SUBMITTED: August 31, 2016

Running head: CONSTRUCT VALIDITY

How well do juvenile risk assessments measure factors to target in treatment? Examining construct validity

Jennifer L. Skeem

University of California, Berkeley

jenskeem@berkeley.edu

Patrick Kennealy

Travis County Community Justice Services

Joseph Tatar, Wisconsin Department of Corrections

Isaias Hernandez, Parajo Valley Unified School Services

Felicia Keith, Uniformed Services University

Corresponding author: Jennifer Skeem, Professor of Social Welfare and Public Policy; University of California, Berkeley, 120 Haviland Hall #7400, Berkeley, CA 94720-7400

* This project was supported by a grant from the California Department of Corrections and Rehabilitation Division of Juvenile Justice (DJJ). We thank Stephanie Clark for her help with data collection and coding, Michael Galloway and Andre Le Banc for their help with references, and DJJ youth, staff, and administrators for their support.

Abstract

There has been a surge of interest in using one type of risk assessment instrument to tailor treatment to juveniles to reduce recidivism. Unlike prediction-oriented instruments, these reduction-oriented instruments explicitly measure variable risk factors as "needs" to be addressed in treatment. There is little evidence, however, that the instruments accurately measure specific risk factors. Based on a sample of 237 serious juvenile offenders (M age=18, SD=1.5), we tested whether California Youth Assessment Inventory (CA-YASI) scores validly assess the risk factors they purport to assess. Youth were assessed by practitioners with good interrater reliability on the CA-YASI, and by research staff on a battery of validated, multi-method criterion measures of target constructs. We meta-analytically tested whether each CA-YASI risk domain score (e.g., Attitudes) related more strongly to scores on convergent measures of theoretically similar constructs (e.g., criminal thinking styles) than to scores on discriminant measures of theoretically distinct constructs (e.g., intelligence, somatization, pubertal status). CA-YASI risk domain scores with the strongest validity support were those that assess criminal history. The only variable CA-YASI risk domain score that correlated more strongly with convergent (Z_r =.35) than discriminant (Z_r =.07) measures was Substance Use. There was little support for the construct validity of the remaining six variable CA-YASI risk domains—including those that ostensibly assess strong risk factors (e.g., "Attitudes," "Social Influence"). Our findings emphasize the need to test the construct validity of reduction-oriented instruments—and refine instruments to precisely measure their targets so they can truly inform risk reduction.

Public Significance: This study suggests that scores on a juvenile risk assessment instrument provide little direction for targeting treatment to reduce young people's risk of reoffending. With the exception of "Substance Abuse," we found little evidence that scores on ostensibly treatment-relevant scales measure the risk factors they purport to measure. The instrument must be refined, to support valid interpretations about risk factors to target in treatment.

How well do juvenile risk assessments measure factors to target in treatment?

Examining construct validity

Over recent years, there has been a surge of interest in the use of risk assessment to scaffold juvenile justice reform efforts (see Vincent, 2015). In jurisdictions throughout the U.S., risk assessment instruments may be used to divert low risk juveniles from incarceration (or from formal justice processing altogether) and to inform supervision and treatment efforts to reduce recidivism among higher risk youth. Today, juvenile probation departments in 34 states are implementing at least one risk assessment instrument (Wachter, 2015).

There are two major types of instruments. Prediction-oriented instruments assess only "risk" and are designed to efficiently and effectively characterize a juvenile's risk of recidivism, compared to other juveniles (e.g., low, medium, high). Reduction-oriented instruments assess risk and explicitly include variable risk factors as "needs" to be addressed in treatment to reduce the risk of recidivism (Monahan & Skeem, 2014). In juvenile justice, reduction-oriented instruments are appealing because they dovetail with the reform movement's recognition that young people are still developing and should be given opportunities for treatment and rehabilitation. According to the National Academy of Sciences (2013), "Assessing the risk of re-arrest *and the intervention needs* of each youth is the necessary first step in achieving the overall goal of a more rational and developmentally appropriate array of preventive interventions in the juvenile justice system" (pg. 5, emphasis added).

Policy and practice are far outpacing research, however, at the intersection between risk assessment and risk reduction (Monahan & Skeem, 2014; Skeem et al., 2013). Reduction-oriented instruments tend to be longer, more complex, less reliable, and harder to effectively implement than prediction-oriented instruments (see Baird et al., 2013; Skeem et al., 2013)—and there is little direct empirical evidence that they add value to prediction-oriented approaches (Monahan & Skeem, in press; Skeem et al., 2013).

Specifically, prediction- and reduction-oriented instruments perform equally well, with respect to predictive utility. Despite heated debate about their relative utility, there is no compelling evidence that one validated tool forecasts recidivism better than another (e.g., Olver et al., 2009; Yang et al., 2010). Predictive utility is the raison d'etre for prediction-oriented instruments (both necessary and sufficient)—but just a basic requirement for reduction-oriented instruments (necessary, but not sufficient). Reduction-oriented instruments are sold on the promise of going *beyond* prediction to inform the risk reduction enterprise—by explicitly assessing variable risk factors to specifically target in treatment (Vincent, 2015).

In this study, we test whether scores on a reduction-oriented instrument validly assess the risk factors they purportedly assess. This is one of several avenues for evaluating the value that a reduction-oriented instrument can add to prediction. For prediction-oriented instruments, Gottfredson and Moriarty have argued that any predictive variable can be used, barring ethical or legal challenge (cf. Hendry et al., 2013). But reduction-oriented instruments theoretically assess constructs that help *explain* the process that leads to recidivism. Specifically, they assess variable risk factors or "needs" like antisocial peer influence and attitudes that maintain criminal behavior. If interpretations derived from these instruments are valid, they can rationalize risk reduction efforts by specifying risk factors to target in treatment. But building a case plan to address misidentified "needs" is wasteful (at best) and harmful (at worst; Gatti, Tremblay, & Vitaro, 2009; Lipsey, 2009). If a young person does not actually have relevant substance abuse problems (despite an instrument's suggestion that s/he does), there is little reason to expect that substance abuse treatment will reduce reoffending.

We could identify only one study that directly examined the construct validity of a reduction-oriented risk assessment instrument. Based on a sample of 192 adult probationers, Andrews, Kiessling, Mickus and Robinson (1986) examined the pattern of correlations between risk factors

assessed by the Levels of Services Inventory (LSI) and a battery of 26 self-report scales that ostensibly assessed similar constructs. Some self-report scales were constructed specifically for this study; other scales were validated in prior research. The authors concluded that "the absolute level of convergent validity was not great, but on average, the convergent estimates exceeded the discriminant ones" (p. 467). Evidence for the validity of score interpretations varied by risk factor. For example, the LSI Alcohol/Drug scale was most strongly supported, given its strong correlation with scores on the convergent measure (r=.57) and weak-moderate associations with scores on the remaining self-report measures (Mr=.18, Range=.10-.32). LSI Antisocial Attitudes scores were weakly associated with scores on both its convergent measure (r=.27) and the remaining self-report measures (Mr=.18, Range=.15-.24).

This variation in validity evidence is important. It is critical to understand the extent to which scores on a particular scale validly assess the risk factors they purports to assess—i.e., whether the scale's label is meaningful. First, this understanding should determine how strongly a particular scale is weighted, in developing a case plan. Second, this understanding should inform efforts to improve the instruments themselves.

In the present study, we examine the construct validity of scores on the California Youth Assessment Screening Inventory (CA-YASI)—a version of the Youth Assessment and Screening Instrument (YASI; Orbis Partners, 2007) that was customized for California state's population of relatively serious juvenile offenders. Versions of the YASI are implemented in over 70 juvenile justice agencies throughout the US ((http://www.orbispartners.com). In independent studies, the CA-YASI (Skeem et al., 2013) and YASI (Baird et al., 2013) risk scores have been shown to predict recidivism about as strongly as scores on other validated instruments. In fact, largely on the basis of information about reliability and predictive utility, Vincent (2015) lists the YASI as one of seven promising or evidence-based instruments for youth.

Although CA-YASI risk scores predict youths' re-arrest (AUC=.66; Skeem et al., 2013), the question of construct validity looms large for this reduction-oriented instrument. This resource-intensive instrument requires staff to integrate interview- and file- information to rate over 100 items that assess twelve risk factors—a process that often requires over 2.5 hours per case. In an earlier study of 78 agency staff who completed four videotaped training cases, we found that only 59% of practitioners could reach "good" levels of agreement (ICC $\geq$ .60; Cicchetti & Sparrow, 1981) with experts' criterion CA-YASI Total scores (Kennealy et al., in press). At the subscale level, practitioners' accuracy was particularly weak for treatment-relevant factors that require substantial judgment (e.g., criminal attitudes, M ICC = .52)—but good for straightforward factors like criminal history (M ICC = .72). This raises an important question: When staff *can* score treatment-relevant factors reliably (i.e., in keeping with experts' correct scores), are those scores valid indicators of variable risk factors that can actually inform risk reduction efforts?

That is the question we take up in the present study. Because an "instrument that is not reliable cannot be valid…" (Latessa & Lovins, 2010, p. 212), we use only the majority of staff (59%) who demonstrated "good" levels of test score reliability on the CA-YASI in our earlier work. As such, this study presents "best case scenario" for field validity—it describes the validity of conclusions that can be drawn from the CA-YASI when it has been accurately scored. Our study has two specific aims:

1. To assess whether CA-YASI scores validly assess the risk factors they purport to assess. Using well-validated criterion measures that span methods (i.e., self-report, clinical rating, cognitive task, record coding), we test whether each risk domain score relates more strongly to scores on convergent measures of theoretically similar constructs than to scores on discriminant measures of theoretically distinct constructs. The strength of correlation between two variables is a function of similarity in constructs and method. If it is valid, CA-YASI domain scores will

correlate more strongly with scores on a measure of the same construct assessed with a different method, than with scores on a measure of a different construct assessed with the same method (e.g., the Attitudes domain will correlate more strongly with a validated self-report measure of criminal cognition than with a professionally-rated measure of substance abuse).

2. To explore the validity of risk/needs interpretations derived from the CA-YASI scale as a whole. Purpose-built risk assessment instruments are designed to predict recidivism rather than assess a particular construct. Nevertheless, CA-YASI Total Scores should be more strongly associated with scores on a well-validated measure commonly used to assess risk than with scores on well-validated measures of different constructs like intelligence.

Method

Study aims were addressed via two assessments of youth: a) a routine CA-YASI assessment completed by an institutional staff member who demonstrated good interrater reliability in an earlier study (Kennealy et al., in press), and (b) an assessment with over 30 well-validated criterion measures completed by research staff. To ensure that associations between the CA-YASI and criterion measures were not attenuated by change over time, criterion assessments were completed within two months of the CA-YASI assessment. Notably, all ICCs reported in this manuscript (and in Kennealy et al., in press) were calculated via a two-way mixed effect model with absolute agreement estimation and a focus on single rater coefficients.

Participants

Participants were 237 male youth incarcerated in state juvenile justice facilities (girls were excluded because there were too few girls in the agency's population to power separate analyses). Study ineligibility criteria included: (a) non-English speaking (n = 3; because most measures were validated in English), (b) older than 22 years (n = 23; because the evaluation focused on youth), and (c) transfer or discharge during the recruitment window (n= 54; because youth were unavailable).

Of eligible participants approached for recruitment (N=325), 27% of youth or their parents refused to participate. Compared to participants (N=237), youth who refused were ethnically similar but modestly more likely to be young, $d = .26$, $t\ (323) = 2.12$, $p < .05$. There were no significant differences between study participants and the institution's population in terms of age, ethnicity, and severity of index offense. Participants were an average of 18 years old ($SD = 1.57$) and most were ethnic minorities (56.1% Hispanic, 27.0% African American, 11.4% Caucasian and 5.5% other).

Practitioners

As noted earlier, participants were eligible for this study only if they had been scored on the CA-YASI by a staff member who demonstrated good test score reliability on the instrument in a prior study. Of the 78 staff members, 59% (n=54) had good reliability. Staff were men and women (57%) with advanced degrees (master's=48%; bachelor's=30%) with an average age of 43 (SD=7) and average number of years working in the agency of 11 (SD=8). Compared to unreliable staff, reliable staff tended to have fewer years of experience—other staff characteristics did not significantly moderate accuracy in scoring the CA-YASI (Kennealy et al., in press).

Procedure

Using IRB-approved procedures, research staff approached eligible youth at each facility and invited them to participate in the study. When a youth provided assent (in person) and his parent/guardian provided informed consent (typically via telephone), youth were enrolled in the study. Each participant completed a three-hour assessment that included a semi-structured interview, self-report measures, and performance tests (including computerized tasks). Due to agency policy, youth could not receive compensation for participation in this study. These assessments were conducted by research staff who had trained to reliability on all measures and procedures. After the assessment, interviewers reviewed youths' records to code information that was relevant to completing the criterion measures.

Measures

Given space limitations, we summarize essential features of the measures here. Detailed descriptions are available in an online supplement.

CA-YASI. The California Youth Assessment and Screening Instrument (CA-YASI; Orbis, 2007) is a 105-item risk and needs assessment that trained staff score on the basis of a semi-structured interview and file review. In the present study, these items were unit weighted and summed to calculate scores on twelve subscales or risk domains: legal history, correctional response, violence-aggression, social influences, substance use, attitudes, social-cognitive skills, family, education-employment, health, community linkages, and community stability (the last two domains are excluded from this study because no well-validated convergent measures could be identified). Definitions of each domain appear in Table 2.

In this study, we used unit-weighted scores rather than Orbis-weighted scores because our focus is on evaluating construct validity, i.e., whether the domains assess the risk factors they say they do. Orbis-weighted scores "neutralize" (i.e., effectively delete) many items to maximize the predictive utility of total scores. Nevertheless, the agency plans to retain all CA-YASI items because as a whole, they theoretically assess needs relevant to treatment and supervision. So our use of unit weights presents a "best case scenario" for the construct validity of CA-YASI scores, given that Orbis-weighted scores tend to relate more weakly to concurrent measures (see Skeem, Kennealy & Hernandez, 2013, Appendix A).

Reliability and predictive utility for the CA-YASI domain scores are acceptable. With respect to reliability, Table 2 displays each domain's (a) internal consistency in the present sample, and (b) interrater reliability *for the reliable subset* of staff selected for this study (see Kennealy et al., in press). With respect to predictive utility, based on a sample of 846 youth followed for at least one year, Skeem et al. (2013) found that Orbis-weighted scores on the CA-YASI moderately to strongly

predicted future institutional infractions (Any AUC=.65, Violent AUC=.75) and moderately predicted future arrests for any crime (AUC=.66) but not violent crime (AUC=.56). (Unit-weighted scores also significantly predicted most criterion measures.) These reliability and predictive utility estimates for CA-YASI scores are similar to those found in independent research on the YASI (see Baird et al., 2013 and Skeem, Barnoski, et al., 2013).

Correlations among the CA-YASI total- and domain- scores are shown in Table 1. As shown there, two groups of domain scores are so strongly inter-correlated that they arguably measure the same entity: (a) legal history and correctional response (both emphasize criminal history), and (b) violence-aggression, attitudes, social-cognitive skills, and social influences (all of which emphasize antisocial features and peers). Also, two domain scores (health and community linkages) are so weakly correlated with the rest of the CA-YASI that they seem independent of the scale.

Discriminant Measures. Discriminant measures largely were held constant across CA-YASI domains. These measures were chosen because they were well-validated, cross methods, and assess constructs that theoretically differ from those assessed by most CA-YASI domains—i.e., somatic distress, head injury, intelligence, and physical maturation. Brief descriptions of the measures and their psychometrics are provided below and in Table 3; detailed descriptions are available online.

Somatization. Current distress about perceived bodily dysfunction (e.g., dizziness, nausea) was assessed using the Somatization subscale of the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983). Unlike other BSI subscales, Somatization repeatedly emerges in factor analytic studies and manifests discriminant validity in capturing physical distress (see Skeem et al., 2006).

Head Injury. Prior experiences of significant head trauma were assessed using a four-item scale developed for the Pathways to Desistance study (Schubert, Mulvey, Steinberg, Cauffman, Losoya, et al., 2004). This measure is associated with theoretically relevant measures that include past

exposure to victimization/violence and current negative emotionality (Vaughn, Salas-Wright, DeLisi & Perron, 2014).

Intelligence. The Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) to estimate intelligence, using two subtests (Vocabulary and Matrix Reasoning). The WASI has high internal consistency (α=.84-.98) and test-retest reliability (*r*s=.87 to .92 across 2-12 week intervals)—and has been shown to strongly predict IQ estimates based on comprehensive intelligence tests (Wechsler, 1999). Intelligence was used as a divergent measure for all but two CA-YASI domains.[i]

Physical Maturation. The Pubertal Development Scale (PDS; Petersen, Crockett, Richards, & Boxer, 1988) was used to assess juvenile's physical maturation. Construct validity of the PDS has been supported in prior research by significant correlations (*r*s ranging from .61 to .67) with Tanner staging measurements via physician examinations (Schmitz et al., 2004).

Convergent Measures. Measures of convergent validity vary by CA-YASI domain, as described below. On the whole, we selected well-validated convergent measures of construct that was similar to, or the same as, a CA-YASI domain (see definitions in Table 2)—and we strove to include different methods. Brief descriptions of the measures are provided below and in Table 3 (format, reliability, etc.); details are available in an online supplement.

Violence-Aggression Domain: Social Information Processing (SIP). The SIP assesses social-cognitive problems that have been shown to robustly predict youths' aggression (Bradshaw, Rodgers, Ghandour & Garbino, 2009). Researchers score youths' responses to four vignettes to capture tendencies to perceive hostile intent in ambiguous situations ("Intent") and generate aggressive responses ("Response"). Based on 6 cases, our raters (n=5) manifested excellent interrater reliability in scoring Intent (ICC=.91) and Response (ICC=.87). The SIP also includes self report items that assess "Attitudes" supportive of aggressive behavior (alpha =.70). SIP scores relate to community violence exposure and predict aggressive behavior (Bradshaw et al., 2009).

Violence-Aggression Domain: Anger. The BSI (Derogatis & Melisaratos, 1983) hostility scale was used to assess Anger. Hostility repeatedly emerges in factor analytic studies of the BSI, selectively relates to other measures of anger, and predicts violence (see Skeem et al., 2006).

Violence-Aggression Domain: Meanness. The Meanness Inventory (MI;Patrick, 2010) assesses tendencies toward callousness, cruelty and predatory aggression. The scale has been shown to relate coherently to other measures of psychopathic traits (Patrick & Venables, 2010).

Multiple Domains: Social Deviance. Subscales of the Psychopathy Checklist: Youth Version (PCL:YV; Forth, Kosson, & Hare, 2003) were used to assess constructs relevant to several CA-YASI domains. We used the Social Deviance scale (Factor 2) as a convergent measure for the Aggression-Violence domain, given that it assesses past criminal behavior and broad traits like antagonism, anger, and poor behavior controls that are not specific to psychopathy, but place people at risk for violence (Skeem et al., 2011). The Social Deviance scale is divisible into two smaller scales: *Impulsive-Irresponsible Lifestyle* (Facet 3), and *Criminal Behavior* (Facet 4). We used Facet 3 as a convergent measure for the Education-Employment domain, given that it includes poor school/work achievement and motivation, including a lack of long-term goals. We used Facet 4 as a convergent measure for the Legal History and Correctional Response domains because it distills similar criminal behavior (see Table 2).

Before the study, interviewers (n=5) trained to reliability on the PCL:YV (i.e., PCL:YV Total Score, ICC $\geq$.80) using four videotaped cases. Based on ten study participants, levels of inter-rater reliability remained acceptable during the study (Total Score ICC=.81; for subscales, see Table 3). The Social Deviance scale has been shown to correlate with maladaptive characteristics and behaviors that include impulsivity, sensation seeking, substance abuse problems, criminal behavior, and aggression (see Skeem et al., 2001; Forth et al., 2003). In fact, Social Deviance accounts for most of the PCL:YV's utility in predicting future violence.

Attitudes Domain: Psychological Inventory of Criminal Thinking Styles (PICTS). The Proactive and Reactive scales of the PICTS (Walters, 1990) were combined to assess thinking styles considered essential to the maintenance of a criminal lifestyle (e.g., entitlement; rationalization; distress intolerance). The PICTS is strongly correlated with other measures of criminal thinking (Mandracchia & Morgan, 2011) and strongly predicts criminal recidivism (Walters, 2011).

Social-Cognitive Domain: Tower of London (ToL). The ToL (Berg & Byrd, 2002) is a neuropsychological test of executive function, i.e., cognitive processes that allow for self-regulation and socially appropriate behavior. The test involves moving colored beads across a grouping or rods to achieve a directed stacking pattern in few moves. We used the ToL to assess "Impulsivity" (time to first move) and problems with "Planning" or working memory (errors/excess moves)—which theoretically relate directly to the Social-Cognitive domain (see Table 2). ToL scores are associated with scores on other measures of cognitive flexibility and they predict antisocial behavior (Dolan, 2012; Ogilvie, Steward, Chan & Shum, 2011; Steinberg, 2010).

Social-Cognitive Domain: Response Inhibition. We used the Go/No Go discrimination task (GNG; Newman & Kosson, 1986) to assess response inhibition, an element of executive function relevant to punishment learning. The task challenges participants to earn points by deciding when to respond—or not respond—to images shown on the screen by pressing a key ("good" and "bad" images earn and lose points, respectively). Stimulus contingencies are changed during the task and the number of commission errors (failure to inhibit a response to a bad image) are calculated to reflect passive avoidance learning. Among measures of executive function, meta-analysis has found that Go/No Go scores relate relatively strongly to antisocial traits and behavior ($d = .44$; Ogilvie et al., 2011).

Social Influences Domain: Peer Delinquent Behavior (PDB). The PDB Scale (Thornberry, Lizotte, Krohn, Farnworth, & Jang, 1994) was used to assess engagement of peers in

antisocial behavior ("Behavior" scale) and efforts of peers to influence youth engagement in antisocial behavior ("Influence" scale). In large studies, the PDB has been shown to manifest a theoretically-coherent pattern of associations with age, susceptibility to peer influence, and neighborhood social factors—and to predict criminal behavior (Chung & Steinberg, 2006; Monahan, Steinberg, & Cauffman, 2009).

Social Influences Domain: Neighborhood Disorganization. The degree of neighborhood disorganization at the youth's most recent home address was measured using census tract data on poverty, unemployment, and cultural heterogeneity. Neighborhood disorganization theoretically relates to the availability of prosocial community role models and has been shown in a variety of studies to relate to violence and other criminal behavior (e.g., Elliott et al.,1996).

Family Domain: Family Background Questionnaire (FBQ). We used an adapted version of the FBQ (see McGee, Wolfe, & Wilson, 1997) to assess youths' familial exposure to maltreatment (both "Psychological" and "Physical") and to "Domestic Violence." Scores on the FBQ are associated with verified histories of incest, parental chemical dependency, clinical status, and socioeconomic status (Melchert, 1998).

Family Domain: Monitoring. We assessed exposure to poor parental monitoring and discipline with the Family Management Scale (FMS) of the Communities that Care Youth Survey (Arthur, Hawkins, Pollard, Catalano & Baglioni, 2002). The FMS has been shown to relate to youth substance use and delinquent behavior (Arthur et al., 2002).

Employment/Education Domain: School Connection. We used the School Connection Scale (SCS Brown, 1999) to assess school commitment and belongingness. The SCS has been shown to predict academic performance and extracurricular activity participation (Brown & Evans, 2002).

Employment/Education Domain: Irresponsible and Impulsive Lifestyle. Facet 3 of the PCL:SV was used to assess an Irresponsible and Impulsive Lifestyle (see above, "Social Deviance").

Substance Use Domain: Alcohol and Drug Dependence. We used the Substance Abuse Subtle Screening Inventory-A2 (SASSI-A2; Miller & Lazowski, 2005) to assess Alcohol Dependence (Face Valid Alcohol scale) and Other Drug Dependence (Face Valid Drug scale). These scales have been shown to predict clinical diagnoses of alcohol and drug dependence disorders with over 90% accuracy in both criminal and non-criminal samples (Miller & Lazowski, 2005).

(Mental) Health Domain: Psychological Distress. Total scores on the BSI (Derogatis & Melisaratos, 1983) were used to assess psychological distress or mental health problems. The BSI is strongly associated with other validated measures of mental illness and is sensitive to change in symptoms over time (for a review, see Skeem et al., 2006).

(Mental) Health Domain: Anxiety. We used total scores on the Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1985, 2000) to assess anxiety. The RCMAS is strongly associated with other measures of anxiety (Lee, Piersel, Friedlander, & Collamer, 1988; Reynolds, 1982) and has also been shown to discriminate between boys with anxiety disorders and those without psychiatric illness (Perrin & Last, 1992).

Legal History Domain: Criminal Behavior. Facet 4 of the PCL:YV was used to assess Criminal Behavior (see "Social Deviance" above for a description).

Legal History Domain: Conduct Disorder. We used Kim-Cohen, Moffitt, Taylor, Pawlby, & Caspi's (2005) self report measure of conduct disorder symptoms, which is strongly associated with other indices of delinquent behavior.

Correctional Response Domain: AWOL/Escape History. We coded youths' institutional files to record past escapes and absences without leave from institutional or community custody (base rate=9.3%).

Results

To determine whether CA-YASI scores assess the risk factors they purport to assess, we examined the convergent and discriminant validity of scores on these domains in two major analytic steps. First, we calculated correlations between each CA-YASI domain and scores on the *individual* measures of convergent and discriminant validity. This provides a fine-grained view of construct validity. Bi-variate results are reported by CA-YASI domain in Table 4. Second, to summarize the strength of association between each CA-YASI domain and scores on the *groups* of convergent versus divergent measures, we meta-analyzed the effect of each domain on those groups of external measures. To do so, we (a) converted each correlation into Fisher's Z_rs, (b) used the inverse variance approach to weight each effect size, and (c) calculated the mean effect size for groups of convergent and discriminant measures for each domain, using Lipsey and Wilson's (2001) general approach and SPSS macros. Because random effects models assume that effect sizes vary by population (Rosenthal & DiMatteo, 2001), and all effect sizes in this meta-analysis were drawn from the same sample, we report the fixed effects model here. Meta-analytic results are reported by domain in Table 5. We use Cohen's (1988) benchmarks for interpreting effect sizes: small or weak (r = 0.10, d=.20); medium or moderate (r= 0.30; d=.50), large or strong (r = 0.50, d=.80).

Violence-Aggression

As shown in Table 5, scores on the CA-YASI Violence-Aggression domain were weakly associated with scores on both the convergent (Z_r = .20) and discriminant measures (Z_r = .13), suggesting that interpretations for Violence-Aggression scores lack specificity. However, there was significant heterogeneity in effect sizes for the convergent measures. As shown in Table 4, the

Violence-Aggression scores were moderately associated with Social Deviance (PCL:SV) scores, not significantly associated with hostile attribution bias (SIP Intent) scores, and weakly associated with scores on the remaining convergent measures. Associations with discriminant measure scores were – as they should be – trivial or weak. But there was no significant difference in the strength with which Violence-Aggression scores correlated with low intelligence versus Social Deviance, Z= 1.05 (Lee & Preacher, 2008)– underscoring limited specificity.

Attitudes

Scores on the Attitudes domain were trivially associated with scores on convergent measures (Z_r = .09) and weakly associated with scores on the discriminant measures (Z_r = .14), although convergent effect sizes were heterogeneous (Table 5). Scores taken from this domain were weakly associated with Aggressive Attitudes/SIP scores and not significantly associated with scores on a well-validated convergent measure of criminal thinking (PICTS; Table 4). The strongest correlates for this domain were scores on the discriminant measures of intelligence and pubertal maturation—providing little or no validity support for the resulting risk score conclusions.

Social-Cognitive Skills

Social Cognitive domain scores were trivially associated with scores on both convergent (Z_r = .07) and discriminant measures (Z_r = .09; Table 5), with heterogeneity among convergent measures. Among convergent measures, Social-Cognitive scores were more associated with low intelligence and poor planning (TOL errors) scores than response inhibition (Go/No Go) or impulsivity (TOL; Table 4) scores. Shifting to discriminant measures, the Social Cognitive scores were associated only with low pubertal maturation scores. In light of the similar patterns of associations for convergent and discriminant measures, there is little evidence that scores on this domain capture risk factors for antisocial behavior like poor executive functioning.

Social Influences

Scores on the Social Influences domain were weakly associated with scores on both convergent (Z_r =.16) and discriminant (Z_r =.11) measures (Table 5), with heterogeneous convergent effect sizes. Among convergent measures, Social Influences scores related weakly to peers' antisocial behavior (PDB) and Neighborhood Disorganization scores, but were not significantly associated with peer influence scores (PDB; Table 4). Shifting to discriminant measures, Social Influences scores correlated weakly with low intelligence scores. Together, there is little evidence that scores on this domain specifically capture peer/social *influence* on antisocial behavior.

Family

Family domain scores weakly correlated with scores from both convergent (Z_r =.12) and discriminant (Z_r =.10) measures, and effects were homogeneous (Table 4). As shown in Table 3, the this scale related as strongly to convergent (Family Monitoring) as divergent (Somatization) measure scores, providing little support that the domain taps family-related risk factors for crime.

Education/Employment

As shown in Table 5, Education/Employment scores related weakly to both convergent (Z_r =.18) and discriminant (Z_r =.20) measure scores, with heterogeneity in divergent measures. Among the convergent measures, Education/Employment scores were trivially to weakly associated with scores on the main convergent measures (School Connection; Impulsive & Irresponsible Lifestyle/PCL; Table 4). Scores on this domain were unassociated or weakly associated with discriminant measure scores, with the exception of low pubertal status (notably, pubertal status was not associated with School Connection). In short, interpretations that this scale assesses Education/Employment problems will have poor specificity.

Substance Use

The Substance Use domain demonstrated a significantly stronger association with scores on the convergent measures (Z_r =.35) than those from the discriminant measures (Z_r =.07), as indicated

by the non-overlapping confidence intervals of these mean effect sizes (Table 5). Discriminant effects were heterogeneous. At the bivariate level (Table 4), scores from this domain were moderately associated with scores from both alcohol- and drug- problems scales assessed by the SASSI, a well-validated self-report measure. With the exception of low intelligence, Substance Use scores were trivially associated with remaining discriminant measure scores. Of CA-YASI scales examined thus far, evidence of construct validity is strongest for the Substance Abuse domain.

(Mental) Health

As noted earlier, (Mental) Health is not scored as a risk factor in the CA-YASI (likely because mental health problems are relatively weak risk factors for recidivism). Health scores were weakly associated with scores on the convergent measure (Z_r =.24) and trivially associated with those on the discriminant (Z_r =.06) measures—but this difference was significant (Table 5). Health scores were weakly- associated with indices of global psychological distress (BSI) and anxiety (RCMAS) and trivially associated with scores for all discriminant measures (see Table 4), providing modest support for the validity of Health score interpretations.

Legal History

Legal History scores correlated significantly more strongly with convergent (Z_r = .36) than discriminant measure scores (Z_r = .03), as indicated by the non-overlapping confidence intervals of these mean effect sizes (Table 5). Still, effect sizes for the convergent measures were heterogeneous. Of the convergent measures, Legal History scores were strongly associated with Criminal Behavior (PCL:YV) scores but only weakly associated with Conduct Disorder (Table 4) scores. In contrast, this scale was not significantly associated with scores on any discriminant measure (including intelligence). These results provide strong support for the validity of the Legal History scale.

Correctional Response

Correctional Response scores were significantly more strongly associated with convergent (Z_r = .45) than discriminant measure scores (Z_r = .09), as indicated by the non-overlapping confidence intervals (see Table 5). However, there was heterogeneity among effect sizes for the convergent measures. Among convergent measures, Correctional Response scores were strongly associated with Criminal Behavior (PCL:YV) scores and moderately associated with the more domain-specific measure of AWOL/Escape (Records). This provides limited support for the validity of the Correctional Response scale (as distinct from Legal History scores).

CA-YASI Total Scores

To contextualize domain-level results, we compared the degree of association between CA-YASI Total Scores and scores on (a) the PCL:YV Social Deviance scale, and (b) the discriminant validity measures. Social Deviance scores are appropriate measures of convergent validity for CA-YASI scores because (a) the PCL family of measures are often applied to assess risk of recidivism because they have been shown to robustly predict recidivism—chiefly as a function of the Social Deviance scale (see Skeem, Polaschek, Patrick & Lilienfeld, 2011), and (b) PCL:YV scores tend to correlate strongly with scores on other validated risk assessment instruments (e.g., Edens et al., 2007; Hilterman et al., 2013).

CA-YASI Total Scores were strongly associated with Social Deviance scores (Table 4, Z_r =.51, *SE*=.07) and weakly associated with scores on the discriminant measures (Table 5, Z_r =.18, *SE*= .03)—and non-overlapping confidence intervals indicate this difference is statistically significant. However, among the heterogeneous discriminant effects, CA-YASI Total scores were moderately correlated with low intelligence scores (Zr=.33)—at a level that does not differ significantly from the strength of association between CA-YASI Total scores and Social Deviance scores, Z = 1.71 (Lee & Preacher, 2008).

In summary, scores on the CA-YASI as a whole are strongly associated with scores on a validated measure of antisocial traits and behavior that has been shown to predict recidivism among youth. This provides some confidence in the convergent validity of CA-YASI risk score interpretations. However, the moderate association with low intelligence scores suggests that the tool may tap irrelevant variance. It seems unlikely that this association is completely explained by individual differences in motivation during IQ testing, which can spuriously inflate the association between scores on measures of intelligence and antisocial behavior (Duckworth et al., 2011). Specifically, IQ scores were significantly more strongly associated with CA-YASI scores (r = .33) than with Social Deviance scores (r = .18), Z = 2.10, p <.05. This underscores concern that conclusions about juveniles' risk of recidivism based on the CA-YASI tap some risk-irrelevant variance related to intelligence.

Discussion

This study is among the first to test whether scores on a reduction-oriented juvenile risk assessment instrument actually assess the risk factors they purport to assess. The results of this study represent a "best case scenario" for the field validity of the CA-YASI scales, in that we (a) included only the 59% of staff who could score the instrument accurately, and (b) used all of the instrument's items (rather than Orbis-weighted formulae that exclude certain items to maximize the measure's predictive utility; see *Method*). Our results provide little evidence that CA-YASI scores can add value to prediction (see Skeem et al., 2013) by validly assessing specific risk factors to target in treatment.

The CA-YASI domain with clearest evidence of construct validity—by far—was Legal History (and the overlapping domain of Correctional Response). It seems relatively safe to interpret scores on these domains as indicators of criminal history. Criminal behavior strongly predicts future criminal behavior, but provides little direction for risk reduction efforts. The only variable CA-

YASI domains that correlated more strongly with convergent- than divergent- measures were those that assess Substance Abuse and (Mental) Health—and the latter is not even scored as a risk domain (given its weak utility in predicting recidivism). There was little support for the validity of risk score interpretations that would be derived from the remaining domains—including those that ostensibly assess strong, treatment-relevant risk factors for recidivism (e.g., "Attitudes," "Social-Cognitive Skills," "Social Influence"; Andrews et al., 2006). These CA-YASI domain scores do not appear to accurately specify treatment targets for young people.

Limitations

Before contextualizing these findings, we note three potential study limitations—and one generalizability consideration—that must be considered before doing so. First, as may be true of other instruments, definitions for some CA-YASI domains are imprecise and/or multi-faceted (see Table 2). For example, the Violence-Aggression domain references past violence (reactive and predatory) and several violence-relevant traits and processes (e.g., anger, hostile attribution biases, aggressive attitudes, callousness). Imprecision in the target domain translates to multi-dimensional measures of convergent validity (see Table 4) and potentially heterogeneous effect sizes (Table 5). In our view, this problem mostly provides direction for instrument refinement, given evidence that improving a scale's precision in assessing a target construct improves its utility in predicting recidivism (Hendry et al., 2013). This problem is unlikely to explain this study's results, as even more specifically-defined CA-YASI domains (e.g., Attitudes) with clear convergent measures (e.g., PICTS) performed poorly in assessing their target constructs. Second, even though only the reliable subset of staff were included in this study, inter-rater reliability for two domain scores (i.e., Substance Use, Social-Cognitive Skills) was only "fair," which could constrain validity estimates. Nevertheless, one of these two domains (Substance Use) was the only variable domain score with evidence of construct validity in this study. Third, the internal consistency of scores on a minority

of criterion scales was limited—but again, this is unlikely to account for our findings. For example, there was no clear difference in the pattern of correlations between CA-YAS Aggression Violence scores and (a) the two criterion measures with weakest internal consistency in this study (SIP Intent and Response), and (b) the four remaining criterion measures with stronger internal consistency (see Tables 3 & 4). On the whole, this study seems to accurately estimate the capacity for the CA-YASI to yield valid risk score interpretations when unit-scored by reliable staff. Finally, as explained later, the extent to which our results will generalize from the CA-YASI to other reduction-oriented instruments is unclear and must be examined in future research.

The Challenge of Validly Assessing Variable Risk Factors

The CA-YASI domain scores with the strongest interrater reliability and validity support are those that assess criminal history (Legal History/Correctional Response). Moreover, scores from the instrument as a whole are strongly associated with scores on a well-validated measure of Social Deviance (PCL:YV Factor 2), which suggests that CA-YASI Total Scores overlap in distilling past criminal behavior and broad externalizing traits like antagonism, anger, and poor behavior controls that strongly predict recidivism (see Skeem et al., 2011). Similar findings may apply to most validated risk assessment instruments, given that criminal history is easy to reliably score (Baird et al., 2013) and scores on risk assessment instruments typically correlate strongly with one another and with PCL measures (e.g., Hilterman, Nicholls, & van Nieuwenhuizen, 2013; Olver et al., 2012; Stockdale, Olver & Wong, 2014).

But the CA-YASI is designed to go beyond distilling "risk as usual" (i.e., prediction), to capture treatment-relevant variable risk factors (i.e., reduction). We found little evidence it does so, with the important exception of Substance Use. Our finding that CA-YASI scores validly assess Substance Use is consistent with Andrews et al.'s (1986) observation that evidence of construct validity was strongest for the LSI's Alcohol/Drug scale score. These findings provide some

confidence that reduction-oriented instruments can help identify youth with substance abuse problems and make pertinent treatment referrals. Among the weakest "need" scales on both instruments, however, were Attitudes (CA-YASI) or Antisocial Attitudes (LSI). Poor assessment of criminal cognition is a serious limitation for reduction-oriented instruments, given that cognitive-behavioral treatment programs that explicitly target criminal cognition are among the most robustly supported evidence-based approaches for reducing recidivism (Lipsey, Landenberger, & Wilson, 2007).

As a group, CA-YASI variable risk domains performed more poorly in this study than the LSI variable risk scales performed in Andrews et al. (1986). Convergent associations, for example, tended to be moderate for the LSI scales—and weak for the CA-YASI scales. These differences may partly reflect differences in evaluation method (e.g., self-report vs. multi-modal criterion measures)—but probably also reflect differences in the ability of different scales to yield valid inferences about a particular construct. We found little evidence that CA-YASI domain scores labeled as Violence-Aggression, Social-Cognitive Skills, Social Influences, Education/Employment, and Family specifically translate to treatment-relevant target constructs of anger/hostility, executive function deficits, antisocial peer influence, poor school/work motivation, or problematic parental discipline and monitoring, respectively.

Implications

Research. Our findings emphasize the need to test how well reduction-oriented risk assessment instruments assess variable risk factors, as we did here with the CA-YASI. More broadly, our findings underscore the need to test the value that reduction-oriented instruments add to prediction-oriented instruments. There are several alternative avenues for doing so. A relatively undemanding avenue involves testing whether relevant scores on an instrument change over time, and whether those changes predict recidivism (see Dixon & Howard, 2013). Or one could

experimentally test whether youth are less likely to recidivate when professionals use a reduction-oriented rather than prediction-oriented instrument. The most rigorous—and treatment-relevant—test would be a randomized controlled trial in which a targeted intervention was shown to be effective in changing scores on a variable risk factor(s) on an instrument, and the resulting changes were shown to reduce the likelihood of post-treatment recidivism (see Monahan & Skeem, in press).

The results of such studies would directly inform instrument refinement to help reduction-oriented instruments precisely assess the variable risk factors they aim to assess. Achieving validity is an ongoing, challenging process that requires iterative improvements.

Practice. There is evidence that CA-YASI Total scores (and Violence-Aggression scores) identify youth with antisocial features who are relatively likely to recidivate (Skeem et al., 2013). These scores can be used to identify youth who should receive intensive supervision and services, according to the "risk" principle of effective correctional intervention (Andrews et al., 2006). Scores on the CA-YASI, however, have little capacity to inform the nature of services that youth should receive. Because most domain scores do not specifically assess the variable risk factors they purport to assess, CA-YASI score interpretations cannot help actualize the "need" principle of correctional intervention. Specifically, CA-YASI Substance Use (and Health) scores can be used to flag youth in need of substance abuse (and mental health) services. But there is little support for interpreting scores on the remaining variable domains – including those that ostensibly tap strong risk factors for recidivism (e.g., antisocial attitudes and peers; Andrews et al., 2006)– as specific indicators of treatment targets.

Policy. Variable risk factors require specific assessment only if there is a realistic likelihood that they subsequently will be addressed with pertinent treatment services (Monahan & Skeem, in press). Given resource limitations and implementation challenges, efficiency and simplicity are to be preferred. A prediction-oriented assessment is sufficient, if the goal is to divert all low risk youth

from the system and/or to provide all high risk youth with the same generic services. Assessment of specific variable risk factors may be added, if a specific type of treatment is available to some, but not to all, high risk offenders. For example, if an anger treatment program is available, it would be efficient and cost effective to use a well-validated measure of anger to refer youth to that program.

If a juvenile justice system is prepared to address most – or all- of the variable risk factors included in a reduction-oriented instrument, then it makes sense to administer that instrument in every case to inform appropriate case referrals. In some promising jurisdictions, there is sufficient momentum to meaningfully target services to offenders' risk factors (Vincent, 2015; see also Vincent et al., 2012). The field's next challenge is to ensure that reduction-oriented instruments can validly scaffold these efforts.

References

Achenbach, T. M. (1991, 2001). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile.* Burlington: University of Vermont, Department of Psychiatry.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency, 52*, 7-27.

Andrews, DA, Kiessling, JJ, Mickus, S, & Robinson, D (1986). The construct validity of interview-based risk assessment in corrections. *Canadian Journal of Behavioural Science,* 18, 460-471.

Arthur, M. W., Hawkins, J. D., Pollard, J. A., Catalano, R. F., & Baglioni Jr., A. J. (2002). Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: The Communities That Care Youth Survey. *Evaluation Review, 26,* 575-601.

Baird, C., Healy, T., Johnson, K, Bogie, A., Dankert, E., & Scharenbroch, C. (2013). *A comparison of risk assessment instruments in juvenile justice.* Madison, WI: National Council on Crime and Delinquency.

Berg, W., & Byrd, D. (2002). The Tower of London spatial problem solving task: Enhancing clinical and research implementation. *Journal of Experimental and Clinical Neuropsychology, 25,* 586–604.

Bostwick, L. (2010). *Mental health screening and assessment in the Illinois juvenile justice system.* Chicago, IL: Illinois Juvenile Justice Commisssion.

Bradshaw, C. P., Rodgers, C., Ghandour, L. & Garbarino, J. (2009). Social-cognitive mediators of the association between community violence exposure and aggressive behavior. *School Psychology Quarterly, 24,* 199-210.

Brown, R. (1999). *The influence of extracurricular activity participation upon youth problem behavior: School connection as a mediator.* Washington, DC: Education Resources Information Center. (Eric Document Reproduction Service No. ED430167).

Brown R., & Evans, W. P. (2002). Extracurricular activity and ethnicity: creating greater school connection among diverse student populations. *Urban Education, 37,* 263-286.

Chung, H. L., & Steinberg, L. (2006). Relations between neighborhood factors, parenting behaviors, peer deviance, and delinquency among serious juvenile offenders. *Developmental Psychology, 42,* 319-331.

Costello, A. J., Edelbrock, C., Kalas, R., Kessler, M. D. & Klaric, S. H. (1982). *The NIMH Diagnostic Interview Schedule for Children (DISC).* Unpublished interview schedule, Department of Psychiatry, University of Pittsburgh.

Derogatis, L., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*, 595-605.

Dolan, M. (2012). The neuropsychology of prefrontal function in antisocial personality disordered offenders with varying degrees of psychopathy. *Psychological Medicine*, *1*, 1-11.

Elliott, D. S., Wilson, W. J., Huizinga, D., Sampson, R. J., Elliott, A., & Rankin, B. (1996). The effects of neighborhood disadvantage on adolescent development. Journal of Research in *Crime and Delinquency, 33,* 389-426.

Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *The Psychopathy Checklist: Youth Version manual.* Toronto: Multi-Health Systems.

Gatti, U., Tremblay, R. E., & Vitaro, F. (2009). Iatrogenic effect of juvenile justice. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 50*, 991–998. http://dx.doi.org/10.1111/j.1469-7610.2008 .02057.x

Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime and Delinquency, 52(1),* 178–200.

Hendry, M. C., Douglas, K. S., Winter, E. A., & Edens, J. F. (2013). Construct measurement quality improves predictive accuracy in violence risk assessment: An illustration using the Personality Assessment Inventory. *Behavioral Sciences & the Law, 31*, 477-493.

Hilterman, E. L., Nicholls, T. L., & van Nieuwenhuizen, C. (2014). Predictive validity of risk assessments in juvenile offenders: Comparing the SAVRY, PCL: YV, and YLS/CMI with unstructured clinical assessments. *Assessment, 21*(3), 324-339.

Howard, P., & Dixon, L. (2013). Identifying change in the likelihood of violent recidivism: Causal dynamic risk factors in the OASys violence predictor. *Law & Human Behavior, 37*, 163-174.

Kennealy, P., Skeem, J., & Hernandez, I. (in press). Does staff see what experts see? Accuracy of frontline workers in scoring juveniles' risk factors. *Law & Human Behavior.*

Kim-Cohen, J., Moffitt, T. E., Taylor, A., Pawlby, S. J., & Caspi, A. (2005). Maternal depression and children's antisocial behavior: nature and nurture effects. *Archives of General Psychiatry, 62,* 173-181.

Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

Lee, S. W., Piersel, W. C., Friedlander, R., & Collamer, W. (1988). Convergent validity of Revised Children's Manifest Anxiety Scale (RCMAS) for adolescents. *Educational and Psychological Measurement, 48,* 429-433.

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders, 4,* 124 –147. http://dx.doi.org/10.1080/15564880 802612573

Lipsey, M. W., Landenberger, N. A., Wilson, S. J., & Lipsey, M. (2007). Effects of cognitive-behavioral programs for criminal offenders. *Campbell Systematic Reviews, 6.*

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

McGee, R. A., Wolfe, D. A., & Wilson, S. K. (1997). Multiple maltreatment experiences and adolescent behavior problems: adolescents' perspectives. *Development and Psychopathology, 9,* 131–149.

Melchert, T. P. (1998). Testing the validity of an instrument for assessing family of origin history. *Journal of Clinical Psychology, 54,* 863–875.

Miller, F. G. & Lazowski, L. E. (2005). Substance Abuse Subtle Screening Inventory for Adolescents-Second Version. In T. Grisso, G. Vincent, D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice,* (pp. 139-151). New York, NY, US: Guilford Press.

Monahan, J., & Skeem, J. (in press). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology.*

Monahan, J., & Skeem, J. (2014). Risk redux: The resurgence of risk assessment in criminal sentencing. *Federal Sentencing Reporter, 26*, 158-166.

Monahan, K. C., Steinberg, L., & Cauffman, E. (2009). Affiliation with antisocial peers, susceptibility to peer influence, and desistance from antisocial behavior during the transition to adulthood. *Developmental Psychology, 45*, 1520-1530.

National Academy of Sciences (2013). *Reforming juvenile justice: A developmental approach.* Washington, DC: The National Academies Press.

Newman, J. P., & Kosson, D. S. (1986). Passive avoidance learning in psychopathic and nonpsychopathic offenders. *Journal of Abnormal Psychology, 95,* 257-263.

Ogilvie, J. M., Stewart, A. L., Chan, R. C. K., & Shum, D. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology, 49*, 1063-1107.

Olver, M. E., Stockdale, K. C., & Wong, S. C. (2012). Short and Long-Term Prediction of Recidivism Using the Youth Level of Service/Case Management Inventory in a Sample of Serious Young Offenders. *Law and Human Behavior, 36,* 331-344.

Olver, M. E., Stockdale, K. C., and Wormith, J. S. (2009). Risk assessment with young offenders A meta-analysis of three assessment measures. *Criminal Justice and Behavior, 36,* 329-353.

Orbis Partners, Inc. (2007). *Youth Assessment and Screening Instrument.* Ottawa, Ontario: Orbis Partners, Inc.

Patrick, C. J. (2010). *Description of brief scales for assessment of boldness, meanness, and disinhibition*. Unpublished manual.

Patrick, C.J., & Venables, N. (2010, March). Boldness, meanness and disinhibition facets of psychopathy: Relations with aggression, substance use, internalizing, and suicidality. Paper presented at the conference of the American Psychology-Law Society (APLS), Vancouver, Canada.

Perrin, S., & Last, C. G. (1992). Do childhood anxiety measures measure anxiety? *Journal of Abnormal Child Psychology, 20,* 567-578.

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence, 2,* 117-133.

Reynolds, C. R. (1982). Convergent and divergent validity of the Revised Children's Manifest Anxiety Scale. *Educational and Psychological Measurement, 42,* 1205-1212.

Reynolds, C. R., & Richmond, B. O. (1985, 2000). *Revised Children's Manifest Anxiety Scale. RCMAS Manual.* Los Angeles, CA: Western Psychological Services.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent development in quantitative methods for literature reviews. *Annual Review of Psychology, 52,* 59-82.

Schmitz, K. E., Hovell, M. F., Nichols, J. F., Irvin, V. L., Keating, K., Simon, G. M., et al. (2004). A validation study of early adolescents' pubertal self-assessments. *Journal of Early Adolescence, 24,* 357–384.

Schubert, C. A., Mulvey, E. P., Steinberg, L., Cauffman, E., Losoya, S., Hecker, T., Chassin, L., et al. (2004). Operational lessons from the Pathways to Desistance project. *Youth Violence and Juvenile Justice, 2,* 237-255.

Skeem, J. L., Kennealy, P. J., Hernandez, I., Clark, S., & Tatar, J. R. (2013). *CA- YASI Predictive Utility: How well do scores and classifications predict youths' infractions and re-arrest?* Report prepared for the California Department of Juvenile Justice. Available: http://risk-resilience.berkeley.edu/content/publications-1

Skeem, J., Kennealy, P., & Hernandez, I. (2013). CA-YASI construct validity: To what extent do the domains measure the risk factors they're supposed to measure? Report prepared for the Division of Juvenile Justice, California Department of Corrections and Rehabilitation (CDCR). Available: http://risk-resilience.berkeley.edu/content/publications-1

Skeem, J., Barnoski, R., Latessa, E., Robinson, D., & Tjaden, C. (2013). *Youth risk assessment approaches: Lessons learned and question raised by Baird et al.'s study.* NCCD Advisory Board Rebuttal prepared for the Office of Juvenile Justice and Delinquency Prevention (OJJDP). Available: http://risk-resilience.berkeley.edu/content/publications-1

Skeem, J., Polaschek, D., Patrick, C., & Lilienfeld, S. (2011). Psychopathic personality: Bridging the gap between scientific evidence and public policy. *Psychological Science in the Public Interest, 12,* 95-162.

Skeem, J. L., Schubert, C. A., Odgers, C., Mulvey, E. P., Gardner, W. Lidz, C. (2006). Psychiatric symptoms and community violence among high-risk patients: A test of the relationship at the weekly level. *Journal of Consulting and Clinical Psychology, 74,* 967-979.

Steinberg, L. (2010). A dual systems model of adolescent risk-taking. *Developmental Psychobiology, Special Issue: Psychobiological models of adolescent risk, 52*, 216-224.

Stockdale, K. C., Olver, M. E., & Wong, S. C. (2014). The Validity and Reliability of the Violence Risk Scale–Youth Version in a Diverse Sample of Violent Young Offenders. *Criminal Justice and Behavior, 41(1),* 114-138.

Vincent, G. (2015). *Application and implementation of risk assessment in juvenile justice for the courts.* National Courts and Science Institute. Available online.

Vincent, G. M., Paiva-Salisbury, M. L.,Cook, N. E., Guy, L. S., Perrault, R. T. (2012). Impact of risk/needs assessment on juvenile probation officers' decision making: Importance of implementation. *Psychology, Public Policy, and Law, 18*, 549-576.

Vaughn, M. G., Salas-Wright, C. P., DeLisi, M., & Perron, B. (2014). Correlates of traumatic brain injury among juvenile offenders: A multi-site study. *Criminal behaviour and mental health, 24(3*), 188-203.

Thornberry, T. P., Lizotte, A. J., Krohn, M. D., Farnworth, M. & Jang, S. J. (1994). Delinquent peers, beliefs, and delinquent behavior: A longitudinal test of interactional theory. *Criminology, 32,* 47-83.

Wachter, A. (April, 2014). Issue Brief: Risk Needs Assessment in Juvenile Probation. JJGPS State Scan. Pittsburgh, PA: National Center for Juvenile Justice

Walters, G. D. (1990). *The criminal lifestyle: Patterns of serious criminal conduct.* Newbury Park, CA: Sage.

Walters, G. D. (2011). Predicting recidivism with the Psychological Inventory of Criminal Thinking Styles and Level of Service Inventory-Revised: Screening Version. *Law & Human Behavior, 35*, 211-220.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence.* The Psychological Corporation: Harcourt Brace & Company. New York, NY.

Yang, M., Wong, S. C., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136(5),* 740-767.

Table 1. CA-YASI Domain and Total Correlations

	Legal History	Correc Resp	Viol-Agg	Social Infl	Sub Use	Attitudes	Soc-Cog	Family	Educ-Employ	Health	Comm Link	Comm Stab
Legal History	1.00	---	---	---	---	---	---	---	---	---	---	---
Correctional Response	.64**	1.00	---	---	---	---	---	---	---	---	---	---
Violence-Aggression	.31**	.41**	1.00	---	---	---	---	---	---	---	---	---
Social Influences	.35**	.35**	.67**	1.00	---	---	---	---	---	---	---	---
Substance Use	.24**	.26**	.38**	.38**	1.00	---	---	---	---	---	---	---
Attitudes	.19**	.31**	.68**	.65**	.10	1.00	---	---	---	---	---	---
Social-Cognitive Skills	.13*	.18**	.65**	.49**	.03	.68**	1.00	---	---	---	---	---
Family	.13*	.33**	.38**	.30**	.20**	.25**	.29**	1.00	---	---	---	---
Education-Employment	.13*	.26**	.47**	.41**	.04	.53**	.52**	.34*	1.00	---	---	---
Health	.11	.25**	.08	-.07	.09	.02	.02	.25**	.22**	1.00	---	---
Community Linkages	.11	.09	.05	.26**	.10	.11	.07	.15*	.14*	.01	1.00	---
Community Stability	.23**	.19**	.32**	.32**	-.05	.15*	.20**	.38**	.27**	.05	.12	1.00
Domain/Total Correlation	.46**	.58**	.83**	.79**	.48**	.76**	.71**	.56**	.70**	.22**	.23**	.47**

** $p < .01$, * $p < .05$

Table 2. CA-YASI Domain Definitions and Psychometric Properties

Subscale	Number of Items	Alpha	ICC	Mean (Standard Deviation)	Definition
Total Score	105	.93	.84	191.4 (36.8)	Unit-weighted total of all items
Legal History	7	.62	.82	6.3 (4.0)	Criminal behavior that is frequent, varied, and with early onset
Correctional Response	11	.65	.72	5.5 (4.2)	Noncompliance with rules of institutional or community placement (misconduct, violations, new offenses)
Aggression-Violence	22	.86	.66	22.7 (7.4)	Past violence, anger/hostility, callousness, attitudes supportive of aggression
Social Influences	8	.89	.62	24.0 (7.7)	Attachment to antisocial peers, absence of constructive adult role models
Substance Use	3	.52	.57	9.5 (4.1)	Frequent alcohol and drug use that can impair functioning
Attitudes	6	.87	.79	13.0 (4.4)	Antisocial attitudes including minimization of responsibility, denial of harm, poor attitudes toward authority
Social-Cognitive Skills	8	.93	.57	19.3 (5.7)	Poor decision-making skills relevant to antisocial behavior (consequential thinking, goal setting, problem-solving, perspective taking)
Family	8	.75	.80	16.1 (4.6)	Poor family relationships/role modeling
Education-Employment	15	.80	.84	40.7 (8.2)	Poor achievement and/or motivation for education and employment
Health	7	.49	NA	6.7 (2.9)	Mental health problems
Community Linkages	4	.65	.85	10.0 (2.2)	Lack of relevant services to address criminogenic needs in the community
Community Stability	6	.62	.67	17.4 (3.7)	Poor finances, accommodation, or transportation

Notes. ICC= Intraclass Correlation Coefficient. There was insufficient information to calculate reliability for the Health Domain

Table 3. Psychometric Properties of Convergent and Discriminant Validity Measures

Validity Measure	Type	Number of Items	Alpha	ICC	Mean (SD)
Discriminant Measures					
Somatization	Self-Report	7	.85	NA	1.8 (3.5)
Head Injury	Self-Report	1	NA	NA	0.4 (0.5)
Intelligence	Cognitive Task	Variable	NA	NA	88.4 (12.3)
Physical Maturation	Self-Report	5	.71	NA	15.3 (2.8)
Convergent Measures					
Violence Aggression Domain					
Social Information Processing					
Intent	Clinical Rating	4	.47	.91	3.6 (1.3)
Response	Clinical Rating	4	.53	.87	4.7 (1.0)
Attitudes	Self-Report	4	.70	NA	9.9 (2.1)
Anger	Self-Report	5	.77	NA	3.5 (3.8)
Meanness	Self-Report	19	.89	NA	56.5 (10.7)
Social Deviance	Clinical Rating	10	.88	.88	13.0 (4.0)
Attitudes Domain					
PICTS	Self-Report	35	.88	NA	68.7 (14.9)
Social-Cognitive Domain					
Tower of London					
Impulsivity/1st Move (ms)	Cognitive Task	21 trials	NA	NA	5303.3 (1981.0)
Planning/Errors (avg)	Cognitive Task	21 trials	NA	NA	0.9 (0.7)
Response Inhibition (Go/No Go)	Cognitive Task	100 trials	NA	NA	17.4 (6.2)
Social Influences Domain					
Peer Delinquent Behavior					
Behavior	Self-Report	12	.93	NA	39.2 (11.0)
Influence	Self-Report	7	.90	NA	16.9 (6.8)
Neighborhood Disorganization (*Z*)	Coded	NA	NA	NA	0.0 (1.0)
Family Domain					
Family Background Questionnaire					
Psychological Abuse	Self-Report	8	.90	NA	6.1 (5.1)
Physical Abuse	Self-Report	3	.77	NA	2.8 (2.6)
Domestic Violence	Self-Report	4	.82	NA	1.4 (2.4)

Family Monitoring	Self-Report	8	.69	NA	14.4 (4.3)
Education/Employment Domain					
School Connection	Self-Report	19	.77	NA	52.8 (8.5)
Irresponsible and Impulsive Lifestyle	Clinical Rating	5	.61	.80	5.9 (2.2)
Substance Use Domain					
Alcohol Dependence	Self-Report	12	.86	NA	6.1 (6.1)
Drug Dependence	Self-Report	12	.91	NA	9.4 (7.5)
(Mental) Health Domain					
Psychological Distress (BSI Total)	Self-Report	53	.96	NA	23.7 (27.1)
Anxiety	Self-Report	37	.85	NA	6.2 (5.0)
Legal History Domain					
Criminal Behavior (PCL:YV)	Clinical Rating	5	.69	.75	7.1 (2.4)
Conduct Disorder	Self-Report	15	.69	NA	8.3 (2.7)
AWOL/Escape History	Coded	NA	NA	NA	0.1 (0.3)

Table 4. *Correlations between CA-YASI Domains and Convergent and Discriminant Measures*

Measure	Type	*r*	Z_r	*SE*	*N*
Total					
PCLYV Social Deviance	Convergent	.47***	.51	.07	235
Somatization	Discriminant	.15*	.15	.07	221
Head Injury	Discriminant	.04	.04	.07	235
Low IQ	Discriminant	.33***	.34	.07	190
Low Pubertal Maturation	Discriminant	.20**	.20	.07	235
Aggression-Violence					
SIP Intent	Convergent	.04	.07	.07	237
SIP Response	Convergent	.17*	.08	.07	237
SIP Attitudes	Convergent	.23***	.23	.07	237
BSI Hostility	Convergent	.15*	.15	.07	222
Meanness	Convergent	.24**	.24	.08	180
Social Deviance	Convergent	.36***	.38	.07	237
Somatization	Discriminant	.06	.06	.07	222
Head Injury	Discriminant	.06	.06	.07	237
Low IQ	Discriminant	.28***	.29	.07	191
Low Pubertal Maturation	Discriminant	.14*	.14	.07	237
Attitudes					
PICTS Total	Convergent	-.12	-.12	.11	86
SIP Attitudes	Convergent	.16*	.16	.07	237
Somatization	Discriminant	.09	.09	.07	222
Head Injury	Discriminant	.03	.03	.07	237
Low IQ	Discriminant	.23***	.23	.07	191
Low Pubertal Maturation	Discriminant	.20*	.20	.07	237
Social-Cognitive Skills					
TOL Errors	Convergent	.17*	.17	.07	200
TOL Impulsivity	Convergent	.05	.05	.07	200
Go/No-Go	Convergent	-.11	-.11	.07	237
Low IQ	Convergent	.20*	.20	.07	191
Somatization	Discriminant	.12	.12	.07	222
Head Injury	Discriminant	-.03	-.03	.07	237
Low Pubertal Maturation	Discriminant	.17*	.17	.07	237
Social Influences					
PDB Behavior	Convergent	.22**	.22	.07	234
PDB Influence	Convergent	.05	.05	.07	235
Neighborhood Disorganization	Convergent	.21**	.21	.08	162
Somatization	Discriminant	.00	.00	.07	222
Head Injury	Discriminant	.06	.06	.07	237
Low IQ	Discriminant	.26**	.27	.07	191
Low Pubertal Maturation	Discriminant	.12	.12	.07	237
Family					
FBQ Psychological	Convergent	.00	.00	.07	231

FBQ Physical	Convergent	.13*	.13	.07	230
FBQ Domestic Violence	Convergent	.16*	.16	.07	230
Family Monitoring	Convergent	.19**	.19	.07	233
Somatization	Discriminant	.18**	.18	.07	222
Head Injury	Discriminant	.02	.02	.07	237
Low IQ	Discriminant	.14	.14	.07	191
Low Pubertal Maturation	Discriminant	.08	.08	.07	237
Education/Employment					
School Connection	Convergent	0.10	.10	.08	153
Impulsive-Irresponsible Lifestyle	Convergent	.17**	.17	.07	237
Low IQ	Convergent	.24**	.24	.07	191
Somatization	Discriminant	.22**	.22	.07	222
Head Injury	Discriminant	.06	.06	.07	237
Low Pubertal Maturation	Discriminant	.31**	.32	.07	237
Substance Use					
Alcohol	Convergent	.32**	.26	.07	237
Drug	Convergent	.35***	.31	.07	237
Somatization	Discriminant	-.05	-.05	.07	222
Head Injury	Discriminant	.10	.10	.07	237
Low IQ	Discriminant	.22**	.22	.07	191
Low Pubertal Maturation	Discriminant	.02	.02	.07	237
Health					
Total BSI	Convergent	.21**	.21	.07	219
Anxiety RCMAS	Convergent	.26**	.27	.07	237
Head Injury	Discriminant	.09	.09	.07	237
Low IQ	Discriminant	.05	.05	.07	191
Low Pubertal Maturation	Discriminant	.05	.05	.07	237
Legal History					
PCLYV Facet 4	Convergent	.50***	.55	.07	237
Conduct Disorder	Convergent	.15*	.15	.07	217
Somatization	Discriminant	.01	.01	.07	222
Head Injury	Discriminant	.02	.02	.07	237
Low IQ	Discriminant	.11	.11	.07	191
Low Pubertal Maturation	Discriminant	.01	.01	.07	237
Correctional Response					
PCLYV Facet 4	Convergent	.52***	.58	.07	237
AWOL/Escape	Convergent	.32***	.33	.07	237
Somatization	Discriminant	.16*	.16	.07	222
Head Injury	Discriminant	.05	.05	.07	237
Low IQ	Discriminant	.19*	.19	.07	191
Low Pubertal Maturation	Discriminant	-.02	-.02	.07	237

Table 5

Mean Effect Sizes of Convergent and Discriminant Measure Groups for All Domains

	k	*Mean Z_r*	*Lower Bound*	*Upper Bound*	*SE*	*Q*	*Q df*
YASI Total							
Discriminant	4	18***	.11	.24	.03	9.79*	3
Aggression-Violence							
Convergent	6	0.20***	.15	.26	.03	14.60*	5
Discriminant	4	0.13***	.06	.20	.03	6.92	3
Attitudes							
Convergent	2	0.09	-.02	.20	.06	4.87*	1
Discriminant	4	0.14***	.07	.20	.03	5.94	3
Social/Cognitive Skills							
Convergent	4	0.07*	.00	.14	.04	13.05*	3
Discriminant	3	0.09*	.01	.16	.04	5.13	2
Social Influences							
Convergent	3	0.16***	.08	.23	.04	4.18	2
Discriminant	4	0.11***	.04	.17	.03	7.82*	3
Family							
Convergent	4	0.12***	.06	.19	.03	4.90	3
Discriminant	4	0.10***	.04	.17	.03	3.37	3
Education/Employment							
Convergent	3	0.18***	.10	.26	.04	1.75	2
Discriminant	3	0.20***	.13	.28	.04	8.10*	2
Substance Use							
Convergent	2	0.35***	.26	.44	.05	0.14	1
Discriminant	4	0.07*	.00	.13	.04	8.39*	3
Health							
Convergent	2	0.24***	.15	.333	.05	0.31	1
Discriminant	3	0.06	-.01	.14	.04	0.24	2
Legal History							
Convergent	2	0.36***	.27	.45	.05	17.72***	1
Discriminant	4	0.03	-.03	.10	.03	1.41	3
Correctional Response							
Convergent	2	0.45***	.36	.54	.05	7.01***	1
Discriminant	4	0.09**	.02	.16	.03	6.29	3

Endnotes

[i] Intelligence was used as a discriminant measure for the CA-YASI, with the exception of the Social-Cognitive and Education/Employment domains—where weak convergent correlations are expected. First, although the Social-Cognitive domain is meant to assess neuropsychological problems relevant to offending, intelligence relates to such problems. Second, although the Educational/Employment domain is meant to assess relevant motivation and achievement, low intelligence relates to poor educational achievement. Intelligence theoretically relates more weakly to both domains than the other convergent measures (e.g., response inhibition, impulsivity and planning for Social Cognitive; school connection for Education/Employment).