Published in final edited form as:

Child Youth Serv Rev. 2023 September ; 152: . doi:10.1016/j.childyouth.2023.107039.

Examining the measurement precision of behavior problems among a sample of primarily rural youth on juvenile probation and their parents

Jill Viglione[a,*], **Kristina K. Childs**[a], **Jennifer H. Peck**[a], **Jason E. Chapman**[b], **Tess K. Drazdowski**[b], **Michael R. McCart**[b], **Ashli J. Sheidow**[b]

[a]Department of Criminal Justice, University of Central Florida, 12805 Pegasus Drive, Orlando, FL 8216, United States

[b]Oregon Social Learning Center (OSLC), 10 Shelton McMurphey Blvd., Eugene, OR 97401, United States

Abstract

Accurate and timely intervention in the justice system is particularly critical in rural communities, given documented barriers to accessible, evidence-based services for youth. As youth in the juvenile justice system have a high prevalence of behavioral health needs, accurate assessment of those needs is a critical first step in linking youth to appropriate care. The goal of the current study is to examine the reliability of a brief assessment (the Brief Problem Checklist [BPC]) among a sample of 222 justice-involved youth and their caregivers who primarily reside in rural communities in the United States. Using a series of reliability analyses and tests of agreement, we examined whether youth and caregiver BPC produces reliable scales, the strength of the convergence among each of the BPC scales, and youth and caregiver agreement on the BPC scales. Findings support the reliability of the BPC, but not inter-rater reliability. Poor agreement between youth and caregiver reports exists for both youth internalizing and externalizing problems. Additionally, the BPC was significantly related to several theoretically relevant constructs, including treatment, substance use disorder severity, and family history of substance use. These findings lend merit to discussions about the need for more research on the reliability and validity of assessment instruments before their widespread use in guiding youth- and agency case planning decisions, along with informing conclusions about program effectiveness.

*Corresponding author. Jill.viglione@ucf.edu (J. Viglione),.

CRediT authorship contribution statement

Jill Viglione: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Kristina K. Childs:** Conceptualization, Writing – original draft, Writing – review & editing. **Jennifer H. Peck:** Formal analysis, Writing – original draft, Writing – review & editing. **Jason E. Chapman:** Conceptualization, Methodology, Data curation. **Tess K. Drazdowski:** Investigation, Writing – review & editing, Supervision. **Michael R. McCart:** Conceptualization, Investigation, Supervision, Project administration. **Ashli J. Sheidow:** Conceptualization, Investigation, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Keywords

Juvenile justice; Rural; Brief Problem Checklist; Youth problem behaviors; Reliability

1. Introduction

Youth involved in the juvenile justice system across the United States disproportionately experience high rates of behavioral health concerns, including mental health and substance use issues (Belenko et al., 2017; Golzari, Hunt, & Anoshiravani, 2006; Fazel, Doll & Langstrom, 2008; Teplin et al., 2002; Wasserman et al., 2010). While the juvenile justice system can serve as an outlet for youth to access needed services they might not otherwise have access to, this process relies on the accurate assessment of youth problems. Recently, juvenile justice organizations have moved towards relying on standardized instruments to gather information about youth problem behaviors and their service needs, guide case planning and service referral decisions, and determine progress towards release (Howell et al., 2017; Lyons, 2009; Wachter, 2015).[1] Accurate assessment is the first step in ensuring youth receive appropriate, needs-based interventions. However, relying on a tool that produces inaccurate or unreliable information will inevitably lead to unintended negative consequences for youth and public safety.

Best practices for achieving “accuracy” while balancing efficiency have not received a great deal of attention in the juvenile justice literature. Research demonstrates that juvenile probation officers (JPOs) are dealing with high caseloads and work-related stress and burnout (Dir et al., 2019; Mack & Rhineberger-Dunn, 2022; White et al., 2015). Valid and reliable assessment of youth needs takes time and resources, especially when assessments are completed from multiple perspectives (e.g., youth and caregiver) and assess multiple risk factors (e.g., substance use, mental health). Thus, the integration of assessment practices into the routine activities of JPOs adds another layer to the responsibilities and time commitments of JPOs (Guy et al., 2014; Shook & Sarri, 2007). As a result, jurisdictions across the country implemented brief assessment tools to balance the benefits of needs assessment with the high workloads already experienced by JPOs (e.g., Belenko et al., 2017; Grisso & Underwood, 2003; Jones et al., 2016).

There are a range of factors present in rural settings across the United States that are important to consider when evaluating a tool’s performance in the field, including limited access to and availability of appropriate treatment services (Fehr et al., 2020; Jensen et al., 2021; Oser et al., 2011), overworked professionals with large caseloads and little available time to conduct assessments (Bethea et al., 2020; Paris & Hoge, 2010), and youth with different and complex intervention needs (e.g., problem behavior patterns: Krupa et al., 2021; cumulative risk: Connolly et al. 2017, Dembo et al. 2020; substance use: National Center on Addiction and Substance Use 2000). Screening and assessment of behavioral problems is an essential first step to linking youth to necessary treatment and services. This is especially true for youth who have a substance use disorder (SUD), as there is

[1]There are various methods for assessing youth problem behaviors. For a detailed review of these instruments, please see Baird et al. (2013), Grisso & Vincent (2005), and Wachter (2015).

greater access to targeted rehabilitation services for youth once under court authority. In rural communities, accurate and timely intervention in the justice system becomes even more critical given the documented barriers to accessible, evidence-based services for youth with behavioral health problems (Fehr et al., 2020; Jensen et al., 2021; Oser et al., 2011). Ensuring that assessment practices (and the specific measures used, such as family history of substance use, prior stays in substance use treatment) can correctly identify the array of behavioral health needs experienced by rural youth involved in the justice system and the types of problem behaviors they engage in, is vital to the success of aligning a youth's needs, treatment, and outcomes.

Accordingly, the purpose of this study is to explore the use of brief multi-informant assessment procedures, specifically among youth who are involved in the juvenile justice system in rural communities. Youth involved in the juvenile justice system have high rates of behavioral health concerns that are diverse, co-occurring, and often vary from one youth to the next. Clear empirical evidence demonstrates that relying on systematic methods for assessing youth functioning is ideal during case planning (Belenko et al., 2017; Grisso et al., 2005; Hunsley & Lee, 2007), although the most accurate and reliable approach for collecting this information has not been resolved. Scholars have continually questioned the widespread use of these tools without convincing support for their psychometric properties (Baird, 2009; Boateng et al., 2018; McCrae et al., 2011). Specifically, the precision of how youth behaviors are measured across contexts, instruments, and raters, as well as the validity of the interpretations made from assessment scores, has not received a great deal of attention (e.g., Baird et al., 2013; De Los Reyes et al., 2015; Singh & Fazel, 2010). In particular, the reliability of brief assessment tools specifically among youth who are involved in the juvenile justice system in rural communities represents a significant gap in this area of research, which then has serious implications for policy and practice for the treatment of justice-involved youth in various geographic locations.

2. Methods for assessing youth problems

According to the Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) the reliability of measurement (i.e., precision of measurement) becomes more essential as the "consequences of decisions and interpretations grow in importance" (pg, 33). Most scholars agree that the quality of information gained from only one source (e.g., only the individual themselves) is not sufficient, especially when it is obtained from children and adolescents (Baird et al., 2013; Bögels et al., 2004; Becker et al., 2004; De Los Reyes et al., 2015). Relying solely on self-reported symptoms and behaviors by youth has been criticized in the literature for several reasons, such as youth being more developmentally immature than adults, having a lack of awareness of symptoms, under- and over-reporting, selecting socially desirable answer choices, and poor memory (Bögels et al., 2004; De Los Reyes et al., 2013; Lewis et al., 2014).

The most widely supported strategy for assessing youth behavioral health needs is the reliance on corroborative sources of information. This approach involves collecting

information from more than one individual who is close to the youth and who spends a significant amount of time observing the youth's behavior (Achenbach, 2006). Strong empirical evidence indicates that multi-informant approaches strengthen the accuracy and quality of information about adolescents' symptoms and behaviors (Klein, Dougherty, & Olino, 2005; Silverman & Ollendick, 2005). One strength of relying on multi-informant approaches is that data from more than one source provides information from different perspectives (e.g., parents, teachers) and across different contexts (e.g., family, school). This means that "…different informants contribute unique value and nonoverlapping predictive information…" to understanding youth functioning (Hourigan et al., 2011, p. 198). Another strength is the ability to cross-validate the information provided, which addresses concerns about under- or over-reporting of symptoms or behaviors (Piehler et al., 2020).

While the use of multi-informant approaches has grown in recent years and is now considered a best practice in clinical guidelines (see American Academy of Child & Adolescent Psychiatry, 2020), decades of research has demonstrated that parents, teachers, and adolescents do not produce high levels of agreement on the severity of youths' problems (Achenbach et al., 1987; De Los Reyes et al., 2015: Hughes & Gullone, 2010; Miller et al., 2014; Youngstrom et al., 2000). Often referred to as "informant discrepancy", these studies focus on the level of divergence among informants and the correlates of higher levels of discordance (De Los Reyes et al., 2013). Informant discrepancy is related to inter-rater reliability, which is the extent to which independent raters produce similar scores on the same test. Given the considerable discretion inherent in gathering information about youth behavior and then making judgements about youth and family intervention needs based on this information, confirming consistency in scores across raters is essential. On a practical level, these types of clinical decisions are common and are largely inconsequential for the agency making decisions. However, for youth and their families, these decisions are critical in shaping the availability of and access to appropriate and necessary services, which in turn, shapes their long-term success during supervision and beyond (e.g., recidivism, well-being).

Studies of inter-rater reliability generally produce moderate congruence across parents and youth, providing some support for the inter-rater reliability of items measuring youths' behavioral problems, although these studies also demonstrate that informant discrepancies are not trivial (Achenbach & Rescorla, 2014; De Los Reyes, 2011; Dimler et al., 2017). For instance, one of the most consistent findings across studies of informant discrepancies is that adolescents report higher rates of problematic behaviors and score lower on measures of youth functioning compared to their parents' reports (Rescorla et al., 2013; van der Ende et al., 2012). Research also suggests that parents and youth tend to agree more often about internalizing problems (e.g., anxiety) compared to externalizing problems (e.g., substance use) (Rescorla et al., 2013; Seiffge-Krenke & Kollmar, 1998). These discrepancies lead to several questions regarding the magnitude of agreement across parents and youth, the types of behaviors that show agreement, whether the behaviors that are often targeted during intervention planning (e.g., substance use, mental health) are consistent across raters, and the implications for relying on youth versus parent reports when ratings do not converge.

2.1. Brief problem checklist

The Brief Problem Checklist (BPC) is a short interview tool that assesses internalizing and externalizing problems among children and adolescents (Chorpita et al., 2010). The tool was designed for use as a quick interview guide used in clinical and research settings to evaluate a youth's intervention progress in a timely manner. Clinical progress is measured by administering a baseline BPC prior to intervention and then routinely assessing the youth with the BPC at specified intervals (e. g., often once per week). The BPC is recognized as a validated and reliable measure of internalizing and externalizing problems (Beidas et al., 2015; Chorpita et al., 2017; National Assembly on School-Based Health Care, 2011). In clinical settings, scores from the BPC are used to make service decisions and measure improvements in symptomology over time. In research settings, changes in BPC scores over time are used as an indicator of treatment effectiveness (Weisz et al., 2011; 2012).

There are two versions of the BPC: youth self-report and caregiver. Each version of the BPC consists of twelve items: six items representing internalizing problems and six items representing externalizing problems. These items were initially selected based on a comprehensive analysis of the internal structure of the Youth Self-Report (YSR; Achenbach & Rescorla, 2001) and Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001).[2] Both instruments are widely used and empirically validated assessments of youth functioning, although research has not found strong cross-informant agreement (e.g., Achenbach et al., 1987; Ebesutani et al., 2011; Huang, 2017). Consistent findings across the YSR and CBCL include low to moderate rates of parent/youth agreement, higher discrepancies for internalizing problems compared to externalizing problems, and variations in parent/youth divergence across important individual and contextual characteristics (e.g., Berg-Nielsen et al., 2003; Ferdinand et al., 2006; Martin et al., 2004; Rey et al., 1992; Sourander et al., 1999). However, most of these studies were conducted with non-justice involved youth (i.e., Achenbach et al. (1987) *meta*-analysis of 119 studies includes just five with justice-involved populations). The focus of the existing research on non-clinical settings is noteworthy given the stakes are much lower for these youth and their families compared to those under the umbrella of the justice system who face additional consequences.

In one of the only published studies of the measurement properties of the BPC, Chorpita et al. (2010) conducted a comprehensive assessment of the caregiver and youth BPC scales by examining reliability (test–retest), correlations with the corresponding YSR and CBCL scales (convergent validity), BPC scores across diagnostic groupings (discriminant validity), and correlations between caregiver and youth responses in the United States. These analyses led the researchers to conclude that the findings were generally "…quite supportive of the psychometric strength of the BPC" (Chorpita et al., 2010, p. 534). However, correlations (r) among caregiver and youth BPC scales were low, ranging from 0.19 for the total BPC scale to 0.31 for the externalizing scale. In addition, correlations cannot provide estimates of agreement between caregiver and youth ratings. More recently, Chorpita et al. (2017) found

[2]The BPC was developed based on analysis of the YSR and CBCL, which are both lengthy instruments consisting of over 100 items that are not freely available. The goal of the BPC was to develop a free and brief screener to identify items most important in identifying youth problem behavior.

that U.S. youth self-reported fewer externalizing problems compared to their caregivers over time, with no differences identified for the internalizing scale or total problems scale between youth and caregivers.

3. Current study

There are several reasons, related to both research and practice, that underscore the necessity of additional tests of the reliability and cross-informant agreement of a brief multi-informant behavioral health assessment for use in juvenile legal system settings, such as the BPC scales. Most important is the reliance on BPC scores to inform judgements about behavioral health needs and treatment progress within a juvenile legal system setting. These judgements carry implications for ensuring public safety and improving the well-being of the youth, family, and community. At the organizational level, aggregate BPC scores also can be helpful in informing choices relating to service provider contracts and the allocation of resources. In evaluations of intervention efficacy and effectiveness, improvement in BPC scores also serves as a critical outcome measure that may extend to conclusions about feasibility and scalability in local jurisdictions (e.g., Chorpita et al., 2017; Weisz et al., 2012).

In particular, none of the three available studies examined the measurement properties of the BPC specifically on juvenile justice or rural youth samples.[3] Furthermore, rural youth are often underrepresented in community-based samples of youth, losing the ability to detect possible differences in the properties of the tool among this specific population. The clear benefits of conducting quality screening and assessment for justice-involved youth (Grisso et al., 2005; Lipsey et al., 2012; National Research Council, 2013; Vincent et al., 2016), coupled with the “brief” format of the instrument (i.e., comprised of only 12 items) and its availability at no cost to juvenile justice organizations, represent clear strengths of the BPC for use in rural juvenile justice settings. These characteristics also underscore the importance of examining its measurement precision among a sample of this unique population. As a result, the current study tests the internal consistency, convergent validity, and youth-caregiver agreement of the three BPC scales among a sample of youth with a SUD currently on probation in rural communities. More specifically, we sought to address the following research questions. First, does the youth and caregiver BPC produce reliable scales representing internalizing problems, externalizing problems, and total problems? Second, how strong is the convergence among each of the youth and caregiver BPC scales and other, relevant treatment measures? Third, do youth and caregivers agree on the BPC scales for internalizing problems, externalizing problems, and total problems?

4. Method

4.1. Study Description

Data for the current study come from an ongoing randomized controlled trial aimed at examining the effectiveness of Contingency Management delivered by juvenile probation

[3]A task for future research is to also examine the measurement precision of the CBCL and YSR with youth who reside in rural populations, but these are much lengthier assessment tools.

officers compared to Probation as Usual. More information on this study can be found in other reports (see [redacted for anonymity]). The current study uses baseline data, collected prior to youth exposure to CM of PAU. From October 2017 to December 2021, these data were collected from youth under probation supervision and their caregivers on a range of measures including engagement in delinquency, substance use, risky sexual behavior, service utilization, family instability, and behavioral risk factors.

4.2. Data collection

Baseline data were collected either in-person or via video conference (e.g., Zoom) by a trained assessor from the [Center name redacted for anonymity]. Prior to conducting the assessments, the assessor obtained caregiver consent and youth assent. The assessor completed the *Structured Adolescent Interview (SAI)* jointly with the youth and caregiver to gather demographic data. Upon completion of the *SAI*, the assessor met individually with the youth to complete the *Global Appraisal of Individual Needs (GAIN)* substance use scales*, Self-Report Delinquency Scale, Sexual Risk Behavior Scale*, and *Brief Problem Checklist*. The assessor then met individually with the caregiver to complete the caregiver version of the *Brief Problem Checklist*. Upon completion of all baseline assessments, youth were randomized into the Contingency Management or Probation as Usual study condition (i.e., participants were not aware of their study status during the baseline interview). For their participation, families were compensated $10 for the initial MINI Kid screening and $20 for participation in the full baseline interview. All study procedures were approved by the [Center name redacted for anonymity] Institutional Review Board.

4.3. Study sample

Youth were referred from 13 agencies across three Western states for participation. Referral eligibility criteria for youth included: 1) age 11–18 with a parent or caregiver who would participate in the study, 2) recent drug or alcohol use, and 3) at least four months left on probation supervision. Those excluded from referral were youth charged with a sexual offense or those actively participating in a drug court program. A total of 360 youth met these criteria and were referred to an assessor from the research center for further eligibility screening. During this initial meeting, the assessor administered the Mini International Neuropsychiatric Interview (MINI Kid; Sheehan et al. 2010) to both the youth and caregiver separately to screen for drug and alcohol use disorders. Youth were eligible for the trial if they met criteria for at least one type of SUD based on theirs or caregiver's MINI Kid assessment. Of the 360 youth referred for screening, 68 declined further screening after hearing more about the study, 13 were no longer on probation supervision, 11 were unable to be contacted, and 19 were ineligible based on the criteria described above, leaving 248 youth. Of the remaining youth, 17 declined participation in the study after screening[4]. The remaining 231 youth were randomized into the study. Because the trial is ongoing, the current analyses report on only those youth and caregivers who participated in the baseline assessments at the time of this study (n = 222). Of participating youth, nearly 70% were

[4]We conducted analyses comparing those that declined to participate in the study to the final sample of youth who agreed to participate. We found no significant differences between the two groups based on sex or race (i.e., the only two measures available for youth that declined).

supervised on probation in a rural area (defined by the Federal Office of Rural Health Policy, 2021). The coding and distribution of all sample characteristics are provided in Table 1.

4.4. Measures

Youth Characteristics.—The *Structured Adolescent Interview (SAI)* (Brown, 1989) collected background information on the youth, caregiver, and family. This information includes demographic data on youth and caregivers, treatment participation, educational program participation, and current medication usage. The sample was 66% male, 91% White race, 44% Hispanic ethnicity, with an average age of fifteen years (SD = 1.44). The majority of youth were currently enrolled in an education program (87%).

Caregiver Characteristics.—Caregiver characteristics were also derived from the *SAI*. Of the caregivers who participated in the baseline assessments, 82% were female, 91% were White, 36% were Hispanic, 72% were the youth's biological mother, and were on average 42 years old (SD = 8.66). Approximately 29% of caregivers had a high school degree/GED, some high school (27%) or some college (26%). The majority of caregivers were employed at the time of study (75%).

Brief Problem Checklist.—Given the focus of the current study is on agreement between youth and caregivers, we examine the similarities across youth and caregiver scores for three different measures of problem behaviors: internalizing problems, externalizing problems, and total problems scores.

The main measure of interest is the *Brief Problem Checklist (BPC)*, a clinical assessment designed to measure the progress of a youth during psychological treatment (Chorpita et al., 2010). The *BPC* is designed to capture an Internalizing scale (six items), an Externalizing scale (six items), and a Total Problems scale (all 12 items). The BPC has versions designed to assess both youth and caregiver perceptions of problems individually. Each item is measured on a three-point scale (0 = not true; 1 = somewhat true; 2 = very true). The six internalizing behaviors include: "I worry a lot"; "I am unhappy, sad, or depressed"; "I feel worthless or inferior"; "I feel too guilty"; "I am self-conscious or easily embarrassed"; and "I am too fearful or anxious". The six externalizing behaviors include: "I argue a lot"; "I destroy things belonging to others"; "I disobey my parents or people at school"; "I am stubborn"; "I have a hot temper"; and "I threaten to hurt people." To assess agreement on the *BPC* scales, we summed the items to create the Internalizing scale (youth: mean = 3.30, SD = 3.08; caregiver: mean = 4.69, SD = 3.12), Externalizing scale (youth: mean = 4.08, SD = 2.58; caregiver: mean = 5.27, SD = 3.25), and Total Problems scale (youth: mean = 7.38, SD = 4.66; caregiver: mean = 9.96, SD = 5.25).

To examine whether the BPC scales were associated with theoretically related concepts (i.e., criterion validity), we utilized several youth and caregiver characteristics identified in previous literature as predictive of youth problem behaviors. First, we included whether youth had participated in SUD treatment in the last 90 days obtained from the *SAI*. Approximately 55% of the sample reported that they had not participated in either inpatient or outpatient substance use treatment in the last ninety days. Second, we included a measure of SUD severity according to the MINI Kid. Youth who reported two or three symptoms

were classified as having a mild SUD, followed by moderate (four or five symptoms), and severe (more than six symptoms). Approximately 46% of youth were classified as having a severe SUD according to the MINI Kid. Lastly, we included a proxy measure of family history of substance use problems from the Living Risk Index (LRI; completed by the youth) on the *Global Appraisal of Individual Needs (GAIN)* (Dennis et al., 2003). Youth were asked to answer whether people they have regularly lived with in the past 30 days "have been in drug or alcohol treatment" or "would describe themselves as being in recovery" on a five-point Likert scale (coded as none, a few, some, most, all of them). Both items were summed with an average score of 2.57 (SD = 1.10, $\alpha = 0.73$).

4.5. Analysis plan

To address the first research question, we estimated the internal reliability for the two BPC subscales (i.e., internal and external behaviors) and the total BPC scale using Cronbach's alpha coefficient and the average inter-item correlation (IIC) (Briggs & Cheek, 1986; Piedmont & Hyland, 1993) (Table 2). Second, to examine whether the theoretically related concepts were associated with the BPC scales (e.g., convergent validity, research question 2), we investigated if certain youth and caregiver characteristics identified in previous literature as predictors of youth problem behaviors were correlated with the BPC scales among our sample. These analyses focused on correlations between a youth's substance use treatment history, youth's substance use severity, and family history of substance use with the BPC scales (Table 3). Point bi-serial correlations were used for dichotomous variables (i.e., treatment history), Spearman's correlation for ordinal variables (i.e., substance use severity) and Pearson's *r* for continuous variables (i.e., history of substance use).

To answer the third research question, we conducted correlations (Pearson's *r)* to examine the linear relationship between youth and adult responses for each BPC measure (i.e., youth responses on the internalizing scale compared to caregiver responses on the same scale). These analyses primarily function as a measure of content or criterion-related validity, as correlations do not provide an estimate of agreement between the youth and caregiver responses (Table 4). Last, we conducted *t*-tests to examine the difference between the means of the youth and caregiver responses for each scale, with associated effect sizes using Cohen's d (Table 4). Effect sizes were interpreted using Cohen's (1988) guidelines: "small" (0.2), "medium" (0.5), and "large" (0.8). However, because *t*-tests can hide poor agreement in the distribution of the differences between two means (Zaki et al., 2012), we also included intraclass correlation coefficients (ICCs), the ideal method for assessing agreement among continuous variables (Bartko, 1991; Cicchetti, 1994). Following the standards outlined by Koo and Li (2016), the following cut-offs were used to interpret ICC results: "excellent" (>0.90), "good" (0.75–0.90), "fair" (0.50–0.75), or "poor" (<0.50). All analyses were conducted in SPSS Version 28.

5. Results

The comparisons of youth and caregiver agreement on internalizing problems, externalizing problems, and the total problems scale are reported in Table 2. First, we examined the internal reliability of the internalizing and externalizing subscales of the BPC and the total

problems scale for both youth and caregiver reports. All six scales demonstrated good reliability based on both the alphas (ranged from 0.74 to 0.85) and inter-item correlations (IICs) (ranged from 0.26 to 0.48).

The correlations between theoretically relevant variables and all three BPC scales using correlation coefficients (*r)* are shown in Table 3. There was a statistically significant and positive correlation between youth who received substance use treatment and all six scales (correlations ranging from 0.21 to 0.32). The association was strongest for youth who received treatment and the youth total problems scale ($r(222) = 0.32$, $p < .01$), followed by the youth internalizing scale ($r(222) = 0.31$, $p < .01$), then youth externalizing scale ($r(222) = 0.21$, $p < .01$). These associations were between weak and moderate in strength, while the relationships with the caregiver scales were all weak (correlations between 0.21 and 0.29). The severity of a youth's substance use was also associated with responses on all three youth problem behavior scales. As a youth's substance use severity increased, there was an increase in internalizing, externalizing, and total problem behaviors. These correlations were between weak and moderate in strength (r ranging from 0.20 to 0.31), with the strongest association for youth internalizing behaviors ($r(222) = 0.31$, $p < .01$). For the last measure, family history of substance use problems was only significantly associated with externalizing behaviors. Specifically, increases in a youth's family history of substance use problems resulted in increases in youth externalizing behaviors ($r(222) = 0.18$, $p < .01$), although this relationship was weak.

Agreement across the youth and caregiver BPC reports was measured by examining the correlations between both caregiver and youth responses for the internalizing, externalizing, and total problems scales. As shown in Table 4, there was a statistically significant, positive correlation between youth and caregiver reports for all three scales. The association was strongest for agreement with externalizing behaviors ($r(222) = 0.48$, $p < .01$), followed by the total problems scale ($r(222) = 0.47$, $p < .01$), then internalizing behaviors ($r(222) = 0.43$, $p < .01$).

Table 4 also presents the means, standard deviations, and *t*-test results to understand the magnitude of differences between how youth and caregivers reported problem behaviors. Results from the *t*-tests indicate that caregivers reported more severe problems compared to youth across all three BPC scales. Specifically, caregiver scores were significantly higher than youth scores for internalizing, $t(222) = 6.25$, $p < .001$, externalizing $t(222) = 5.86$, $p < .001$, and the total problems scale, $t(222) = 7.52$, $p < .001$. Based on Cohen's *d*, the effect sizes ranged from small to medium (0.39 to 0.50). The largest effect size was identified for the total problems scale ($d = 0.50$), followed by internalizing problems ($d = 0.42$), then externalizing problems ($d = 0.50$).

Lastly, to assess agreement between youth and caregiver reports among the three BPC measures, we reported interclass correlation coefficients (ICCs). These results are also presented in Table 4. The ICC was 0.39 (95% CI, 0.234 to 0.516) for internalizing behaviors, 0.43 (95% CI, 0.286 to 0.554) for externalizing behaviors, and 0.41 (95% CI, 0.224 to 0.558) for the total problems scale. Using Koo and Li's (2016) guidelines, all three ICCs reflect poor agreement between youth and caregivers across the three BPC scales.

6. Discussion

The overall goal of this study was to examine youth and caregiver agreement on problem behavior as measured by a brief, feasible, and free tool, the Brief Problem Checklist. Our findings suggest that while the BPC is a reliable measure of problem behaviors, youth and their caregivers did not report a high degree of agreement. Caregivers tended to report that youth engaged in more severe internalizing and externalizing problems compared to youth self-reports. The implications of these findings are somewhat concerning, given the value placed on these types of assessments during case planning. Whether decision-makers place more weight on parent or youth reports and how these weighting decisions are made (e.g., practitioner discretion or agency policy) can have serious consequences for providing appropriate and timely intervention. Most juvenile justice agencies in the United States are working towards integrating the Risk, Need, Responsivity (RNR) framework into agency policies to address the intervention needs of justice-involved youths. The RNR framework requires the use of validated assessment practices to inform *who* receives treatment (Risk principle), *what* factors treatment targets (Need principle), and *how* treatment is delivered (Responsivity principle) (Bonta & Andrews, 2016). This move is largely based on overwhelming empirical evidence that RNR methods produce positive results for youth (e.g., reduce criminogenic needs, Baglivio et al., 2018) and public safety (e.g., recidivism, Brogan et al., 2015). However, the effectiveness of these methods is wholly based on the quality of the information used to carry out these methods.

Analyses demonstrated strong reliability for the total BPC and the two internalizing and externalizing subscales. Additionally, in examining convergent validity, our analyses suggested both youth and caregiver BPC scales were significantly related to theoretically relevant constructs. Specifically, all six BPC scales were significantly and positively correlated with youth who received SUD treatment in the last 90 days. That is, both caregivers and youth were more likely to report higher internalizing and externalizing problems if youth had received any form of treatment recently. Substance use severity was also positively associated with all three youth BPC scales, with youth with more severe SUDs reporting higher internalizing and externalizing problems. Lastly, youth with a family history of substance use problems reported higher externalizing problem behaviors. It is interesting that youth who were aware of a family history of substance use problems did not report higher internalizing symptoms, perhaps an indication of the youth's resilience. It also is interesting that none of the caregiver BPC scales were significantly related to the youth's SUD severity or the youth's report of family history of substance use problems. This may indicate a lack of relation between these phenomena, but it also may be an indication of the parent's lack of awareness of the youth's actual problem severity, whereas for a youth in treatment, the parent could potentially have greater awareness of the problems, resulting in helping the youth engage in recent treatment, or perhaps they became more aware of the extent of problems during the recent treatment experience.

In analyzing the correlations between youth and caregiver BPC reports, we identified statistically significant, positive correlations that were moderate in strength between youth and their caregivers on all three scales. The correlation coefficients detected in this study were higher than those identified in previous studies of cross-informant agreement on the

BPC (Chorpita et al., 2010). For example, Chorpita and colleagues (2010) reported cross-informant correlations for internalizing (r = 0.22), externalizing (r - 0.31), and total score (r = 0.19), while Achenbach and colleagues (1987) reported an average cross-informant correlation of 0.25. The current study identified higher cross-informant agreement across all three scales (ranging from 0.43 to 0.48). One might think that families who have youth engaged in the justice system would be more disconnected, but this may not be true. Further, when youth have become involved in the juvenile justice system, their parents have been provided some evidence of their behaviors and problems. It also may be the case that rural parents in rural communities have more insight into their teen's behaviors, although comparison with an urban sample is not possible in the present study. Regardless, the correlation coefficients remain lower than what would be preferred for identifying levels of fairly observable phenomena, especially for externalizing behaviors.

As would be expected, our correlation analyses followed similar trends found in previous literature with higher cross-informant correlations identified for externalizing behaviors compared to internalizing (Achenbach et al., 1987; Chorpita et al., 2010). Independent t-tests revealed on average, caregivers reported significantly higher scores for the total BPC scale as well as internalizing and externalizing behaviors. In fact, the greatest mean difference was identified for externalizing behaviors, with caregivers reporting the highest scores. This finding is not surprising, given external behaviors are those youth outwardly display and thus are more observable to their caregivers and is in line with previous research (Becker et al., 2004; Hope et al., 1999; Theunissen et al., 2019).

In assessing the level of agreement between youth and caregivers on the three BPC measures, we identified poor agreement. This is largely due to caregivers reporting significantly higher perceptions of engagement in both internalizing and externalizing problems. The observed poor agreement between youth and caregivers on the BPC brings up an important methodological question for consideration – in the assessment of youth problems, which perspective (youth or caregiver) matters, and for what purpose? We argue that our findings demonstrate the importance of assessing both perspectives (youth and caregiver) because caregivers may overestimate youth internal problems (e.g., feelings of guilt, inferiority, sadness) and youth may underestimate their external behaviors (e.g., argumentative, disobedient, stubborn). Thus, when BPC scores are used to evaluate behavioral health needs or treatment progress, both perspectives could be informative.

There are also several possible justifications for the misalignment of parent and youth ratings. Previous research examining the agreement between caregiver and youth report on adolescent substance use found greater disagreement when youth had greater involvement with probation/parole (McGillicuddy & Eliseo-Arras, 2012). Given the current sample of youth are all on probation, this could partly explain the poor agreement on the BPC identified here. That is, it is possible caregivers overestimate both internal and external problems because youth are already involved with the criminal justice system.

Some differences may simply be due to parents having a limited awareness of their child's behavior while unsupervised (e.g., McGillicuddy et al., 2007; McGillicuddy & Eliseo-Arras, 2012). Prior studies have also suggested that variations in correspondence among parent and

youth reports of youth functioning are not random (De Los Reyes & Kazdin, 2005; Piehler et al., 2020). For example, parent/youth discrepancies in internalizing problems are often found to be larger among girls and older adolescents, whereas inconsistencies in reporting externalizing problems are higher among boys and younger adolescents (Barker et al., 2007; Grills & Ollendick, 2002; van der Meer et al., 2008; Achenbach et al., 1987; De Los Reyes et al. 2015). Several parent-specific risk factors related to parental mental health, distress and trauma, parental monitoring, and family structure have also been found to correlate with lower rates of agreement (Berg-Nielsen et al., 2003; McGillicuddy et al., 2007; Sourander et al., 1999). Parents' ratings of their own behavior and distress are also associated with parent/youth discrepancies above and beyond characteristics of the youth (Shemesh et al., 2005; Kassam-Adams et al., 2006). These findings may signal an important inaccuracy in relying on parent reports of their child's behavior – the parents' own behavioral health needs. Additional research, using larger samples of youth with sufficient power, is needed to investigate meaningful patterns in (dis)agreement among youth and caregiver reports. Additionally, qualitative research examining caregiver and youth explanations of their responses may further illuminate drivers of disagreements between caregivers and youths. These types of studies are critical to making conclusions about the reliability of muti-informant assessment processes.

7. Strengths, limitations, and future research

Much research on rural youth is based on small samples, given the challenges associated with conducting research with this hard-to-reach population. However, a strength of the current study is a relatively large sample size of rural youth on probation (N = 222). Still though, for certain analyses, the sample would be considered small, and it represents a distinct population – rural youth on probation with a SUD. While the sample size may limit the generalizability of our findings, it is also imperative to study rural youth on probation, given justice-involved youth often lack access to evidence-based behavioral health services, especially in rural communities (Click et al., 2018; White et al., 2019). Examining this understudied population is particularly important as the use of validated screeners are not as common in justice settings as they are in clinical settings. Thus, identifying support for a brief, feasible, and free screener that can help probation agencies better identify youth needs, match them to services, and case plan is an important practical implication of this research.

The current study uses cross-sectional data, which cannot account for the potential changes in perspectives of youth problems behavior among caregivers and adolescents; however, the point of this study was to compare agreement at a single point in time. Future research should examine whether and how youth and caregiver perceptions of internalizing and externalizing behavior fluctuates over time. Additionally, we are unable to identify whether youth are accurately reporting their experiences and if caregivers are overreporting their perceptions. While future research should explore explanations for differential perceptions of internalizing and externalizing behaviors, we argue caregiver perceptions of problems are still informative in giving a holistic picture of adolescent problem behaviors. Given the youth in their care are currently justice-involved and have a SUD, caregiver perceptions of problems may be appropriate indicators of treatment needs and/or progress in treatment. Finally, although our findings provide initial support for the BPC as a feasible and free

tool to use, local validation is encouraged prior to its implementation. In particular, the reliability and validity of the BPC in other countries has not been examined. The significant differences in youths' behavioral health needs, access to effective treatment, parent/child interactions, cultural expectations, and juvenile justice system processes that exist around the world may lead to differences in the performance of the BPC across countries. Thus, more research on the reliability and validity of the BPC among international samples of justice-involved youth across different population densities (i.e., rural, suburban, urban) is needed.

7.1. Implications for research

For high-risk youth involved in the juvenile justice system, few things could be more consequential than determining the trajectory of a youth's life. Accurate and timely assessment of youth problem behaviors and needs is the first step in the decision-making process that can have significant consequences for youth and their families. The current study provided support for the reliability of the BPC scales but not inter-rater reliability of the scales. These findings lend merit to discussions about the need for more research on the reliability and validity of assessment instruments before their widespread use in guiding youth- and agency case planning decisions, along with informing conclusions about program effectiveness. There is a need for additional studies to address questions surrounding the extent to which instruments designed to measure youths' behavioral problems produce reliable and valid estimates. It is likely that the reliability and validity of most instruments vary according to the needs of the population being assessed, methods of assessment, availability of services, and environmental circumstances within which the assessment is completed.

7.2. Practical implications

Research finds relatively low rates of screening and assessment in justice-involved agencies (Bowser et al., 2019; Scott et al., 2019), particularly in community-based organizations that serve multi-need youth (Belenko et al., 2017). For example, a nationally representative study of juvenile probation agencies in the United States found that 53% screen for substance use and behavioral health problems, with even fewer using standardized assessment tools (Scott & Dennis, 2015). In rural settings, these challenges are exacerbated due to fewer resources (Ingoldsby, 2010) and a greater likelihood that youth may not be appropriately matched to treatment due to a lack of available services (Haqanee et al., 2015; Kapoor et al., 2018). While the BPC is a feasible tool, especially for programs low on resources – it is free, relatively simple (requiring only 12 questions for each respondent) and can produce reliable scores for a variety of problem behaviors – questions remain as to whether it is the *best* tool for use with this vulnerable population. More work is needed to understand the value of administering both the parent and youth versions of a brief tool such as the BPC, whether parent and/or youth reports are more beneficial to the goals of the juvenile justice intervention, and whether a brief tool such as the BPC can produce accurate assessments of behavioral problems among the high stakes decision-making processes for justice-involved youth from rural settings.

Acknowledgements

The authors extend their appreciation to participating youth, families, and juvenile probation officers.

Funding

This article was supported by funding from the National Institute on Drug Abuse (K23DA048161, R01DA041434). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Data availability

Data will be made available on request.

References

Achenbach TM, McConaughy SH, & Howell CT (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. Psychological Bulletin, 101(2), 213–213. [PubMed: 3562706]

Achenbach TM, & Rescorla LA (2001). Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Achenbach TM, & Rescorla LA (2014). The Achenbach system of empirically based assessment (ASEBA) for ages 1.5 to 18 years. In Maruish ME (Ed.), The use of psychological testing for treatment planning and outcomes assessment (pp. 179–214). Routledge.

Achenbach TM (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. Current Directions in Psychological Science, 15(2), 94–98.

American Academy of Child & Adolescent Psychiatry (2020). AACAP Practice Parameters. Retrieved from Practice Parameters (aacap.org) on August 17th, 2020.

Baglivio MT, Wolff KT, Howell JC, Jackowski K, & Greenwald MA (2018). The search for the holy grail: Criminogenic needs matching, intervention dosage, and subsequent recidivism among serious juvenile offenders in residential placement. Journal of Criminal Justice, 55, 46–57.

Baird C (2009). A question of evidence: A critique of risk assessment models used in the justice system. Madison, WI: National Council on Crime and Delinquency.

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

Barker ET, Bornstein MH, Putnick DL, Hendricks C, & Suwalsky JT (2007). Adolescent-mother agreement about adolescent problem behaviors: Direction and predictors of disagreement. Journal of Youth and Adolescence, 36(7), 950–962.

Bartko JJ (1991). Measurement and reliability: Statistical thinking considerations. Schizophrenia Bulletin, 17(3), 483–489. [PubMed: 1947873]

Becker A, Hagenberg N, Roessner V, Woerner W, & Rothenberger A (2004). Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? European Child & Adolescent Psychiatry, 13(2), ii17–ii24. [PubMed: 15243782]

Beidas RS, Stewart RE, Walsh L, Lucas S, Downey MM, Jackson K, & Mandell DS (2015). Free, brief, and validated: Standardized instruments for low-resource mental health settings. Cognitive and Behavioral Practice, 22(1), 5–19. [PubMed: 25642130]

Belenko S, Knight D, Wasserman GA, Dennis ML, Wiley T, Taxman FS, Oser C, Dembo R, Robertson AA, & Sales J (2017). The Juvenile Justice Behavioral Health Services Cascade: A new framework for measuring unmet substance use treatment services needs among adolescent offenders. Journal of Substance Abuse Treatment, 74, 80–91. [PubMed: 28132705]

Berg-Nielsen TS, Vika A, & Dahl AA (2003). When adolescents disagree with their mothers: CBCL-YSR discrepancies related to maternal depression and adolescent self-esteem. Child Care, Health and Development, 29(3), 207–213. [PubMed: 12752611]

Bethea A, Samanta D, Kali M, Lucente FC, & Richmond BK (2020). The impact of burnout syndrome on practitioners working within rural healthcare systems. The American Journal of Emergency Medicine, 38(3), 582–588. [PubMed: 31706660]

Bowser D, Henry BF, & McCollister KE (2019). An overlapping systems conceptual framework to evaluate implementation of a behavioral health intervention for justice–involved youth. Health services insights, 12, 1178632919855037.

Brogan L, Haney-Caron E, NeMoyer A, & DeMatteo D (2015). Applying the risk-needs-responsivity (RNR) model to juvenile justice. Criminal Justice Review, 40(3), 277–302.

Brown SA (1989). Life events of adolescents in relation to personal and parental substance use. American Journal of Psychiatry, 146, 484–489. [PubMed: 2929749]

Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, & Young SL (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. Frontiers in Public Health, 6, 149–167. [PubMed: 29942800]

Briggs SR, & Cheek JM (1986). The role of factor analysis in the development and evaluation of personality scales. Journal of Personality, 54(1), 106–148.

Chorpita BF, Reise S, Weisz JR, Grubbs K, Becker KD, & Krull JL (2010). Evaluation of the Brief Problem Checklist: Child and caregiver interviews to measure clinical progress. Journal of Consulting and Clinical Psychology, 78(4), 526–536. [PubMed: 20658809]

Chorpita BF, Daleiden EL, Park AL, Ward AM, Levy MC, Cromley T, Chiu AW, Letamendi AM, Tsai KH, & Krull JL (2017). Child STEPs in California: A cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. Journal of Consulting and Clinical Psychology, 85(1), 13–25. [PubMed: 27548030]

Cicchetti DV (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6, 284–290.

Click IA, Basden JA, Bohannon JM, Anderson H, & Tudiver F (2018). Opioid prescribing in rural family practices. Substance Use & Misuse, 53, 533–540. [PubMed: 28857643]

Connolly EJ, Lewis RH, & Boisvert DL (2017). The effect of socioeconomic status on delinquency across urban and rural contexts: Using a genetically informed design to identify environmental risk. Criminal Justice Review, 42(3), 237–253.

De Los Reyes A (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. Journal of Clinical Child & Adolescent Psychology, 40(1), 1–9. [PubMed: 21229439]

De Los Reyes A, & Kazdin AE (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. Psychological Bulletin, 131, 483–509. [PubMed: 16060799]

De Los Reyes A, Thomas SA, Goodman KL, & Kundey SM (2013). Principles underlying the use of multiple informants' reports. Annual review of Clinical Psychology, 9, 123–149.

De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DA, Burgers DE, & Rabinowitz J (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. Psychological Bulletin, 141(4), 858–900. [PubMed: 25915035]

Dembo R, Schmeidler J, Wareham J, & Faber J (2020). Gender and Urbanity Group Comparisons of the Impact of a Health Risk Factor on Juvenile Recidivism. American Journal of Criminal Justice, 47, 1–21.

Dennis ML, Titus JC, White MK, Unsicker JI, & Hodgkins D (2003). Global appraisal of individual needs: Administration guide for the GAIN and related measures. Bloomington, IL: Chestnut Health Systems.

Dimler LM, Natsuaki MN, Hastings PD, Zahn-Waxler C, & Klimes-Dougan B (2017). Parenting effects are in the eye of the beholder: Parent-adolescent differences in perceptions affects adolescent problem behaviors. Journal of Youth and Adolescence, 46(5), 1076–1088. [PubMed: 27848126]

Dir AL, Saldana L, Chapman JE, & Aalsma MC (2019). Burnout and mental health stigma among juvenile probation officers: The moderating effect of participatory atmosphere. Administration

and Policy in Mental Health and Mental Health Services Research, 46(2), 167–174. [PubMed: 30392147]

Ebesutani C, Bernstein A, Martinez JI, Chorpita BF, & Weisz JR (2011). The youth self report: Applicability and validity across younger and older youths. Journal of Clinical Child & Adolescent Psychology, 40(2), 338–346. [PubMed: 21391029]

Fazel S, Doll H, & Långström N (2008). Mental disorders among adolescents in juvenile detention and correctional facilities: a systematic review and metaregression analysis of 25 surveys. Journal of the American Academy of Child & Adolescent Psychiatry, 47(9), 1010–1019. [PubMed: 18664994]

Federal Office of Rural Health Policy (2021). Defining rural population. Health Resources & Services Administration. Retrieved from hrsa.gov/rural-health/about-us/what-is-rural.

Fehr KK, Leraas BC, & Littles MM (2020). Behavioral health needs, barriers, and parent preferences in rural pediatric primary care. Journal of Pediatric Psychology, 45 (8), 910–920. [PubMed: 32766670]

Ferdinand RF, van der Ende J, & Verhulst FC (2006). Prognostic value of parent–adolescent disagreement in a referred sample. European Child & Adolescent Psychiatry, 15(3), 156–162. [PubMed: 16424962]

Golzari M, Hunt SJ, & Anoshiravani A (2006). The health status of youth in juvenile detention facilities. Journal of Adolescent Health, 38(6), 776–782.

Grills AE, & Ollendick TH (2002). Issues in parent-child agreement: The case of structured diagnostic interviews. Clinical Child & Family Psychology Review, 5(1), 57–83. [PubMed: 11993545]

Grisso T, & Underwood L (2003). Screening and Assessing Mental Health and Substance Use Disorders Among Youth in the Juvenile Justice System. Policy Research Associates: National Center for Mental Health and Juvenile Justice.

Grisso T, Vincent G, & Seagrave D (Eds.). (2005). Mental health screening and assessment in juvenile justice. Guilford Press.

Guy LS, Nelson RJ, Fusco-Morin SL, & Vincent GM (2014). What do juvenile probation officers think of using the SAVRY and YLS/CMI for case management, and do they use the instruments properly? International Journal of Forensic Mental Health, 13(3), 227–241.

Haqanee Z, Peterson-Badali M, & Skilling T (2015). Making “what works” work: Examining probation officers’ experiences addressing the criminogenic needs of juvenile offenders. Journal of Offender Rehabilitation, 54(1), 37–59.

Hope TL, Adams C, Reynolds L, Powers D, Perez RA, & Kelley ML (1999). Parent vs. self-report: Contributions toward diagnosis of adolescent psychopathology. Journal of Psychopathology and Behavioral Assessment, 21(4), 349–363.

Hourigan SE, Goodman KL, & Southam-Gerow MA (2011). Discrepancies in parents’ and children’s reports of child emotion regulation. Journal of Experimental Child Psychology, 110(2), 198–212. [PubMed: 21458826]

Howell JC, Wilson JJ, Sickmund MH, Hodges NJ, & Howell MQ (2017). Caught in the act: States doing some things right in juvenile justice. Juvenile and Family Court Journal, 68(4), 25–42.

Huang C (2017). Cross-informant agreement on the Child Behavior Checklist for youths: A meta-analysis. Psychological Reports, 120(6), 1096–1116. [PubMed: 28675325]

Hughes EK, & Gullone E (2010). Discrepancies between adolescent, mother, and father reports of adolescent internalizing symptom levels and their association with parent symptoms. Journal of Clinical Psychology, 66(9), 978–995. [PubMed: 20694961]

Hunsley J, & Lee CM (2007). informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. Professional Psychology: Research and Practice, 38(1), 21.

Ingoldsby EM (2010). Review of interventions to improve family engagement and retention in parent and child mental health programs. Journal of child and family studies, 19(5), 629–645. [PubMed: 20823946]

Jensen EJ, Mendenhall T, Futoransky C, & Clark K (2021). Family Medicine Physicians’ Perspectives Regarding Rural Behavioral Health Care: Informing Ideas for Increasing Access to High-Quality Services. The Journal of Behavioral Health Services & Research, 1–12. [PubMed: 33230653]

Jones NJ, Brown SL, Robinson D, & Frey D (2016). Validity of the youth assessment and screening instrument: A juvenile justice tool incorporating risks, needs, and strengths. Law and Human Behavior, 40(2), 182. [PubMed: 26752015]

Kapoor A, Peterson-Badali M, & Skilling T (2018). Barriers to service provision for justice-involved youth. Criminal Justice and Behavior, 45(12), 1832–1851.

Kassam-Adams N, Garcia-Espana JF, Miller VA, & Winston F (2006). Parent-child agreement regarding children's acute stress: The role of parent acute stress reactions. Journal of the American Academy of Child & Adolescent Psychiatry, 45(12), 1485–1493. [PubMed: 17135994]

Klein DN, Dougherty LR, & Olino TM (2005). Toward guidelines for evidence-based assessment of depression in children and adolescents. Journal of Clinical Child and Adolescent Psychology, 34(3), 412–432. [PubMed: 16026212]

Krupa JM, Dembo R, Schmeidler J, Wolff J, & Wareham J (2021). An Examination of Risk Profiles among Justice-Involved Girls: Do Race and Place Matter? Deviant Behavior, 1–20. online first.

Lewis AJ, Bertino MD, Bailey CM, Skewes J, Lubman DI, & Toumbourou JW (2014). Depression and suicidal behavior in adolescents: A multi-informant and multi-methods approach to diagnostic classification. Frontiers in Psychology, 5(766), 1–9. [PubMed: 24474945]

Lipsey MW, & Howell JC (2012). A broader view of evidence-based programs reveals more options for state juvenile justice systems. Criminology & Public Policy, 11, 515–591.

Lyons JS (2009). Knowledge creation through total clinical outcomes management: A practice-based evidence solution to address some of the challenges of knowledge translation. Journal of the Canadian Academy of Child and Adolescent Psychiatry, 18 (1), 38. [PubMed: 19270847]

Mack K, & Rhineberger-Dunn G (2022). What matters most? Comparing the impact of individual, job, and organizational factors on job stress and job satisfaction among juvenile justice personnel. Criminal Justice Studies, 35(1), 18–37.

Martin JL, Ford CB, Dyer-Friedman J, Tang J, & Huffman LC (2004). Patterns of agreement between parent and child ratings of emotional and behavioral problems in an outpatient clinical setting: When children endorse more problems. Journal of Developmental & Behavioral Pediatrics, 25(3), 150–155. [PubMed: 15194898]

McCrae RR, Kurtz JE, Yamagata S, & Terracciano A (2011). Internal consistency, retest reliability, and their implications for personality scale validity. Personality and Social Psychology Review, 15(1), 28–50. [PubMed: 20435807]

McGillicuddy NB, Rychtarik RG, Morsheimer ET, & Burke-Storer MR (2007). Agreement between parent and adolescent reports of adolescent substance use. Journal of Child & Adolescent Substance Abuse, 16(4), 59–78. [PubMed: 19180245]

McGillicuddy NB, & Eliseo-Arras RK (2012). Parent–adolescent report correspondence on adolescent substance abuse among teens in residential rehabilitation. Addictive Behaviors, 37(4), 456–462. [PubMed: 22251413]

Miller LD, Martinez YJ, Shumka E, & Baker H (2014). Multiple informant agreement of child, parent, and teacher ratings of child anxiety within community samples. The Canadian Journal of Psychiatry, 59(1), 34–39. [PubMed: 24444322]

National Assembly on School-Based Health Care (2011). Summary of Free Assessment Measures, Center for School Mental Health. Retrieved from The National Assembly on School-Based Health Care (nasbhc.org) on August 17th, 2021.

National Center on Addiction and Substance Abuse. (2000). No place to hide: substance abuse in mid-size cities and rural America. New York.

National Research Council. (2013). Health and incarceration: A workshop summary.

Oser CB, Leukefeld CG, Staton Tindall M, Garrity TF, Carlson RG, Falck R, & Booth BM (2011). Rural drug users: Factors associated with substance abuse treatment utilization. International Journal of Offender Therapy and Comparative Criminology, 55(4), 567–586. [PubMed: 20463206]

Paris M, & Hoge MA (2010). Burnout in the mental health workforce: A review. The Journal of Behavioral Health Services & Research, 37(4), 519–528. [PubMed: 20013066]

Piedmont RL, & Hyland ME (1993). Inter-item correlation frequency distribution analysis: A method for evaluating scale dimensionality. Educational and Psychological Measurement, 53(2), 369–378.

Piehler TF, Lee SK, Stockness A, & Winters KC (2020). The correspondence of parent-reported measures of adolescent alcohol and cannabis use with adolescent-reported measures: A systematic review. Substance Abuse, 41(4), 437–450. [PubMed: 31809653]

Rescorla LA, Ginzburg S, Achenbach TM, Ivanova MY, Almqvist F, Begovac I, & Verhulst FC (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. Journal of Clinical Child & Adolescent Psychology, 42(2), 262–273. [PubMed: 23009025]

Rey JM, Schrader E, & Morris-Yates A (1992). Parent-child agreement on children's behaviours reported by the Child Behaviour Checklist (CBCL). Journal of Adolescence, 15(3), 219–230. [PubMed: 1447409]

Scott CK, & Dennis ML (2015). Juvenile Justice National Survey preliminary findings and juvenile justice community supervision agency report: Feedback for participating agencies. Chicago, IL: Chestnut Health Systems.

Scott T, Brown SL, & Skilling TA (2019). Predictive and convergent validity of the youth assessment and screening instrument in a sample of male and female justice-involved youth. Criminal Justice and Behavior, 46(6), 811–831.

Seiffge-Krenke I, & Kollmar F (1998). Discrepancies between mothers' and fathers' perceptions of sons' and daughters' problem behavior: A longitudinal analysis of parent–adolescent agreement on internalizing and externalizing problem behavior. The Journal of Child Psychology and Psychiatry and Allied Disciplines, 39(5), 687–697. [PubMed: 9690932]

Sheehan DV, Sheehan KH, Shytle RD, Janavs J, Bannon Y, Rogers JE, … Wilkinson B (2010). Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID). The Journal of clinical psychiatry, 71(3), 17393.

Shemesh E, Newcorn JH, Rockmore L, Shneider BL, Emre S, Gelb BD, Rapaport R, Noone SA, Annunziato R, Schmeidler J, & Yehuda R (2005). Comparison of parent and child reports of emotional trauma symptoms in pediatric outpatient settings. Pediatrics, 115(5). e582–e589. [PubMed: 15867023]

Shook JJ, & Sarri RC (2007). Structured decision making in juvenile justice: Judges' and probation officers' perceptions and use. Children and Youth Services Review, 29 (10), 1335–1351.

Silverman WK, & Ollendick TH (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. Journal of Clinical Child and Adolescent Psychology, 34(3), 380–411. [PubMed: 16026211]

Singh JP, & Fazel S (2010). Forensic risk assessment: A metareview. Criminal Justice and Behavior, 37(9), 965–988.

Sourander A, Helstelä L, & Helenius H (1999). Parent-adolescent agreement on emotional and behavioral problems. Social Psychiatry and Psychiatric Epidemiology, 34 (12), 657–663. [PubMed: 10703276]

Teplin LA, Abram KM, McClelland GM, Dulcan MK, & Mericle AA (2002). Psychiatric disorders in youth in juvenile detention. Archives of General Psychiatry, 59 (12), 1133–1143. [PubMed: 12470130]

Theunissen MH, de Wolff MS, & Reijneveld SA (2019). The strengths and difficulties questionnaire self-report: A valid instrument for the identification of emotional and behavioral problems. Academic Pediatrics, 19(4), 471–476. [PubMed: 30639760]

van der Ende J, Verhulst FC, & Tiemeier H (2012). Agreement of informants on emotional and behavioral problems from childhood to adulthood. Psychological Assessment, 24(2), 293–300. [PubMed: 21928910]

Van der Meer M, Dixon A, & Rose D (2008). Parent and child agreement on reports of problem behaviour obtained from a screening questionnaire, the SDQ. European child & adolescent psychiatry, 17, 491–497. [PubMed: 18431536]

Vincent GM, Guy LS, Perrault RT, & Gershenson B (2016). Risk assessment matters, but only when implemented well: A multisite study in juvenile probation. Law and Human Behavior, 40(6), 683–696. [PubMed: 27797547]

Wachter A (2015). Statewide risk assessment in juvenile probation. JJGPS StateScan. Pittsburgh, PA: National Center for Juvenile Justice.

Wasserman GA, McReynolds LS, Schwalbe CS, Keating JM, & Jones SA (2010). Psychiatric disorder, comorbidity, and suicidal behavior in juvenile justice youth. Criminal Justice and Behavior, 37(12), 1361–1376.

Weisz JR, Chorpita BF, Frye A, Ng MY, Lau N, Bearman SK, … Hoagwood KE (2011). Youth Top Problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. Journal of Consulting and Clinical Psychology, 79(3), 369–380. [PubMed: 21500888]

White LM, Aalsma MC, Salyers MP, Hershberger AR, Anderson VR, Schwartz K, Dir JL, & McGrew JH (2019). Behavioral health service utilization among detained adolescents: A meta-analysis of prevalence and potential moderators. Journal of Adolescent Health, 64(6), 700–708.

Weisz JR, Chorpita BF, Palinkas LA, Schoenwald SK, Miranda J, Bearman SK, … & Research Network on Youth Mental Health. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. Archives of General Psychiatry, 69(3), 274–282. [PubMed: 22065252]

White LM, Aalsma MC, Holloway ED, Adams EL, & Salyers MP (2015). Job-related burnout among juvenile probation officers: Implications for mental health stigma and competency. Psychological Services, 12(3), 291. [PubMed: 26009896]

Youngstrom E, Loeber R, & Stouthamer-Loeber M (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. Journal of Consulting and Clinical Psychology, 68(6), 1038–1050. [PubMed: 11142538]

Zaki R, Bulgiba A, Ismail R, & Ismail NA (2012). Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. PloS one, 7(5), Article e37908. [PubMed: 22662248]

Further reading

Achenbach TM, & Edelbrock C (1991). Child behavior checklist. Burlington (Vt), 7, 371–392.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). Standards for educational and psychological testing. American Educational Research Association.

Arslan G (2019). Development and validation of the Youth Externalizing Behavior Screener: A brief and effective measure of behavioral problems. International Journal of School & Educational Psychology, 7(sup1), 64–74.

Barber BK, Olsen JE, & Shagle SC (1994). Associations between parental psychological and behavioral control and youth internalized and externalized behaviors. Child Development, 65(4), 1120–1136. [PubMed: 7956469]

Bender HA, Auciello D, Morrison CE, MacAllister WS, & Zaroff CM (2008). Comparing the convergent validity and clinical utility of the Behavior Assessment System for Children-Parent Rating Scales and Child Behavior Checklist in children with epilepsy. Epilepsy & Behavior, 13(1), 237–242. [PubMed: 18448391]

Bonta J, & Andrews DA (2016). The psychology of criminal conduct (6th Edition). Routledge: New York.

Bourdon KH, Goodman R, Rae DS, Simpson G, & Koretz DS (2005). The Strengths and Difficulties Questionnaire: US normative data and psychometric properties. Journal of the American Academy of Child & Adolescent Psychiatry, 44(6), 557–564. [PubMed: 15908838]

Brumariu LE (2015). Parent–child attachment and emotion regulation. New Directions for Child and Adolescent Development, 2015(148), 31–45. [PubMed: 26086126]

Bögels SM, & van Melick M (2004). The relationship between child-report, parent self-report, and partner report of perceived parental rearing behaviors and anxiety in children and parents. Personality and Individual Differences, 37(8), 1583–1596.

Croft S, Stride C, Maughan B, & Rowe R (2015). Validity of the strengths and difficulties questionnaire in preschool-aged children. Pediatrics, 135(5). e1210–e1219. [PubMed: 25847804]

Duckworth AL, & Kern ML (2011). A meta-analysis of the convergent validity of self-control measures. Journal of Research in Personality, 45(3), 259–268. [PubMed: 21643479]

Flanagan DP, Alfonso VC, Primavera LH, Povall L, & Higgins D (1996). Convergent validity of the BASC and SSRS: Implications for social skills assessment. Psychology in the Schools, 33(1), 13–23.

Fulkerson JA, Pasch KE, Perry CL, & Komro K (2008). Relationships between alcohol related informal social control, parental monitoring and adolescent problem behaviors among racially diverse urban youth. Journal of community health, 33(6), 425–433. [PubMed: 18607698]

Gomez R, Vance A, & Gomez RM (2014). Analysis of the convergent and discriminant validity of the CBCL, TRF, and YSR in a clinic-referred sample. Journal of Abnormal Child Psychology, 42(8), 1413–1425. [PubMed: 24824190]

Goodman R (2001). Psychometric properties of the strengths and difficulties questionnaire. Journal of the American Academy of Child & Adolescent Psychiatry, 40 (11), 1337–1345. [PubMed: 11699809]

Gottfredson SD, & Moriarty LJ (2006). Statistical risk assessment: Old problems and new applications. Crime & Delinquency, 52(1), 178–200.

Goodman A, Lamping DL, & Ploubidis GB (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. Journal of Abnormal Child Psychology, 38(8), 1179–1191. [PubMed: 20623175]

Hunsley J, & Mash EJ (2007). Evidence-based assessment. Annual Review of Clinical Psychology, 3, 29–51.

Lanktree CB, Gilbert AM, Briere J, Taylor N, Chen K, Maida CA, & Saltzman WR (2008). Multi-informant assessment of maltreated children: Convergent and discriminant validity of the TSCC and TSCYC. Child Abuse & Neglect, 32(6), 621–625. [PubMed: 18584866]

Li Y, Babcock SE, Stewart SL, Hirdes JP, & Schwean VL (2021). August). Psychometric evaluation of the Depressive Severity Index (DSI) among children and youth using the interRAI Child and Youth Mental Health (ChYMH) assessment tool. Child & Youth Care Forum, 50(4), 611–630.

Merydith SP (2001). Temporal stability and convergent validity of the Behavior Assessment System for Children. Journal of School Psychology, 39(3), 253–265.

Morris AS, Silk JS, Steinberg L, Myers SS, & Robinson LR (2007). The role of the family context in the development of emotion regulation. Social development, 16 (2), 361–388. [PubMed: 19756175]

Nakamura BJ, Ebesutani C, Bernstein A, & Chorpita BF (2009). A psychometric analysis of the child behavior checklist DSM-oriented scales. Journal of Psychopathology and Behavioral Assessment, 31(3), 178–189.

Poythress NG, Douglas KS, Falkenbach D, Cruise K, Lee Z, Murrie DC, & Vitacco M (2006). Internal consistency reliability of the self-report antisocial process screening device. Assessment, 13(1), 107–113. [PubMed: 16443722]

Renshaw TL, & Cook CR (2018). Initial development and validation of the youth internalizing problems screener. Journal of Psychoeducational Assessment, 36(4), 366–378.

Robertson AA, Hiller M, Dembo R, Dennis M, Scott C, Henry BF, & Elkington KS (2019). National survey of juvenile community supervision agency practices and caregiver involvement in behavioral health treatment. Journal of child and family studies, 28(11), 3110–3120. [PubMed: 31749598]

Robinson EA, Eyberg SM, & Ross AW (1980). The standardization of an inventory of child conduct problem behaviors. Journal of Clinical Child & Adolescent Psychology, 9(1), 22–28.

Ron Nelson J, Benner GJ, Reid RC, Epstein MH, & Currin D (2002). The convergent validity of office discipline referrals with the CBCL-TRF. Journal of Emotional and Behavioral Disorders, 10(3), 181–188.

Salbach-Andrae H, Klinkowski N, Lenz K, & Lehmkuhl U (2009). Agreement between youth-reported and parent-reported psychopathology in a referred sample. European Child & Adolescent Psychiatry, 18(3), 136–143. [PubMed: 19129966]

Sieving RE, Beuhring T, Resnick MD, Bearinger LH, Shew M, Ireland M, & Blum RW (2001). Development of adolescent self-report measures from the National Longitudinal Study of Adolescent Health. Journal of Adolescent Health, 28 (1), 73–81.

Stanger C, & Lewis M (1993). Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. Journal of Clinical Child Psychology, 22(1), 107–116.

Steinberg L, Fletcher A, & Darling N (1994). Parental monitoring and peer influences on adolescent substance use. Pediatrics, 93(6), 1060–1064. [PubMed: 8197008]

Van Roy B, Veenstra M, & Clench-Aas J (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. Journal of Child Psychology and Psychiatry, 49(12), 1304–1312. [PubMed: 19120709]

Table 1

Description of Variables (N = 222).

Variables	Value	N	%
Youth Characteristics			
Sex	0 – Male	146	66
	1 – Female	76	34
Race	0 – White	202	91
	1 – Non-White	20	9
Ethnicity	0 – Non-Hispanic/Latino	124	56
	1 – Hispanic/Latino	98	44
Age	Number (low to high)	M = 15.37 SD = 1.44 Range = 11 – 18	
SUD[a] Treatment	0 – No	121	55
	1 – Yes	101	45
Severity of SUD	0 – Mild	58	26
	1 – Moderate	61	28
	2 – Severe	103	46
Education Program	0 – No	28	13
	1 – Yes	194	87
Caregiver Characteristics			
Sex	0 – Male	39	18
	1 – Female	183	82
Race	0 – White	201	91
	1 – Non-White	21	9
Ethnicity	0 – Non-Hispanic/Latino	143	64
	1 – Hispanic/Latino	79	36
Relationship to Youth	0 – Other	61	28
	1 – Mother	161	72
Age	Number (low to high)	M = 42.44 SD = 8.66 Range = 19 – 77	
Education	0 – Some High School	61	27
	1 – H.S. Degree/GED	64	29
	2 – Some College	57	26
	3 – AA/College/Graduate Degree	40	18
Employment Status	0 – Not Employed	55	25
	1 – Employed	167	75
History of Substance Abuse	Number (low to high)	M = 2.57 SD = 1.10 Range = 2–10	
BPC Scales[b]			
Youth – Internalizing	Number (low to high)	M = 3.30 SD = 3.08 Range = 0 – 12	

Variables	Value	N %
Youth – Externalizing	Number (low to high)	M = 4.08 SD = 2.58 Range = 0 – 11
Youth – Total Problems	Number (low to high)	M = 7.38 SD = 4.66 Range = 0 – 19
Caregiver – Internalizing	Number (low to high)	M = 4.69 SD = 3.12 Range = 0 – 12
Caregiver – Externalizing	Number (low to high)	M = 5.27 SD = 3.25 Range = 0 – 12
Caregiver – Total Problems	Number (low to high)	M = 9.96 SD = 5.25 Range = 0 – 22

[a]SUD = Substance Use Disorder.

[b]BPC = Brief Problem Checklist M = Mean; SD = Standard Deviation.

Table 2

Reliability Analyses for the Brief Problem Checklist Scales for Youth and Caregivers.

	Internalizing		Externalizing		Total Problems	
	Youth	Caregiver	Youth	Caregiver	Youth	Caregiver
Cronbach's Alpha	0.83	0.81	0.74	0.85	0.81	0.84
Average IIC	0.44	0.41	0.33	0.48	0.26	.30

IIC = Inter-Item Correlations.

N = 222.

Table 3

Correlations between Youth and Caregiver Substance Use Measures and the Brief Problem Behavior Checklist Scales.

		1	2	3	4	5	6	7	8	9
1.	SUD[a] Treatment	–								
2.	Severity of SUD[a]	0.15*	–							
3.	History of Substance Abuse	−0.01	−0.12	–						
4.	Youth – Internalizing	0.31***	0.31***	0.04	–					
5.	Youth – Externalizing	0.21***	0.20***	0.18**	0.35***	–				
6.	Youth – Total	0.32***	0.29***	0.12	0.86***	0.79***	–			
7.	Caregiver – Internalizing	0.26***	0.13	0.01	0.43***	0.21***	0.40***	–		
8.	Caregiver – Externalizing	0.21***	0.03	0.07	0.17*	0.48***	0.38***	0.36***	–	
9.	Caregiver – Total	0.29***	0.09	0.05	0.36***	0.43***	0.47***	0.82***	0.83***	–

N = 222

[a]SUD = Substance Use Disorder

*p <.05,

**p <.01,

***p <.00.

Table 4

Youth and Caregiver Correlations between Internalizing, Externalizing, and Total Problems Scales.

	Youth		Caregiver					
	Mean	**SD**	**Mean**	**SD**	***r***	***t***	***d***	**ICC**
Internalizing Scale	3.30	3.08	4.69	3.12	0.43**	6.25***	0.42	0.39***
Externalizing Scale	4.08	2.58	5.27	3.25	0.48**	5.86***	0.39	0.43***
Total Problems Scale	7.38	4.66	9.96	5.25	0.47**	7.52***	0.50	0.41***

N = 222.

SD = Standard Deviation; ICC = Interclass Correlation Coefficients

* $p < .05$,

** $p < .01$,

*** $p < .001$.