

# Detecting inconsistent responding on the youth psychopathic traits inventory-short form

Citation for published version (APA):

Kelley, S., Edens, J. F., Donnellan, M. B., Somma, A., Fossati, A., de Ruiter, C., Eisenbarth, H., & Vaughn, M. G. (2019). Detecting inconsistent responding on the youth psychopathic traits inventory-short form. *Assessment*, 26(3), 465-478. <https://doi.org/10.1177/1073191117738048>

## Document status and date:

Published: 01/04/2019

## DOI:

[10.1177/1073191117738048](https://doi.org/10.1177/1073191117738048)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# Detecting Inconsistent Responding on the Youth Psychopathic Traits Inventory–Short Form

Assessment  
2019, Vol. 26(3) 465–478  
© The Author(s) 2017  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1073191117738048  
journals.sagepub.com/home/asm



Shannon E. Kelley<sup>1</sup>, John F. Edens<sup>1</sup>, M. Brent Donnellan<sup>1</sup>,  
Antonella Somma<sup>2,3</sup>, Andrea Fossati<sup>2,3</sup>, Corine de Ruiter<sup>4</sup>,  
Hedwig Eisenbarth<sup>5</sup>, and Michael G. Vaughn<sup>6</sup>

## Abstract

The Youth Psychopathic Traits Inventory–Short Form (YPI-S) is a convenient measure for assessing psychopathy in settings with constraints on resources. However, the YPI-S does not contain a means of detecting careless or random response styles. The present study describes the development and evaluation of an inconsistent responding scale for the YPI-S using five archival samples that vary in language (English, German, Italian, Dutch) and other participant characteristics (juvenile offenders, adolescent students). Inconsistency scores resulting from the new scale effectively distinguished genuine participant responses from randomly generated cases (area under the ROC curve [AUC] = .85-.90) and from cases in which 50% of original responses were replaced with random data (AUC = .75-.82). The associations between the YPI-S and theoretically relevant correlates were reduced among participants exceeding proposed cutoff scores for profile validity compared with associations among more consistent respondents.

## Keywords

inconsistent responding, psychopathy assessment, Youth Psychopathic Traits Inventory–Short Form, validity scales

Although there is a long history of assessing antisocial and psychopathic traits using structured interviews and file reviews that rely on trained examiners (e.g., First, Spitzer, Gibbon, Williams, & Benjamin, 1997; Hare, 2003), various self-report inventories have demonstrated considerable utility in quantifying these attributes. Though not without some controversy, self-report measures have been designed to assess important features of psychopathy such as remorselessness, grandiosity, superficial charm, irresponsibility, and impulsivity and have the advantage of being relatively easy to administer (see Sellbom, Lilienfeld, Fowler, & McCrary, in press). One potential vulnerability of this approach, however, concerns the negative effect of random and careless responding on the psychometric properties and ultimate validity of self-report measures (e.g., Curran, 2016; Huang, Liu, & Bowling, 2015; Meade & Craig, 2012).

The Youth Psychopathic Traits Inventory (YPI; Andershed, Kerr, Stattin, & Levander, 2002) is a widely used self-report measure of psychopathy designed for adolescent populations, although studies support the validity of the measure among young adults as well (e.g., Campbell, Doucette, & French, 2009). This instrument has been extensively researched across various languages and nationalities (e.g., Fossati et al., 2016; Oshukova et al., 2015; Pechorro, Andershed, Ray, Maroco, & Gonçalves, 2015). Scores on

the YPI are associated with a range of theoretically relevant criterion variables, including externalizing psychiatric disorders, aggression, offending behaviors, and substance use (e.g., Colins, Bijttebier, Broekaert, & Andershed, 2014; Dolan & Rennie, 2007; Hillege, Das, & de Ruiter, 2010). Moreover, given the inclusion of “limited prosocial emotions” as a new specifier for the diagnosis of conduct disorder in the most recent *Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition* (American Psychiatric Association, 2013), callous–unemotional features of psychopathy are increasingly likely to be emphasized in clinical and forensic assessments of juveniles. The YPI provides one means of informing this diagnostic determination, as the measure contains item content specifically referring to

<sup>1</sup>Texas A&M University, College Station, TX, USA

<sup>2</sup>Vita-Salute San Raffaele University, Milan, Italy

<sup>3</sup>San Raffaele Turro Hospital, Milan, Italy

<sup>4</sup>Maastricht University, Maastricht, the Netherlands

<sup>5</sup>University of Southampton, Southampton, UK

<sup>6</sup>Saint Louis University, Saint Louis, MO, USA

## Corresponding Author:

Shannon E. Kelley, Texas A&M University, 4235 TAMU, College Station, TX 77843, USA.

Email: shannonkelley@tamu.edu

three of the four specifier criteria (remorselessness, callousness, and shallow affect), although it does not directly assess concern about performance (see Colins, 2016; Colins & Andershed, 2015).

When using self-report inventories, however, the validity of responses may be compromised by confounds such as inattentive or random completion of the measure. Even conservative estimates suggest that careless responding occurs rather frequently among college student populations (10-20%; Curran, 2016; Faust, Faust, Baker, & Meyer, 2012) and juvenile offenders (5-10%; Cook, Faust, Meyer, & Faust, 2016). These rates may be heightened when measures are administered anonymously, in the context of extensive survey batteries, or in distracting environments. Additionally, concerns about the influence of reading comprehension, attentiveness, and motivation on response validity may be especially applicable to juvenile populations.

Despite the ostensible need to identify “invalid” self-report data, the utility of validity scales has been the focus of some controversy and criticism. For example, measures of response style are open to different interpretations and do not consistently moderate criterion-related validity (McGrath, Mitchell, Kim, & Hough, 2010; Piedmont, McCrae, Riemann, & Angleitner, 2000; cf. Anderson, Sellbom, Wygant, & Edens, 2013; Morey, 2012). However, more recent studies counter that careless responding in particular can adversely affect factor structures (Benning & Freeman, 2017), produce spurious or attenuated validity coefficients (Kelley et al., 2017; Marcus, Church, O’Connell, & Lilienfeld, 2016; Meade & Craig, 2012), and otherwise alter the psychometric properties of measures (Cornell, Klein, Konold, & Huang, 2012; Curran, 2016; Huang et al., 2015; Wood, Harms, Lowman, & DeSimone, 2017). Furthermore, random responding on common measures of psychological functioning creates the appearance of greater maladjustment in comparison with typical scores from valid profiles (Benning & Freeman, 2017; Cook et al., 2016; Keeley, Webb, Peterson, Roussin, & Flanagan, 2016; Kelley et al., 2016).

Likewise, inattentive and random responding has important implications for evaluations conducted in applied settings to inform clinical and forensic decision-making (Cook et al., 2016). For example, careless responding could contribute to misleading impressions concerning the presence of psychopathic traits, which may in turn have negative consequences on case recommendations and the availability of intervention services for justice-involved youth (Edens, Mowle, Clark, & Magyar, 2016; Vidal & Skeem, 2007; Viljoen, McLachlan, & Vincent, 2010). Including an assessment of this response style in real-world psychological evaluations is needed to avoid serious errors in interpretation, especially as inattentiveness could be difficult to visually detect from responses and produce overall scores that

are not obviously implausible, aside from extreme fixed responding (e.g., “all false” or “all true” responses).

Because no validity scales were originally included in the development of the full-length YPI, Penson et al. (2017) recently constructed a measure of inconsistent responding for this instrument. The authors applied procedures resembling those used to construct inconsistency scales for more comprehensive personality inventories (e.g., Minnesota Multiphasic Personality Inventory–2, Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989; Personality Assessment Inventory, Morey, 1991), as well as for measures of psychopathy, including the Psychopathic Personality Inventory–Revised (PPI-R; Lilienfeld & Widows, 2005), the abbreviated PPI-R-40 (Kelley et al., 2016), and the Triarchic Psychopathy Measure (Mowle et al., 2017). The resulting inconsistency scale, the Screening Procedure for Inconsistent or Careless Item Endorsement (SPICIE), appeared to satisfactorily distinguish between conscientious and random or partially random responding on the YPI (Penson et al., 2017).

Although the length of the YPI (50 items) is not particularly cumbersome, a shortened version of the instrument (Youth Psychopathic Traits Inventory–Short Form [YPI-S]; van Baardewijk et al., 2010) has received attention in the literature (e.g., Colins & Andershed, 2016; Colins, Noom, & Vanderplasschen, 2012; Fossati et al., 2016). This 18-item short-form is strongly correlated with scores on the full-length inventory and possesses similar psychometric properties. Although short-forms may suffer from some psychometric limitations (e.g., attenuated validity coefficients) relative to their parent scales (see Credé, Harms, Niehorster, & Gaye-Valentine, 2012; Smith, McCarthy, & Anderson, 2000, for a review), instruments such as the YPI-S may be valuable for assessing psychopathic traits in contexts necessitating brevity (e.g., large-scale screening for psychopathic traits in community surveys), which are contexts that often raise concerns about inattentive responding. Of note, researchers have increasingly referenced the promising clinical utility of the YPI and YPI-S and begun to investigate the performance of these instruments in applied juvenile forensic contexts (Colins, 2016; Colins et al., 2017; Gillen, MacDougall, Forth, Barry, & Salekin, 2017). Perhaps due to this accelerating interest and knowledge base, the YPI-S has been incorporated into routine evaluations of juvenile offenders in applied settings (Vahl et al., 2014), suggesting the measure currently influences clinical decision-making in real-world contexts. Unfortunately, most of the items contained in the full-length YPI that were used in constructing the SPICIE were omitted in the development of the short-form, making it impossible to apply this inconsistency scale to the YPI-S. Therefore, the focus of the present research was to develop and evaluate a new scale from the 18 existing YPI-S items that effectively identifies protocols of questionable validity due to careless or

inconsistent responding. To accomplish this objective, we examined five archival data sets where the YPI or YPI-S was administered in English, German, Italian, and Dutch to culturally diverse samples of juvenile offenders and youth in the community.

### **Current Study**

To develop a viable inconsistency scale for the YPI-S, we first identified pairs of items showing high correspondence across our five samples. Next, we computed inconsistency scores for study participants by summing the absolute values of differences between paired items. For each of our samples, participant inconsistency scores were evaluated against those obtained when YPI-S responses were generated completely at random ( $N = 1,000$ ). Because careless responding would not necessarily manifest as entirely random response selection for many or most respondents, we also created comparison samples from our data sets in which 50% of participants' original responses to YPI-S items were randomly selected and replaced using a procedure described by Handel, Ben-Porath, Tellegen, and Archer (2010). We then examined the criterion validity of inconsistency scores with available indices of inattention and response style from other measures of psychological functioning. Finally, we used proposed cutoff scores for profile invalidity to divide participants into two groups and, for a number of theoretically relevant correlates, examined if and how associations with the YPI-S were affected by inconsistent responding.

### **Method**

This research relies on multiple archival samples to identify and cross-validate a new inconsistency scale for the YPI-S. Three of these samples were used in the initial construction of the aforementioned SPICIE (Penson et al., 2017), which was developed for the full 50-item version of the instrument. We restricted our analyses across all samples to only those 18 items included in the short-form.

#### **Sample 1: U.S. Juvenile Justice**

The first sample (Shook, Vaughn, Goodkind, & Johnson, 2011) consisted of juvenile offenders ( $N = 253$ ) from two gender-specific residential placement facilities in Western Pennsylvania operated by private nonprofit organizations. The majority of participants were male (60.1%) and African American (51.8%), with a mean age of 16.2 years ( $SD = 1.70$ ). One participant was excluded from analyses due to missing data. Participants received an explanation of the original study purposes prior to assenting (or consenting for those 18-19 years old) to participate. Trained graduate research assistants conducted one-on-one interviews with participating youth

using techniques from the computer-assisted survey interview. These procedures involved reading each question to the participant and providing supplementary response cards. Participants were administered the full-length YPI and were compensated with \$10 credited to their institutional account.

#### **Sample 2: U.S. Middle School**

The second sample (DeLisi et al., 2011; Vaughn et al., 2011) was composed of seventh- and eighth-grade students ( $N = 432$ ; 43% female) from ethnically diverse backgrounds (43% Hispanic; 40% African American). The 50-item YPI was administered to groups of students with reading assistance. The administrator read each statement and participants provided a response from among four options. For the present study, participants with any missing responses ( $n = 69$ ) were excluded from further analyses, resulting in a final sample of 363 students. No compensation was provided for participation.

#### **Sample 3: German Vocational Training School**

This sample included students ( $N = 339$ ) attending vocational training school in Germany who participated in a more extensive study investigating psychopathic traits in the community (Eisenbarth & Centifanti, 2017). Thirty-eight participants were excluded from analyses due to missing data. The resulting sample ( $N = 301$ ) was 18.6 years of age, on average ( $SD = 1.92$ ) and the majority identified German as their primary language (86%). Those not primarily using German described themselves as sufficiently proficient in the language to complete vocational courses and follow instructions at outside placements. Prior to attending vocational training school, most participants (93.7%) had completed 10 or more years of education. Data were collected in the classroom setting using paper-pencil questionnaires, including the German translation of the 50-item YPI. No compensation was provided for participating in this research.

#### **Sample 4: Italian Public High School**

Participants in this sample were adolescents ( $N = 1,761$ ) attending specialized and/or professional programs at public high schools in Italy (Fossati et al., 2016). The participants were primarily male (51.6%) and were 16.0 years of age, on average ( $SD = 1.45$ ). Twelve participants were excluded from analyses due to missing data. Following an explanation of the original study, participants gave written consent to participate, and parents of participants younger than 18 years additionally provided written informed consent to allow participation. Participants were administered the Italian versions of the YPI-S and other study questionnaires in a randomized order while attending class. The measures were administered by research assistants outside the presence of teachers and completed anonymously.

### Sample 5: Dutch Secondary School

This sample consisted of 740 adolescents (53.2% female) from two secondary schools in rural areas of the Netherlands (Hillege et al., 2010). Most participants were Dutch (88.4%) and the remaining represented diverse ethnic backgrounds (e.g., North African, Middle Eastern, East Asian). The mean age was 15.6 years ( $SD = .94$ ). Participants with missing data on the YPI-S ( $n = 93$ ) were excluded from further analyses, resulting in a final sample size of 647. The original study purpose and procedures were explained to participants, who provided written informed consent, and to their parents, who were given the opportunity to decline participation. Participants anonymously completed the 50-item YPI and other study questionnaires during class under the supervision of a teacher, who was provided with standardized instructions and a glossary containing explanations of various terms referenced in the questionnaires.

### Measures

In addition to the YPI/YPI-S, our archival samples contained various criterion measures that were of theoretical interest in relation to examining either (a) the convergent validity of our newly developed inconsistent responding scale or (b) the effects of purported inconsistent responding on criterion-related validity (e.g., the relationship between psychopathy and measures of externalizing behavior). Descriptive statistics for the following study measures are provided as online supplemental materials.

**Youth Psychopathic Traits Inventory—Short Version (van Baardewijk et al., 2010).** This abbreviated version of the YPI consists of 18 items measuring core psychopathic features in adolescence (see van Baardewijk et al., 2010, for a listing of the specific items). The structure of the YPI-S corresponds to the organization of the full-length inventory and is consistent with the three-factor model of psychopathy (Cooke & Michie, 2001). In addition to general psychopathy, three dimensions, each composed of six items, are assessed: Grandiose–Manipulative (G-M), Callous–Unemotional (C-U), and Impulsive–Irresponsible (I-I). Participants rated items on a 4-point scale from *does not apply at all* (1) to *applies very well* (4). For each sample, internal consistency estimates for the total score ( $\alpha = .77-.85$ ) and subscale scores (G-M mean interitem correlation [MIC] = .27-.42; C-U MIC = .23-.33; I-I MIC = .17-.26) were generally in a range ( $\alpha \geq .80$ , MIC  $\geq .15$ ) that would be considered acceptable (see, e.g., Clark & Watson, 1995; Nunnally & Bernstein, 1994), although as noted by Schmitt (1996) there is no “sacred” level of Cronbach’s alpha that is required for research purposes. Reliability estimates combining across samples ( $N = 3,312$ ), were in similar ranges (Total  $\alpha = .83$ ; G-M MIC = .40; C-U MIC = .31; I-I MIC = .24).

### Sample 1: U.S. Juvenile Justice

**Massachusetts Youth Screening Instrument–2 (MAYSI-2; Grisso & Barnum, 2000).** The MAYSI-2 is a 52-item self-report questionnaire designed to broadly screen for psychopathology and behavioral dysfunction in juvenile justice settings. Participants endorsed items dichotomously as *yes* (1) or *no* (0). For analyses concerning the effect of response inconsistency on validity coefficients, we examined two scales of the MAYSI-2 that conceptually and empirically co-occur with psychopathic features (Colins et al., 2017; Patrick, Fowles, & Krueger, 2009; Vahl et al., 2014). The Alcohol/Drug Use scale (MIC = .33) contains eight items assessing characteristics and consequences of substance use and the Angry-Irritable scale (MIC = .24) consists of nine items capturing the extent of cognitive, affective, and behavioral experiences of anger and frustration. Additionally, we examined a short index of distractibility and inattention developed by Penson et al. (2017) composed of three items from the MAYSI-2 (MIC = .22). Findings from principal components analysis suggested these items were representative of a common dimension (i.e., difficulties with concentration).

**Self-Report of Delinquency (SRD).** The SRD is based on a similar measure used in the National Youth Survey (Elliott, Huizinga, & Menard, 1989) and assesses a range of delinquent activities, including property, drug, and physically violent offenses. Participants responded to 14 items concerning the frequency of various offending behaviors in the 12 months preceding residential placement ( $\alpha = .77$ ). These items were rated on a 9-point scale from *never* (0) to *2-3 times a day* (8).

**Moral Disengagement Index.** The Moral Disengagement Index is a 15-item self-report questionnaire ( $\alpha = .77$ ) that encompasses mechanisms of moral disengagement, including advantageous comparison, displacement and diffusion of responsibility, and moral justification (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). Participants rated items on a 5-point scale from *strongly agree* (1) to *strongly disagree* (5) with lower scores indicating greater moral disengagement.

**Legal Cynicism (Sampson & Bartusch, 1998).** This scale contains five items (MIC = .24) assessing beliefs about the legitimacy of the legal system and compliance with societal norms (e.g., “It’s okay to do anything you want as long as you don’t hurt anyone”). Participants rated items on a 5-point scale from *strongly agree* (1) to *strongly disagree* (5) with lower scores indicating greater legal cynicism.

**Coping Strategies Inventory.** This 35-item self-report inventory assesses the use of adaptive and maladaptive coping

strategies in adolescent populations. Participants indicated the frequency of engagement in various coping strategies on a 5-point scale from *never* (0) to *most of the time* (4). This measure is derived from the Adolescent Coping Orientation for Problem Experiences (Patterson & McCubbin, 1987) and organizes patterns of coping into five domains. We focused on the five-item domain of “Acting Out” (MIC = .48) for this study based on the conceptual association between acting out coping behaviors (e.g., picking a fight, saying mean things) and psychopathic traits in youth.

### Sample 3: German Vocational Training School

*Psychopathic Personality Inventory–Revised* (Lilienfeld & Widows, 2005). Students in vocational training completed the German translation of the PPI-R (Alpers & Eisenbarth, 2008). The PPI-R is an extensively researched self-report measure of psychopathy appropriate for use in community and forensic settings. The questionnaire contains 154 items that are rated on a 4-point scale as *false* (1), *mostly false* (2), *mostly true* (3), or *true* (4). Two validity scales embedded in the PPI-R assess inconsistent responding using either 15 (IR-15) or 40 (IR-40) pairs of highly correlated items. Previous studies have demonstrated the utility of the IR-15 and IR-40 in capturing patterns of inattentive or random responding and identifying profile validity (e.g., Kelley et al., 2016; Lilienfeld & Widows, 2005; Nikolova, Hendry, Douglas, Edens, & Lilienfeld, 2012).

### Sample 4: Italian Public High School

*Self-Report of Delinquency Scale* (SRDS; Elliott & Ageton, 1980). The SRDS is a 40-item self-report questionnaire concerning the frequency of delinquent behaviors during the past year. The Italian translation of this measure was administered to a subgroup of adolescents ( $n = 868$ ). Participants endorsed items on a 6-point scale from *never* (1) to *20 times or more* (6). Composite scores were created by summing responses ( $\alpha = .92$ ).

*Reactive–Proactive Aggression Questionnaire* (RPQ; Raine et al., 2006). The RPQ is a 23-item self-report questionnaire of aggression in adolescents. Participants ( $n = 868$ ) were administered the Italian translation of the RPQ and rated items on a 3-point scale as *never* (0), *sometimes* (1), or *often* (2). The measure contains two separate but interrelated subscales ( $r = .56, p < .01$ ): Reactive Aggression (11 items;  $\alpha = .81$ ) and Proactive Aggression (12 items,  $\alpha = .83$ ).

*Big Five Inventory* (BFI; John & Srivastava, 1999). The BFI is a self-report questionnaire consisting of 44 items that represent the core personality traits of the Big Five domains: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (MICs = .21-.36). The

Italian translation of this measure was administered to a second subgroup of adolescents ( $n = 881$ ). Participants rated items on a 5-point scale from *disagree strongly* (1) to *agree strongly* (5). For study analyses, we focused on Extraversion, Agreeableness, and Conscientiousness as these are the primary dimensions representative of psychopathy using the five-factor model (Miller, Lynam, Widiger, & Leukefeld, 2001).

### Sample 5: Dutch Secondary School

*Interpersonal Checklist–Revised* (ICL-R; de Jong, van den Brink, & Jansma, 2000). The Dutch version of the ICL-R is a 160-item self-report questionnaire based on the circumplex model of interpersonal functioning ( $\alpha = .85$ ). Items are rated dichotomously as *yes* (1) or *no* (0) and are broadly organized around two axes characterizing personality in terms of dominance versus submissiveness and affiliation versus hostility.

*Drug Use Disorders Identification Test* (DUDIT; Berman, Bergman, Palmstierna, & Schlyter, 2002). The DUDIT consists of 11 items intended to screen for drug use and drug-related problems ( $\alpha = .92$ ). Participants self-reported the frequency of various experiences concerning drug use on a 3- or 5-point scale on the Dutch version of the instrument.

*Screening of Alcohol Use* (SAU; Hillege et al., 2010). The SAU is a screening instrument developed in the Dutch language to assess motivations for and frequency of alcohol use. The self-report questionnaire is composed of nine items rated on a 3- or 5-point scale (MIC = .08).

## Results

### Scale Derivation

Given the limited number of nonredundant pairings available across the 18 YPI-S items, we initially focused on identifying item pairs that demonstrated relatively high interitem correlations across Samples 1 to 3.<sup>1</sup> To avoid redundancies, items could be included in only one pair, reducing the maximum number of possible pairs to nine. There is no definitive standard concerning what constitutes a “meaningful” correlation in the creation of an inconsistency scale, although several studies (e.g., Mowle et al., 2017) have used a provisional criterion of  $\geq .35$  as a starting point for selecting potential item pairs. Because of the abbreviated length of the YPI-S, we identified potentially useful item pairs by simply rank ordering all interitem correlations from highest to lowest, rather than by applying a specific cutoff.

Table 1 lists the seven nonoverlapping item pairs that collectively demonstrated the highest interitem correlations. As

**Table 1.** YPI-S Item Pair Correlations and Mean Absolute Values of Differences Between Items.

Item pairs	Sample									
	1. U.S. Juvenile Justice		2. U.S. Middle School		3. German Vocational		4. Italian Youth		5. Dutch Adolescent	
	<i>r</i>	<i>M</i>	<i>r</i>	<i>M</i>	<i>r</i>	<i>M</i>	<i>r</i>	<i>M</i>	<i>r</i>	<i>M</i>
1. 12 and 25	.63	0.47	.52	0.57	.61	0.59	.37	0.58	.37	0.59
2. 14 and 38	.54	0.66	.45	0.61	.47	0.63	.48	0.42	.49	0.54
3. 15 and 20	.47	0.71	.42	0.61	.47	0.62	.44	0.57	.49	0.65
4. 17 and 44	.24	0.60	.30	0.72	.30	0.65	.24	0.45	.33	0.54
5. 18 and 32	.51	0.74	.51	0.56	.60	0.51	.49	0.67	.63	0.52
6. 19 and 41	.34	1.01	.27	0.75	.40	0.63	.42	0.62	.44	0.56
7. 39 and 45	.35	0.91	.34	0.70	.37	0.86	.27	0.74	.30	0.73

Note. YPI-S = Youth Psychopathic Traits Inventory–Short Form; *M* = mean absolute value of the difference between the two items; *r* = Pearson correlation coefficient.

can be seen, multiple pairs demonstrated fairly large inter-item correlations, often approximating or exceeding .50. The four items on the YPI-S not represented in these pairings generally demonstrated weak patterns of association with each other (e.g., *r*s < .20) and therefore could not contribute additional item pairs to the provisional inconsistency scale.<sup>2</sup>

Although interitem correlations are important to consider in the derivation of an inconsistency scale, further attention must be given to the mean differences between paired items in their rates of endorsement. Inconsistency scales are typically scored by summing the absolute values of the within-pair differences between item responses. Accordingly, the two items forming each pair should also have overall mean differences approaching zero. If, hypothetically, two items correlated at .75, but had appreciably disparate mean scores of 0.50 and 2.50, such a pair would contribute a considerable amount (on average) to the summed total inconsistency scale score. The variance tapped by this item pair would largely reflect genuine differences in the absolute magnitude of how individuals respond to the two items rather than indicate inconsistent responding across the two items.

Table 1 also includes columns reporting the absolute mean difference between item responses for each pair. As can be seen, mean differences between responses for paired items were mostly  $\leq .75$ , with almost none exceeding 1.00. (For comparison purposes, the mean absolute difference between randomly generated responses for two items using a 4-point rating scale would be approximately 1.25). Based on these findings, summing the absolute values of the differences between these paired items produces scale scores that are not strongly biased by large within-pair differences in the mean rates of item endorsement.

Next, within each sample, we calculated total scores as described for this new scale, which we labeled the Detection of Inconsistent Response Presentation (DIRP). Considering

**Table 2.** Performance of the DIRP Scale in Identifying Randomly Generated Data.

Sample	<i>t</i>	Cohen's <i>d</i>	AUC	SE
100% Random data				
1. U.S. Juvenile Justice	21.07	1.51	.85	.013
2. U.S. Middle School	30.01	1.74	.89	.010
3. German Vocational	26.03	1.80	.90	.011
4. Italian Youth	46.32	1.84	.90	.006
5. Dutch Adolescent	35.88	1.81	.89	.008
50% Random data				
1. U.S. Juvenile Justice	10.75	0.96	.75	.021
2. U.S. Middle School	12.43	0.92	.76	.018
3. German Vocational	12.40	1.03	.77	.019
4. Italian Youth	38.53	1.30	.82	.007
5. Dutch Adolescent	21.65	1.20	.81	.012

Note. AUC = area under the curve; SE = standard error; DIRP = Detection of Inconsistent Response Presentation scale. All *p* values < .001.

the possible range of DIRP scores (0-21), the means for total scores were relatively low across our five samples (U.S. Juvenile Justice, *M* = 5.09, *SD* = 2.39; U.S. Middle School, *M* = 4.52, *SD* = 2.44; German Vocational, *M* = 4.50, *SD* = 2.27; Italian Youth, *M* = 4.05, *SD* = 2.61; Dutch Adolescent, *M* = 4.13, *SD* = 2.66).

For comparison purposes, we randomly generated YPI-S responses for 1,000 cases, which produced significantly elevated scores on the DIRP (*M* = 8.77; *SD* = 2.49; see Table 2 for *t* values). We next conducted Receiver Operating Characteristic (ROC) analyses to determine the utility of the DIRP in accurately classifying cases as genuine or randomly generated. Here, the area under the ROC curve (AUC) value represents the probability that a case randomly selected from the set of computer-generated protocols would obtain a higher inconsistency score than a case randomly selected from the sample of actual participants.

**Table 3.** Classification Accuracy Across DIRP Cutoff Scores.

DIRP	Sensitivity (%)		Specificity (%)			
	100% Random Data	1. U.S. Juvenile Justice	2. U.S. Middle School	3. German Vocational	4. Italian Youth	5. Dutch Adolescents
≥6	90	58	71	72	73	72
≥7	82	73	81	83	82	83
≥8	70	87	90	90	89	89
≥9	53	92	94	94	94	94

Note. DIRP = Detection of Inconsistent Response Presentation scale. Values are rounded to the nearest whole percentage.

Although significant AUCs  $> .50$  indicate greater than chance classification accuracy, effects are generally considered “strong” at values  $> .70$  (Rice & Harris, 2005). Consistent with comparisons of mean differences, the DIRP strongly differentiated between the original and random sources of data (AUCs = .85-.90).

To further assess the performance of the DIRP, we randomly selected and replaced 50% of YPI-S item responses in the original data sets using a procedure described by Handel et al. (2010). This approach may better approximate participant inattentiveness and aligns with previous findings that when 40-50% of responses are randomly produced, scores on common inconsistency scales typically exceed the threshold for validity (Handel et al., 2010). On average, DIRP scores were significantly higher for samples with 50% random data relative to their original counterparts (see Table 2 for  $t$  values). We again conducted ROC analyses comparing the original data sets to those with 50% random responding. The resulting AUC values expectedly decreased relative to findings using fully random data for comparison; however, the range of values obtained (.75-.82) continued to suggest strong discrimination between genuine YPI-S responses and those containing partially random data.

### Classification Accuracy

Although global indicators of performance (e.g.,  $d$ , AUC) are important when evaluating the utility of a scale such as the DIRP, for practical applications it is necessary to consider the effectiveness of specific cutoff scores in identifying profile validity. Sensitivity and specificity rates for several potential DIRP cutoff scores are presented in Table 3 for randomly generated data and archival samples, respectively. Notably, a cutoff score of  $\geq 8$  corresponds to specificity rates approximating 90% (87-90%), with 70% of randomly generated YPI-S protocols correctly classified as problematically inconsistent.

This cutoff score may be advantageous when attempting to separate attentive from inattentive respondents given previous estimates that 10-12% of research participants engage in careless responding. That is, the 10-13% of “genuine”

participants in our samples exceeding a cutoff score of  $\geq 8$  for profile validity may actually represent accurate detections of responding with neglect to item content. Of course, depending on the context of administration and consequences of certain misclassifications, alternative cutoff scores may be preferable. For example, using a cutoff score of  $\geq 6$  would increase the percentage of random profiles accurately identified to 90%, although with considerably reduced specificity rates (58-73%).

### Criterion-Related Validity Coefficients for Inconsistent Versus Consistent Respondents

We next examined how the correlations between the three factors of the YPI-S were affected by inconsistency in responding using a combined sample of study participants ( $N = 3,297$ ). For this combined sample, the G-M factor moderately correlated with both the C-U ( $r = .47, p < .01$ ) and I-I ( $r = .44, p < .01$ ) dimensions. Similarly, C-U and I-I were correlated at  $r = .34, p < .01$ . When using DIRP cutoff scores of  $\geq 7$  and  $\geq 8$  for classification purposes, the strength of these relationships among the more consistent participants remained largely unchanged; however, sizable reductions in correlation coefficients were observed among participants exceeding these thresholds. Specifically, correlations among respondents with DIRP scores  $\geq 8$  ( $n = 354$ ) were most attenuated for G-M and C-U ( $r = .14, p < .01$ ), followed by C-U and I-I ( $r = .13, p < .05$ ), and then G-M and I-I ( $r = .26, p < .01$ ). Based on comparisons using the Fisher  $r$ -to- $z$  transformation, these coefficients were significantly different from the corresponding magnitudes found for more consistent respondents ( $ps < .01$ ). Equivalent patterns of significant differences ( $ps < .01$ ) were also observed when using a more conservative cutoff score of  $\geq 7$  ( $n = 613$ ; G-M and C-U  $r = .21, p < .01$ ; C-U and I-I  $r = .14, p < .01$ ; G-M and I-I  $r = .27, p < .01$ ).

Additionally, we examined the effect of inconsistent responding on associations between the YPI-S and several theoretically relevant criterion measures available from our archival samples, including substance use, delinquent conduct, aggression, psychopathic traits, and basic dimensions of general personality. Table 4 presents these correlations

**Table 4.** Profile Classification and YPI-S Correlations With External Measures.

	Correlation with YPI-S total						
	Total sample	DIRP < 7	DIRP ≥ 7	z	DIRP < 8	DIRP ≥ 8	z
Sample 1: U.S. Juvenile Justice	<i>n</i> = 252	<i>n</i> = 183	<i>n</i> = 69		<i>n</i> = 219	<i>n</i> = 33	
MAYSI-2 Alcohol/Drug Use	.27**	.30**	.23	0.52	.27**	.32	-0.28
MAYSI-2 Angry-Irritable	.31**	.35**	.08	1.98*	.32**	.19	0.71
Self-Report Of Delinquency	.34**	.35**	.22	0.99	.32**	.20	0.66
Moral Disengagement	-.42**	-.48**	-.30*	-1.48	-.42**	-.28	-0.82
Legal Cynicism	-.45**	-.43**	-.19	-1.86*	-.43**	-.23	-1.16
Acting Out Coping	.41**	.44**	.14	2.30*	.43**	.03	2.21*
Sample 3: German Vocational	<i>n</i> = 282	<i>n</i> = 237	<i>n</i> = 45		<i>n</i> = 256	<i>n</i> = 26	
PPI-R Total	.71**	.73**	.45**	2.65**	.71**	.50**	1.55
PPI-R FD	.28**	.25**	.17	0.50	.26**	.19	0.34
PPI-R SCI	.64**	.67**	.39**	2.38**	.65**	.41*	1.56
PPI-R CH	.32**	.32**	.23	0.58	.30**	.33	-0.15
PPI-R IR-15	-.04	-.12	.08	-1.20	-.09	.07	-0.74
PPI-R IR-40	.08	.02	.06	-0.24	.04	.11	-0.32
Sample 4: Italian ( <i>n</i> = 868)—Males	<i>n</i> = 285	<i>n</i> = 230	<i>n</i> = 55		<i>n</i> = 256	<i>n</i> = 29	
Self-Report of Delinquency	.52**	.40**	.49**	-0.73	.38**	.62**	-1.58
RPQ-R	.49**	.45**	.44**	0.08	.45**	.49**	-0.25
RPQ-P	.48**	.41**	.40**	0.08	.39**	.46*	-0.42
Sample 4: Italian ( <i>n</i> = 868)—Females	<i>n</i> = 583	<i>n</i> = 515	<i>n</i> = 68		<i>n</i> = 542	<i>n</i> = 41	
Self-Report of Delinquency	.52**	.48**	.19	2.51**	.47**	.18	1.95*
RPQ-R	.55**	.46**	.31*	1.34	.44**	.42**	0.15
RPQ-P	.52**	.45**	.43**	0.19	.47**	.45**	0.15
Sample 4: Italian ( <i>n</i> = 881)—Males	<i>n</i> = 618	<i>n</i> = 470	<i>n</i> = 148		<i>n</i> = 526	<i>n</i> = 92	
BFI Agreeableness	-.26**	-.26**	-.30**	0.46	-.27**	-.29**	0.19
BFI Conscientiousness	-.17**	-.18**	-.13	-0.54	-.18**	-.22*	0.36
BFI Extraversion	.24**	.30**	.16*	1.56	.29**	.27**	0.19
Sample 4: Italian ( <i>n</i> = 881)—Females	<i>n</i> = 263	<i>n</i> = 217	<i>n</i> = 46		<i>n</i> = 238	<i>n</i> = 25	
BFI Agreeableness	-.30**	-.30**	-.32*	0.13	-.29**	-.27	-0.10
BFI Conscientiousness	-.20**	-.17*	-.22	0.31	-.16*	-.46*	1.51
BFI Extraversion	.26**	.28**	.38**	-0.67	.27**	.39	-0.61
Sample 5: Dutch Males	<i>n</i> = 298	<i>n</i> = 225	<i>n</i> = 73		<i>n</i> = 254	<i>n</i> = 44	
ICL-R Dominance	.23**	.26**	.14	0.91	.26**	.09	1.04
ICL-R Affiliation	-.23**	-.29**	.02	-2.32*	-.25**	-.01	-1.46
DUDIT Total	.24**	.26**	.18	0.61	.22**	.27	-0.32
SAU	.23**	.18**	.27*	-0.69	.19**	.25	-0.37
Sample 5: Dutch Females	<i>n</i> = 340	<i>n</i> = 304	<i>n</i> = 36		<i>n</i> = 316	<i>n</i> = 24	
ICL-R Dominance	.24**	.21**	.19	0.11	.22**	.23	-0.05
ICL-R Affiliation	-.24**	-.19**	-.08	-0.61	-.21**	.11	-1.44
DUDIT Total	.25**	.12*	.34	-1.27	.28*	-.03	1.41
SAU	.25**	.23**	.10	0.73	.22*	.11	0.50

Note. YPI-S = Youth Psychopathic Traits Inventory–Short Form; DIRP = Detection of Inconsistent Response Presentation Scale; BFI = Big Five Inventory; ICL-R = Interpersonal Checklist–Revised; DUDIT = Drug Use Disorders Identification Test; SAU = Screening of Alcohol Use; MAYSI-2 = Massachusetts Youth Screening Instrument–2; PPI-R = Psychopathic Traits Inventory–Revised; FD = Fearless Dominance; SCI = Self-Centered Impulsivity; CH = Coldheartedness; RPQ-R = Reactive Aggression; RPQ-P = Proactive Aggression.

\**p* < .05. \*\**p* < .01.

for total samples and for subgroups of participants separated by profile validity based on DIRP cutoff scores we previously identified as demonstrating favorable sensitivity and specificity (i.e.,  $\geq 7$  and  $\geq 8$ ). The Italian and Dutch

samples were sufficiently large enough to report findings separately for male and female participants, given the potential for these relationships to vary by gender. Generally, associations among “consistent” participants only slightly

deviated from criterion-related validity coefficients observed for the total sample, at most increasing the strength of the correlation by .08 when DIRP < 7 and .05 when DIRP < 8 ( $M = -0.01$ ,  $SD = 0.05$ ;  $M = -0.02$ ,  $SD = 0.04$ , respectively). In contrast, many of the associations were weakened for participants exceeding DIRP cutoff scores. In some instances, inconsistent responding affected the significance and/or direction of association compared with correlation coefficients for consistent respondents; however, differences in the magnitude of association between these two subgroups infrequently reached statistical significance using Fisher's  $r$ -to- $z$  transformation (Table 4; DIRP  $\geq 7 = 21.9\%$  of comparisons were significant at  $p < .05$ ; DIRP  $\geq 8 = 6.3\%$ ). The attenuation of criterion-related validity coefficients for "inconsistent" participants was particularly consequential among juvenile offenders from the United States and youth from the Netherlands. For example, the moderate association between psychopathic traits self-reported on the YPI-S and endorsement of acting out coping strategies observed for the total sample of juvenile offenders ( $r = .41$ ) was nonexistent among participants scoring  $\geq 8$  on the DIRP ( $r = .03$ ). For our samples of Italian adolescents, however, the impact of inconsistent responding on the predictive validity of the YPI-S was not as evident.

### Convergent Validity With External Measures of Inattention

We further investigated the validity of the DIRP in relation to external measures of attention. First, we examined the association between the DIRP and the brief measure of concentration difficulties developed from three items on the MAYSI-2 (Penson et al., 2017). Among juvenile offenders, DIRP scores were not significantly correlated with this index of inattentiveness ( $r = .09$ ,  $p = .17$ ). Second, we examined the convergence of DIRP scores with the embedded inconsistency scales of the PPI-R among German youth completing both measures ( $n = 286$ ). Consistent with previous findings for the SPICIE, DIRP scores were modestly associated with the IR-15 ( $r = .15$ ,  $p = .01$ ) and the IR-40 ( $r = .22$ ,  $p < .01$ ).

Embedded inconsistency scales developed for measures of psychopathy tend to demonstrate small to moderate relationships with self-reported psychopathic traits (e.g., Mowle et al., 2017; Penson et al., 2017). We similarly found a significant association between DIRP and YPI-S total scores, ( $r = .47$ ,  $p < .01$ ) for our combined sample of study participants ( $N = 3,297$ ). Notably, the magnitude of this association was substantially reduced for those with DIRP scores exceeding potential cutoff scores for profile validity (DIRP  $\geq 7$   $r = .19$ ,  $p < .05$ ; DIRP  $\geq 8$   $r = .11$ ,  $p < .05$ ). The inconsistent responding scales of the PPI-R, however, did not significantly correlate with YPI-S scores, regardless of standing on the DIRP (Table 4).

## Discussion

The current study developed an embedded scale for detecting inconsistent responding on the YPI-S, an abbreviated self-report measure of psychopathy in youth. The YPI-S has increasingly been administered in a variety of contexts in which there are likely to be concerns about response validity, including applied juvenile justice settings (Vahl et al., 2014). Accordingly, the ability to identify problematic completion of the measure due to poor motivation, distraction, or other cognitive factors has important implications for the fidelity of research using the instrument as well as the utility of scores for individual evaluations. Using five archival data sets containing English, German, Italian, and Dutch versions of the YPI-S, we found supportive evidence that our newly constructed inconsistent responding scale is sensitive to such a response style on this measure.

First, scores on the DIRP effectively distinguished original participant responses from randomly generated cases and from cases in which 50% of participant responses were randomly selected and replaced. Second, these scores demonstrated modest associations with the inconsistent responding scales of the PPI-R, although not with a brief measure of concentration difficulties extracted from the MAYSI-2. Third, the magnitude of associations among the three YPI-S factors was reduced significantly for participants identified as excessively inconsistent using two possible DIRP cutoff scores. Last, the strength of associations between the YPI-S and numerous theoretically relevant correlates was noticeably attenuated among participants exceeding proposed thresholds for profile validity. The adverse effect of inconsistent responding on criterion-related validity coefficients was particularly apparent among juvenile offenders from the United States and youth from the Netherlands, yet was not readily discernible for Italian participants in predicting personality, delinquency, and aggression from the YPI-S. The relative stability of associations among Italian participants classified as inconsistent respondents may be attributable to the tendency for erratic responding to create an appearance of greater pathology on self-report questionnaires (e.g., Benning & Freeman, 2017; Keeley et al., 2016), elevating scores on the YPI-S and correspondingly on criterion measures. However, greater mean scores on outcomes of interest among inconsistent participants were not universally observed for every measure (e.g., BFI) and co-occurred with effect size attenuation under certain circumstances (e.g., SRDS for Italian females and U.S. juvenile offenders). Although we cannot explain these variations in findings with certainty, the importance of inconsistent responding for criterion-related validity was typically supported across different measures and participant characteristics.

The utility of the DIRP may not be readily discernible by solely examining the effect of removing inconsistent respondents on criterion-related validity, as this screening

process did not significantly improve associations with study variables. This may also be because the effect of careless responding is uncertain, sometimes strengthening validity coefficients, sometimes weakening them, and sometimes appearing negligible when these opposing effects counteract (Credé, 2010; Curran, 2016). In fact, excluding problematic respondents commonly appears to make no difference, particularly for large effects; regardless, noise and error are reduced in important ways. For instance, detecting careless responding may be advantageous for research concentrating on mean scores or extreme groupings, as inconsistent participants appear to endorse more pathology, on average (e.g., Benning & Freeman, 2017; Keeley et al., 2016).

One additional finding meriting further consideration is the moderate correlation between DIRP and YPI-S total scores. This may suggest an association between psychopathic traits and tendencies toward responding carelessly; however, the covariation of scores might also be explained by the structure of the DIRP and the distribution of psychopathy scores in our samples. Theoretically, DIRP and YPI-S scores exhibit a curvilinear relationship, reflecting that participants consistently denying or consistently endorsing psychopathic traits would generally obtain lower scores on the DIRP. Contrastingly, participants self-reporting moderate levels of psychopathy are more likely to respond inconsistently by endorsing some statements more strongly than others, as even highly correlated item pairs may contain important distinctions in severity and/or content. Because few participants from our samples presented as highly psychopathic, however, the restricted range of YPI-S scores may create a potentially misleading appearance of a monotonic function. Importantly, this association is significantly weakened among participants exceeding DIRP cutoff scores for profile validity, indicating that a broad range of YPI-S scores can produce the same problematic degree of inconsistent responding. In summary, use of the DIRP should not automatically misclassify genuinely psychopathic participants as inconsistent responders.

This measure of response inconsistency represents one of many techniques for identifying inattentive respondents, and would optimally be used as part of an iterative screening process (Curran, 2016; Wood et al., 2017). For example, Wood et al. (2017) advocate for a two-step system that removes participants responding unrealistically rapidly and those demonstrating excessive inconsistency. Despite the association between response speed and consistency, there are undoubtedly times when someone trying to complete the measure as quickly as possible appears consistent (e.g., fixed responding) and times when someone answering items carelessly does so rather slowly. The use of multiple screening approaches with more conservative cutoff scores for invalidity minimizes the danger of excluding valid data

while supporting greater detection of clearly problematic respondents, and is not overly burdensome.

What actions follow identification of inconsistent responding? Because careless responding threatens the fidelity of research in numerous ways, as previously described, we would argue that these participants ultimately should be removed from the data set. Current recommendations urge analysis of the data before and after cleaning to provide a comparison of findings (Curran, 2016; Wood et al., 2017). Researchers may also wish to preregister their analytic plans and decision rules for identifying and handling inconsistent responding to reduce the possibility of selecting procedures based on the results of different approaches. These preliminary investigations can contribute key knowledge about the effect of inconsistent responding on analyses, as might reexamination of archival YPI-S data sets using the DIRP, particularly when previously reported findings were contrary to expectations. For clinical purposes, evaluators may address problematic response style by readministering the measure, when appropriate, or by augmenting the battery with another measure that is expected to provide generally convergent information.

Of note, inconsistency scores on the YPI-S provide no information outright concerning the source or pervasiveness of inconsistent responding across different parts of a survey or assessment battery. Elevations on the DIRP certainly raise concerns about the validity of other instruments from the same administration; however, these scores are not necessarily indicative of problematic responding beyond the YPI-S because respondent attention and motivation can vary throughout a testing session. When conducting an evaluation, closer inspection of examinee responses may uncover a specific source of inconsistency (e.g., poor reading or language comprehension, response sheet misalignment) from which to gauge the pervasiveness of invalidity and guide adjustments to further assessments. Decisions about interpreting other measures and continuing testing with or without modifications may also benefit from providing feedback to the examinee, inquiring about possible causes of inconsistency, and discussing with the examinee the importance of providing attentive and honest responses.

Collectively, results from our validation approaches support the utility of the DIRP for detecting excessive inconsistent responding across multiple translations of the YPI-S (English, German, Italian, and Dutch), differing participant characteristics (juvenile offenders, adolescent students), and varying administration procedures. In addition, the psychometric properties observed for the DIRP parallel findings from accumulating research on the development of inconsistency scales for measures of psychopathic traits using comparable methods and criteria for item pair selection (e.g., Kelley et al., 2016; Mowle et al., 2017; Penson et al., 2017). Despite these strengths, further study is needed to address the performance of the DIRP under more diverse

conditions. This may include, for example, investigation of the generalizability of our results to other languages and cultures, particularly those outside the United States and Western Europe. The utility of the DIRP among psychiatric populations and not yet adjudicated juvenile offenders requires future examination, as does the replicability of findings using different administration procedures (e.g., computer-administered).

In particular, the performance of the DIRP should be more broadly examined for administrations of the YPI-S specifically, as opposed to extracting these items for analysis from the full-length measure. We did not identify differences unique to our sample of Italian adolescents to suggest that administration versus extraction of the YPI-S affects the rate of inconsistent responding or the utility of the DIRP, although these comparisons are confounded by culture, language, and other administration procedures. To determine the potential for any differences with greater certainty, future studies must directly compare these two possible sources of the YPI-S/DIRP using more equivalent samples. Relatedly, side-by-side investigation of the YPI-S/DIRP and YPI/SPICIE may provide important insights concerning the influence of survey length on the presence and detection of response inconsistency. Although there is accumulating research on the prevalence and identification of inconsistent responding, few studies have investigated antecedents of this response style, for example, by counterbalancing lengthy survey batteries or comparing short- and long-form versions to assess the impact of fatigue and boredom, particularly among adolescent samples.

Other methodological issues to highlight include our removal of participants with missing responses on the YPI-S prior to analyses, which raises the possibility that some inconsistent respondents were excluded in advance when inattentive completion of the measure contributed to missing data. Under circumstances of missing data, problematic responding may not be detectable using the DIRP, as this brief scale is not amenable to prorating and the use of other techniques (e.g., response speed) may be necessary to assess the validity of scores. Also, because we relied on cross-sectional data, our findings cannot comment on the stability of response style and validity classification over time, or how well scores on the DIRP will correlate with criterion measures given during a separate administration. Finally, the validity of the DIRP may be further substantiated by evaluating the extent to which these inconsistency scores relate to indicators of response distortion or inattentiveness using methods beyond self-report.

### Authors' Note

M. Brent Donnellan is now at Michigan State University, East Lansing, MI, USA.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### Supplementary Material

Supplementary material is available for this article online.

### Notes

1. These item pairs subsequently were examined in Samples 4 and 5 to ensure that their performance was comparable to what was observed in Samples 1 to 3 (see Table 1 for full results across all 5 samples).
2. Complete correlation matrices are available from the second author on request.

### References

- Alpers, G. W., & Eisenbarth, H. (2008). *Psychopathic Personality Inventory—Revised (PPI-R)*. Göttingen, Germany: Hogrefe.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Andershed, H., Kerr, M., Stattin, H., & Levander, S. (2002). Psychopathic traits in non-referred youths: Initial test of a new assessment tool. In E. Blaauw & L. Sheridan (Eds.), *Psychopaths: Current international perspectives* (pp. 131-158). The Hague, Netherlands: Elsevier.
- 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
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71*, 364-374. doi:10.1037/0022-3514.71.2.364
- Benning, S. D., & Freeman, A. J. (2017). The Multidimensional Personality Questionnaire's inconsistency scales identify invalid profiles through internal statistics and external correlates. *Psychological Assessment*. Advance online publication. doi:10.1037/pas0000452
- Berman, A. H., Bergman, H., Palmstierna, T., & Schlyter, F. (2002). Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a Swedish population sample. *European Addict Research, 11*, 22-31. doi:10.1159/000081413
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press
- Campbell, M. A., Doucette, N. L., & French, S. (2009). Validity and stability of the Youth Psychopathic Traits Inventory in a nonforensic sample of young adults. *Journal of Personality Assessment, 91*, 584-592. doi:10.1080/00223890903228679

- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319. doi:10.1037/1040-3590.7.3.309
- Colins, O. F. (2016). The clinical usefulness of the *DSM-5* specifier for conduct disorder outside of a research context. *Law and Human Behavior, 40*, 310-318. doi:10.1037/lhb0000173
- Colins, O. F., & Andershed, H. (2015). The *DSM-5* with limited prosocial emotions specifier for conduct disorder among detained girls. *Law and Human Behavior, 39*, 198-207. doi:10.1037/lhb0000108
- Colins, O. F., & Andershed, H. (2016). The Youth Psychopathic Traits-Short Version in a general population sample of emerging adults. *Psychological Assessment, 28*, 449-457.
- Colins, O. F., Bijttebier, P., Broekaert, E., & Andershed, H. (2014). Psychopathic-like traits among detained female adolescents: Reliability and validity of the Antisocial Process Screening Device and the Youth Psychopathic Traits Inventory. *Assessment, 21*, 195-209. doi:10.1177/1073191113481997
- Colins, O. F., Fanti, K. A., Andershed, H., Mulder, E., Salekin, R. T., Blokland, A., & Vermeiren, R. R. J. M. (2017). Psychometric properties and prognostic usefulness of the Youth Psychopathic Traits Inventory (YPI) as a component of a clinical protocol for detained youth: A multiethnic examination. *Psychological Assessment, 29*, 740-753. doi:10.1037/pas0000437
- Colins, O. F., Noom, M., & Vanderplassen, W. (2012). Youth Psychopathic Traits Inventory-Short Version: A further test of the internal consistency and criterion validity. *Journal of Psychopathology and Behavioral Assessment, 34*, 476-486. doi:10.1007/s10862-012-9299-0
- Cook, N. E., Faust, D., Meyer, J. F., & Faust, K. A. (2016). The impact of careless and random responding on juvenile forensic assessment: Susceptibility of commonly used measures and implications for research and practice. *Journal of Forensic Psychology Practice, 5*, 425-447. doi:10.1080/15228932.2016.1234146
- Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment, 13*, 171-188. doi:10.1037/1040-3590.13.2.171
- Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment, 24*, 21-35. doi:10.1037/a0024824
- Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement, 70*, 596-612. doi:10.1177/0013164410366686
- Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology, 102*, 874-888. doi:10.1037/a0027403
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19. doi:10.1016/j.jesp.2015.07.006
- de Jong, C. A. J., van den Brink, W., & Jansma, A. (2000). *ICL-R: Handleiding bij de vernieuwde versie van de Interpersonal Checklist (ICL)* [Manual of the revised Dutch version of the Interpersonal Checklist (ICL)]. Sint-Oedenrode, Netherlands: Novadic.
- DeLisi, M., Vaughn, M. G., Beaver, K. M., Wexler, J., Barth, A., & Fletcher, J. (2011). Fledgling psychopathy in the classroom: ADHD subtypes, psychopathy, and educational outcomes in a community sample of adolescents. *Youth Violence and Juvenile Justice, 9*, 43-58. doi:10.1177/1541204010371932
- Dolan, M. C., & Rennie, C. E. (2007). The relationship between psychopathic traits measured by the Youth Psychopathic Trait Inventory and psychopathology in a UK sample of conduct disordered boys. *Journal of Adolescence, 30*, 601-611. doi:10.1016/j.adolescence.2006.07.002
- Edens, J. F., Mowle, E. N., Clark, J. W., & Magyar, M. S. (2016). "A psychopath by any other name?" Juror perceptions of the *DSM-5* "Limited Prosocial Emotions" specifier. *Journal of Personality Disorders, 31*, 90-109. doi:10.1521/pedi\_2016\_30\_239
- Eisenbarth, H., & Centifanti, L. M. (2017). *Dimensions of psychopathic traits in a community sample: Implications from different measures for impulsivity and delinquency*. Manuscript submitted for publication.
- Elliott, D. S., & Ageton, S. (1980). Reconciling race and class differences in self-reported and official estimates of delinquency. *American Sociological Review, 45*, 95-110.
- Elliott, D. S., Huizinga, D., & Menard, S. (1989). *Multiple problem youth: Delinquency, substance use, and mental health problems*. New York, NY: Springer.
- Faust, K. A., Faust, D., Baker, A. M., & Meyer, J. F. (2012). Refining video game use questionnaires for research and clinical application: Detection of problematic response sets. *International Journal of Mental Health and Addiction, 10*, 936-947. doi:10.1007/s11469-012-9390-5
- First, M. B., Spitzer, R. L., Gibbon, M., Williams, G. B. W., & Benjamin, L. (1997). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. Washington, DC: American Psychiatric Association.
- Fossati, A., Somma, A., Borroni, S., Frera, F., Maffei, C., & Andershed, H. (2016). The factor structure of the short version of the Youth Psychopathic Traits Inventory in two independent samples of nonreferred adolescents. *Assessment, 23*, 683-697. doi:10.1177/1073191115593628
- Gillen, C. T. A., MacDougall, E. A. M., Forth, A. E., Barry, C. T., & Salekin, R. T. (2017). Validity of the Youth Psychopathic Traits Inventory-Short Version in justice-involved and at-risk adolescents. *Assessment*. Advance online publication. doi:10.1177/1073191117700723
- Grisso, T., & Barnum, R. (2000). *Massachusetts Youth Screening Instrument-2 (MAYSI-2): User's manual and technical report*. Worcester: University of Massachusetts Medical School.
- Handel, R., Ben-Porath, Y. S., Tellegen, A., & Archer, R. P. (2010). Psychometric functioning of the MMPI-2-RF VRIN-r and TRIN-r scales with varying degrees of randomness, acquiescence, and counter-acquiescence. *Psychological Assessment, 22*, 87-95. doi:10.1037/a0017061
- Hare, R. D. (2003). *The Psychopathy Checklist-Revised (PCL-R)* (2nd ed.). Toronto, Ontario, Canada: Multi-Health Systems.
- Hillege, S., Das, J., & de Ruiter, C. (2010). The Youth Psychopathic Traits Inventory: Psychometric properties and its relation to substance use and interpersonal style in a Dutch sample of non-referred adolescents. *Journal of Adolescence, 33*, 83-91. doi:10.1016/j.adolescence.2009.05.006

- Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology, 100*, 828-845. doi:10.1037/a0038510
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York, NY: Guilford Press.
- Keeley, J. W., Webb, C., Peterson, D., Roussin, L., & Flanagan, E. H. (2016). Development of a response inconsistency scale for the Personality Inventory for DSM-5. *Journal of Personality Assessment, 98*, 351-359. doi:10.1080/00223891.2016.1158719
- 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
- Kelley, S. E., van Dongen, J. M., Donnellan, M. B., Edens, J. F., Eisenbarth, H., Fossati, A., . . . Sörman, K. (2017). Examination of the Triarchic Assessment Procedure for Inconsistent Responding in six non-English language samples. *Psychological Assessment*. Advance online publication. doi:10.1037/pas0000485
- Lilienfeld, S. O., & Widows, M. R. (2005). *Psychopathic Personality Inventory—Revised: Professional manual*. Lutz, FL: Psychological Assessment Resources.
- Marcus, D. K., Church, A. S., O'Connell, D., & Lilienfeld, S. O. (2016). Identifying careless responding with the Psychopathic Personality Inventory—Revised validity scales. *Assessment*. Advance online publication. doi:10.1177/1073191116641507
- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136*, 450-470. doi:10.1037/a0019216
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437-455. doi:10.1037/a0028085
- Miller, J. D., Lynam, D. R., Widiger, T. A., & Leukefeld, C. (2001). Personality disorders as extreme variants of common personality dimensions: Can the Five-Factor Model adequately represent psychopathy? *Journal of Personality, 69*, 253-276. doi:10.1111/1467-6494.00144
- Morey, L. C. (1991). *The Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Morey, L. C. (2012). Detection of response bias in applied assessment: Comment on McGrath et al. (2010). *Psychological Injury and Law, 5*, 153-161. doi:10.1007/s12207-012-9131-x
- Mowle, E. N., Kelley, S. E., Edens, J. F., Donnellan, M. B., Smith, S. T., Wygant, D. B., & Sellbom, M. (2017). Development of an inconsistent responding scale for the Triarchic Psychopathy Measure. *Psychological Assessment, 29*, 990-1000. doi:10.1037/pas0000395
- Nikolova, N. L., Hendry, M. C., Douglas, K. S., Edens, J. F., & Lilienfeld, S. O. (2012). The inconsistency of inconsistency scales: A comparison of two widely used measures. *Behavioral Sciences & the Law, 30*, 16-27. doi:10.1002/bsl.1996
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Oshukova, S., Kaltiala-Heino, R., Miettunen, J., Marttila, R., Tani, P., Aronen, E. T., . . . Lindberg, N. (2015). Self-reported psychopathic traits among non-referred Finnish adolescents: Psychometric properties of the Youth Psychopathic Traits Inventory and the Antisocial Process Screening Device. *Child and Adolescent Psychiatry and Mental Health, 9*, 1-11. doi:10.1186/s13034-015-0047-6
- Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and Psychopathology, 21*, 913-938. doi:10.1017/S0954579409000492
- Patterson, J. M., & McCubbin, H. I. (1987). Adolescent coping style and behaviors: Conceptualization and measurement. *Journal of Adolescence, 10*, 163-186. doi:10.1016/S0140-1971(87)80086-6
- Pechorro, P., Andershed, H., Ray, J. V., Maroco, J., & Gonçalves, R. A. (2015). Validation of the Youth Psychopathic Traits Inventory and Youth Psychopathic Traits Inventory—Short Version among incarcerated juvenile delinquents. *Journal of Psychopathology and Behavioral Assessment, 37*, 576-586. doi:10.1007/s10862-015-9490-1
- Penson, B. N., Ruchensky, J. R., Edens, J. F., Donnellan, M. B., Eisenbarth, H., & Vaughn, M. (2017). Development and initial validation of an inconsistent responding scale for the Youth Psychopathic Traits Inventory. *Journal of Personality Disorders*. Advance online publication. doi:10.1521/pedi\_2017\_31\_287
- Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78*, 582-593. doi:10.1037/0022-3514.78.3.582
- Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., . . . Liu, J. (2006). The Reactive-Proactive Aggression Questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior, 32*, 159-171. doi:10.1002/ab.20115
- 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. doi:10.1007/s10979-005-6832-7
- Sampson, R. J., & Bartusch, D. J. (1998). Legal cynicism and (subcultural?) tolerance of deviance: The neighborhood context of racial differences. *Law & Society Review, 32*, 777-804.
- Sellbom, M., Lilienfeld, S. O., Fowler, K., & McCrary, K. L. (in press). The self-report assessment of psychopathy: Challenges, pitfalls, and promises. In C. J. Patrick (Ed.), *Handbook of psychopathy* (2nd ed.). New York, NY: Guilford Press.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350-353. doi:10.1037/1040-3590.8.4.350
- Shook, J. J., Vaughn, M., Goodkind, S., & Johnson, H. (2011). An empirical portrait of youthful offenders who sell drug. *Journal of Criminal Justice, 39*, 224-231. doi:10.1016/j.jcrimjus.2011.02.014

- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102-111. doi:10.1037/1040-3590.12.1.102
- Vahl, P., Colins, O. F., Lodewijks, H. P. B., Markus, M. T., Doreleijers, T. A. H., & Vermeiren, R. R. J. M. (2014). Psychopathic-like traits in detained adolescents: Clinical usefulness of self-report. *European Child & Adolescent Psychiatry, 23*, 691-699. doi:10.1007/s00787-013-0497-4
- van Baardewijk, Y., Andershed, H., Stegge, H., Nilsson, K. W., Scholte, E., & Vermeiren, R. (2010). Development and tests of short versions of the Youth Psychopathic Traits Inventory and the Youth Psychopathic Traits Inventory-child version. *European Journal of Psychological Assessment, 26*, 122-128. doi:10.1027/1015-5759/a000017
- Vaughn, M. G., DeLisi, M., Beaver, K. M., Wexler, J., Barth, A., & Fletcher, J. (2011). Juvenile psychopathic features are associated with poor reading achievement. *Psychiatric Quarterly, 82*, 177-190. doi:10.1007/s11126-010-9162-y
- Vidal, S., & Skeem, J. L. (2007). Effect of psychopathy, abuse, and ethnicity on juvenile probation officers' decision-making and supervision strategies. *Law and Human Behavior, 31*, 479-498. doi:10.1007/s10979-006-9077-1
- Viljoen, J. L., McLachlan, K., & Vincent, G. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment, 17*, 377-395. doi:10.1177/1073191109359587
- Wood, D., Harms, P. D., Lowman, G. H., & DeSimone, J. A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science, 8*, 454-464. doi:10.1177/1948550617703168