styles confusion-revised, *PICTS Df-r* psychological inventory of criminal thinking styles defensiveness-revised, *MMPI-2-RF* Minnesota Multiphasic Personality Inventory-2-Restructured Form, *F-r* infrequent responses, *Fp-r* infrequent psychopathology responses, *Fs* infrequent somatic responses, *FBS-r* symptom validity; *RBS* response bias

***$p < .001$; **$p < .01$; *$p < .05$

receiver-operating characteristic (ROC) curve determined the sensitivity, specificity, and the positive and negative predictive values comparing individuals who had a completely valid MMPI-2-RF profile (no scales invalid $n = 87$) to those who invalidated any of the over-reporting scales ($n = 78$).

Results

Correlations between PICTS' Cf-r and MMPI-2-RF scales were conducted (see Table 2). PICTS' Cf-r and MMPI-2-RF over-reporting scales of F-r, Fp-r, Fs, and RBS had low positive correlations ($r = 0.26$–0.39; $p < 0.01$). PICTS' Cf-r scale and MMPI-2-RF FBS-r's scale were not significantly correlated ($r = 0.10$; $p > 0.05$).

Classification Agreement

Participants' responses on MMPI-2-RF were classified based on the MMPI-2-RF standard cut scores (F-r ≥ 120; Fp-r ≥ 100; Fs-r ≥ 100; FBS-r ≥ 100; RBS ≥ 100). Participants were grouped as either valid or invalid responders on MMPI-2-RF, and these groups produced large effect differences across PICTS' over-reporting scale (Cf-r; $M_d = 0.51$; $d = 0.82$ [Fp-r] to 1.33 [F-r]). PICTS' Cf-r was most likely to elevate among those with an invalid F-r scale, evidencing the strong pathology competent to what is assessed on PICTS' Cf-r. Except for FBS-r, the Cf-r scale had similar rates of invalidity across scales assessing invalid psychopathological (F-r, Fp-r), somatic (Fs), and cognitive (RBS) symptom endorsement. While PICTS identified the same number of individuals ($\chi^2(1) = 13.92$), there were some discrepancies in the rates of individuals classified (Table 4). In discrepant cases (i.e., those where there was invalidation of either MMPI-2-RF or PICTS but not both), participants were more likely to invalidate the MMPI-2-RF than PICTS' Cf-r. Thus, many of the individuals identified as engaging in a form of positive impression management were not detected on Cf-r, using its standard cut-score.

Cf-r Classification Accuracy

We created another criterion-group based on a pass all/fail any approach across the five MMPI-2-RF criterion over-reporting scales using manual-specified standard cut-scores (pass all [$n = 87$; 52.7%] or fail any [$n = 78$; 47.3%]). Cf-r

Table 3 Means and standard deviations of measures and differences on the PICTS Cf-r between participants with valid versus invalid MMPI-2-RF scales

	Valid				**Invalid**				*t* (163)	*p*	*d*
	M	SD	% ≥ RCS	*n*	*M*	SD	% ≥ RCS	*n*			
F-r	62.5	14.7	13.2	143	81.7	12.3	55.0	22	5.8	< .001	1.33
Fp-r	59.9	14.0	10.5	95	72.0	15.6	30.0	70	5.2	< .001	0.82
Fs-r	63.5	15.1	15.6	147	77.6	16.1	44.4	18	3.7	< .001	1.10
FBS-r	64.9	15.7	18.3	164	92	-	100.0	1	-	-	-
RBS	63.4	15.3	15.6	147	78.2	14.1	44.4	18	3.9	< .001	0.97

M mean, *SD* standard deviation, *RCS* recommended cut score, *PICTS Cf-r* psychological inventory of criminal thinking styles confusion-revised, *MMPI-2-RF* Minnesota Multiphasic Personality Inventory-2-Restructured Form, *F-r* infrequent responses, *Fp-r* infrequent psychopathology responses, *Fs* infrequent somatic responses, *FBS-r* symptom validity, *RBS* response bias scale

Table 4 Diagnostic efficacy statistics for Cf-r in classifying over-reporting responses

Fail any MMPI-2-RF over-reporting scale								
PICTS Cf-r score	SN	SP	BR = 0.15		BR = 0.30		BR = 0.50	
			PPV	NPV	PPV	NPV	PPV	NPV
T > 90	0.17	0.95	0.39	0.87	0.61	0.73	0.78	0.53
T > 85	0.22	0.93	0.36	0.87	0.58	0.74	0.76	0.54
***T* > 80**	0.31	0.92	0.40	0.88	0.62	0.76	0.79	0.57
T > 75	0.46	0.87	0.39	0.90	0.61	0.79	0.78	0.62
T > 70	0.55	0.77	0.30	0.91	0.51	0.80	0.71	0.63
T > 65	0.69	0.67	0.27	0.92	0.47	0.83	0.68	0.68
T > 60	0.72	0.60	0.24	0.92	0.43	0.83	0.64	0.68

The value in bold (T>80) is the current recommended cut-score for the PICTS Cf-r

PICTS Cf-r psychological inventory of criminal thinking styles confusion-revised, *SN* sensitivity, *SP* specificity, *PPV* positive predictive value, *NPV* negative predictive value, *BR* base rate

produced a large classification effect (area under curve [AUC] = 0.72, 95% CI = [0.64, 0.80]) based on this criterion (Rice & Harris, 2005; Salgado, 2018). Sensitivity, specificity, and positive/negative predictive values across a variety of Cf-r scores and base rates are provided in Table 4. At Cf-r's recommended cut score (≥ 80 T), PICTS' over-reporting scale has low sensitivity (0.31) but good specificity (0.92). At base rates of 0.15, 0.3, and 0.5, this cut score has modest positive predictive values (0.40, 0.88, and 0.62, respectively), as well as strong negative predictive values (0.76, 0.79, and 0.57, respectively). Lower Cf-r cut-scores do not maintain high specificity estimates.

Discussion

This study evaluated the effectiveness of PICTS in detecting over-reporting in justice-involved persons in a correctional residential treatment facility. Despite PICTS being widely used in correctional assessments (Bonta & Andrews, 2016), and despite invalid responding being a major concern during those same evaluations (Pollock et al., 1997), research validating PICTS' validity scales was lacking. The previous literature has supported PICTS' over-reporting scales, with moderate correlations between PICTS' Cf-r validity scale and PAI's over-reporting scales (e.g., PIM and NIM; Walters & Geyer, 2005). Our study found approximately similar magnitude correlations between PICTS' over-reporting validity scales and those of the MMPI-2-RF. In general, our findings provide support for Cf-r to detect and classify over-reporting. Two notable patterns in over-reporting (i.e., Cf-r) detection were observed and are described below.

First, standard PICTS' Cf-r cut-score (*T* ≥ 80) results in better specificity (i.e., identifies those without over-reporting responses) than sensitivity (i.e., detects those with over-reporting responses). The recommended PICTS' Cf-r cut-score is better at identifying those with open and honest responding (i.e., negative predictive value [NPV] > positive predictive value [PPV]). This performance is consistent with the rule-out approach common to SVTs (e.g., Armistead-Jehle et al., 2020; Morris et al., 2021), due to the focus on lowering false positives. The sensitivity and specificity estimates for Cf-r at recommended cut-scores fell within expected and preferred ranges (e.g., ≥ 90% specificity and ~ 30% sensitivity; Wygant et al., 2011). Lowering PICTS' cut score, such as to *T* ≥ 75, produces substantial increases in sensitivity while only modestly impacting specificity, although specificity falls just below the recommended 0.9 threshold commonly accepted as a standard guideline (Rogers & Bender, 2018; Sherman et al., 2020). Given that the base rate of exaggerated responding in correctional settings can approximate to 32% (Pollock et al., 1997), providing increased overall classification accuracy (e.g., increasing either sensitivity/specificity while also not notably impacting the other) is an important goal of validity scale refinement (Rogers et al., 2003).

Second, these results are broadly consistent with not only the broader patterns of over-reporting scale effectiveness, but also with the MMPI-2-RF's performance in this study specifically. Because Cf-r identified a majority (63%) of the individuals in the same manner as standard MMPI-2-RF interpretation (see Ben-Porath, 2012), our findings offer some promise of the convergence between these two measures of over-reported criminal risk and psychopathology.

Cf-r has mostly medium associations with other over-reporting SVTs, which is lower than the large effects often observed in correctional studies (e.g., Chmielewski et al., 2017; Wygant et al., 2011). While conceptually related (e.g., over-reporting), it is possible that the lower association with SVTs may be caused by differences in the types of things measured on the MMPI-2-RF (psychopathology and personality) and PICTS (criminal risk).

Limitations and Future Directions

Our findings are preliminary in nature given limited extant literature on PICTS' Cf-r scale. Because of the broad use of PICTS in correctional contexts, our results provide important implications to ensuring valid assessment of, and treatment recommendations for, justice-involved persons. However, this study is not without its limitations. This study integrated PICTS with only one SVT criterion measure, although the MMPI-2-RF contains numerous scales measuring over-reporting (e.g., F-r, Fp-r). As a result of our outcome criterion focusing on a measure of personality and psychopathology (MMPI-2-RF), the sensitivities and specificities observed may differ compared when contrasted to other SVTs with less directly symptom-focused content. As with all validity tests (e.g., both performance and symptom validity, as well as alternative versions of the MMPI [MMPI-2/MMPI-3]), the MMPI-2-RF does not contain infallible validity scales. Thus, replication with other criteria is critical. Our study is most aptly applied to justice-involved persons within in-patient residential treatment facilities with limited violent criminal history and substance-use diagnoses. Generalizability to other correctional settings, including those which are not rehabilitative in nature, warrants further study. Despite the differences between these settings, the base rates of MMPI-2-RF validity failure in our treatment sample approximate that in other non-treatment correctional settings (e.g., Wall et al., 2015). Finally, the question of validity scale implementation on criminal-thinking measures (e.g., PICTS) remains unanswered. Namely, it is unclear if these constructs may be assessed in the same manner, expecting the same effect sizes, given the differential nature of the ideas being assessed (e.g., psychopathology and criminogenic thinking are distinct constructs; Morgan et al., 2012). Similarly, endorsement on these measures may have distinct outcomes in the justice system (i.e., saying one engages in criminal thinking is unlikely to offer the same type of benefit as endorsing higher levels of mental illness).

Conclusion

This study provides an evaluation of the content-based validity scales on a leading measure of criminal thinking (PICTS; Walters, 2001). Research on MMPI-2-RF among those with ongoing litigation has found the somatic and cognitive over-reporting scales to be particularly effective for detecting invalid physical complaints (e.g., Sharf et al., 2017). Within an incarcerated sample not selected for their litigation involvement, we found that PICTS' Cf-r scale produced approximately equitable elevations across MMPI-2-RF criterion scales assessing somatic, cognitive, and psychopathological symptom sets. As such, elevations on Cf-r reflect probable over-endorsement but not of a specific subset of symptoms. Civil litigation and/or disability occurs regularly in incarcerated populations (e.g., as a function of public versus private prison settings or as a function of the Prison Litigation Reform Act; see Gunderson, 2022; Little et al., 2022). The engagement in disability evaluation may change the probability of elevation for these different types of feigned symptoms, altering base rates of specific symptom sets. Additionally, given the release of MMPI-3 (Ben-Porath & Tellegen, 2020), it is important to note that high correlations between MMPI-2-RF and MMPI-3 validity scales suggest our results are likely to generalize when using the MMPI-3 as a criterion for over-reporting ($r = 0.95$ [Fp/Fp-r] to 1.00 [FBS/FBS-r]).

Data Availability The data that support the findings of this study are available from the corresponding author, Sarah Hirsch, upon reasonable request.

Declarations

Competing Interests Dr. Paul Ingram receives regular research support from the University of Minnesota Press, Test Division, publisher of the MMPI-2-RF and Pearson Clinical Assessments, publisher and distributor of the MMPI-2-RF. The authors have no other competing interests to declare.

References

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16*(1), 39–55. https://doi.org/10.1037/a0018362

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency, 52*(1), 7–27. https://doi.org/10.1177/0011128705281756

Aparcero, M., Picard, E. H., Nijdam-Jones, A., & Rosenfeld, B. (2022). Comparing the ability of MMPI-2 and MMPI-2-RF validity scales to detect feigning: A meta-analysis. *Assessment*. https://doi.org/10.7319/11211067535

Armistead-Jehle, P., Lippa, S. M., & Grills, C. E. (2020). The impact of self-efficacy and health locus of control on performance validity testing. *Archives of Clinical Neuropsychology, 35*(7), 1162–1167. https://doi.org/10.1093/arclin/acaa027

Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF*. University of Minnesota Press.

Ben-Porath, Y. S., & Tellegen, A. (2020). Minnesota multiphasic personality inventory-3 (MMPI-3): Manual for administration, scoring, and interpretation.

Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18–32). Sage Publications Inc.

Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation, 6*(1), 1–22.

Bonta, J., & Andrews, D. A. (2016). The psychology of criminal conduct. *Routledge*. https://doi.org/10.4324/9781315677187

Burchett, D., & Bagby, R. M. (2021). Assessing negative response bias: A review of the noncredible overreporting scales of the MMPI-2-RF and MMPI-3. *Psychological injury and law*, 1–15. https://doi.org/10.1007/s12207-021-09435-9

Carson, E. A., & Anderson, E. (2016). *Prisoners in 2015*. Bureau of justice statistics. https://bjs.ojp.gov/content/pub/pdf/p15.pdf

Chmielewski, M., Zhu, J., Burchett, D., Bury, A. S., & Bagby, R. M. (2017). The comparative capacity of the Minnesota multiphasic personality inventory–2 (MMPI–2) and MMPI–2 restructured form (MMPI-2-RF) validity scales to detect suspected malingering in a disability claimant sample. *Psychological Assessment, 29*(2), 199. https://doi.org/10.1037/pas0000328

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Gendreau, P. (1996). Offender rehabilitation: What we know needs to be done. *Criminal Justice and Behavior, 23*(1), 144–161. https://doi.org/10.1177/0093854896023001010

Gunderson, A. (2022). Inmate litigation, legal access, and prison privatization. *Journal of Law and Courts*, 1–21.

Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2-restructured form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist, 30*(4), 473–496. https://doi.org/10.1080/13854046.2016.1187769

Little, K., Griffin III, O. H., Butler, H. D., & Morgan, K. D. (2022). A review of inmate litigation challenging the constitutionality of solitary confinement. *Corrections*, 1–19.

Mandracchia, J. T., & Morgan, R. D. (2010). The relationship between status variables and criminal thinking in an offender population. *Psychological Services, 7*(1), 27. https://doi.org/10.1037/a0016194

McGee Ng, S. A., Bagby, R. M., Goodwin, B. E., Burchett, D., Sellbom, M., Ayearst, L. E., ... & Baker, S. (2016). The effect of response bias on the personality inventory for DSM–5 (PID–5). *Journal of Personality Assessment, 98*(1), 51–61.

Morey, L. C. (1991). *Personality assessment inventory*. Odessa, FL: Psychological assessment resources.

Morgan, A. (2018). *How much does prison really cost?: Comparing the costs of imprisonment with community corrections.* Australian institute of criminology.

Morgan, R. D., Flora, D. B., Kroner, D. G., Mills, J. F., Varghese, F., & Steffan, J. S. (2012). Treating offenders with mental illness: A research synthesis. *Law and Human Behavior, 36*(1), 37–50. https://doi.org/10.1037/h0093964

Morris, N. M., Mattera, J., Golden, B., Moses, S., & Ingram, P. B. (2021). Evaluating the performance of the MMPI-3 over-reporting scales: Sophisticated simulators and the effects of comorbid conditions. *The clinical neuropsychologist*, 1–9. https://doi.org/10.1080/13854046.2021.1968037

Pollock, P. H., Quigley, B., Worley, K. O., & Bashford, C. (1997). Feigned mental disorder in prisoners referred to forensic mental health services. *Journal of Psychiatric and Mental Health Nursing, 4*(1), 9–15.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*, 615–620. https://doi.org/10.1007/s10979-005-6832-7

Roger, R., & Bender, S. D. (Eds.). (2018). *Clinical assessment of malingering and deception* (4th ed.). Guilford Press.

Rogers, R., Gillard, N. D., Berry, D. T., & Granacher, R. P. (2011). Effectiveness of the MMPI-2-RF validity scales for feigned mental disorders and cognitive impairment: A known-groups study. *Journal of Psychopathology and Behavioral Assessment, 33*(3), 355–367.

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, *10*(2), 160–177.

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research, 2*(4), 320–332. https://doi.org/10.1504/ijbhr.2011.043414

Salgado, J. F. (2018). Transforming the area under the normal curve (AUC) into Cohen's d, Pearson's rpb, odds-ratio, and natural log odds-ratio: Two conversion tables. *European Journal of Psychology Applied to Legal Context, 10*(1), 35–47.

Schlager, M. D., & Pacheco, D. (2011). An examination of changes in LSI-R scores over time: Making the case for needs-based case management. *Criminal Justice and Behavior, 38*(6), 541–553. https://doi.org/10.1177/0093854811402300

Sellbom, M., & Bagby, R. M. (2010). Detection of overreported psychopathology with the MMPI-2 RF form validity scales. *Psychological Assessment*, *22*(4), 757–767. https://doi.org/10.1037/a0020825

Sharf, A. J., Rogers, R., Williams, M. M., & Henry, S. A. (2017). The effectiveness of the MMPI-2-RF in detecting feigned mental disorders and cognitive deficits: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment, 39*(3), 441–455. https://doi.org/10.1007/s10862-017-9590-1

Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology : THe Official Journal of the National Academy of Neuropsychologists, 35*(6), 735–764. https://doi.org/10.1093/arclin/acaa019

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(4), 545–561.

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., ... & Conference Participants. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The clinical neuropsychologist*, *35*(6), 1053–1106.

Tellegen, A., & Ben-Porath, Y. S. (2008b/2011). Technical manual citation.

Tonks, S., & Stephenson, Z. (2020). Critique of the psychological inventory of criminal thinking styles. *International Journal of Offender Therapy and Comparative Criminology, 64*(9), 903–921. https://doi.org/10.1177/0306624X19870646

U.S. Census Bureau. (2022, April 15). *About the Hispanic population and its origin*. Census.gov. https://www.census.gov/topics/population/hispanic-origin/about.html

Wall, T. D., Wygant, D. B., & Gallagher, R. W. (2015). Identifying overreporting in a correctional setting: Utility of the MMPI-2 restructured form validity scales. *Criminal Justice and Behavior, 42*(6), 610–622. https://doi.org/10.1177/0093854814556881

Walters, G. (2013). The psychological inventory of criminal thinking styles (PICTS) professional manual. Center for Lifestyle Studies.

Walters, G. D. (1995a). The psychological inventory of criminal thinking styles: Part I: Reliability and preliminary validity. *Criminal Justice and Behavior, 22*(3), 307–325. https://doi.org/10.4324/9781315677187

Walters, G. D. (1995b). The psychological inventory of criminal thinking styles: Part II: Identifying simulated response sets. *Criminal Justice and Behavior, 22*(4), 437–445. https://doi.org/10.1177/0093854895022004007

Walters, G. D. (2001). Revised validity scales for the psychological inventory of criminal thinking styles (PICTS). *Journal of Offender Rehabilitation, 32*(4), 1–13. https://doi.org/10.1300/J076v32n04_01

Walters, G. D. (2002). The psychological inventory of criminal thinking styles (PICTS): A review and meta-analysis. *Assessment, 9*(3), 278–291. https://doi.org/10.1177/1073191102009003007

Walters, G. D. (2019). Criminal thinking: Theory and practice. *The Wiley international handbook of correctional psychology*, 637–653.

Walters, G. D., & Geyer, M. D. (2005). Construct validity of the psychological inventory of criminal thinking styles in relationship to the PAI, disciplinary adjustment, and program completion. *Journal of Personality Assessment, 84*(3), 252–260. https://doi.org/10.1207/s15327752jpa8403_04

Walters, G. D., & Lowenkamp, C. T. (2016). Predicting recidivism with the psychological inventory of criminal thinking styles (PICTS) in community-supervised male and female federal offenders. *Psychological Assessment, 28*(6), 652–659. https://doi.org/10.1037/pas0000210

Wygant, D. B., Anderson, J. L., Sellbom, M., Rapier, J. L., Allgeier, L. M., & Granacher, R. P. (2011). Association of the MMPI-2 restructured form (MMPI-2-RF) validity scales with structured malingering criteria. *Psychological Injury and Law, 4*(1), 13–23. https://doi.org/10.1007/s12207-011-9098-z

Yang, Y., Knight, K., Joe, G. W., Rowan-Szal, G. A., & Lehman, W. E. K. (2013). The influence of client risks and treatment engagement on recidivism. *The Journal of Offender Rehabilitation, 52*(8), 544–564. https://doi.org/10.1080/10509674.2013.840353

Yukhnenko, D., Sridhar, S., & Fazel, S. (2019). A systematic review of criminal recidivism rates worldwide: 3-year update. *Wellcome open research*, *4*, 28. https://doi.org/10.12688/wellcomeopenres.14970.3