Published in final edited form as:

Int J Forensic Ment Health. 2012 ; 11(3): 150–164. doi:10.1080/14999013.2012.737405.

Pilot Implementation and Preliminary Evaluation of START:AV Assessments in Secure Juvenile Correctional Facilities

Dr. Sarah L. Desmarais,

North Carolina State University, Psychology, Raleigh, NC United States

Brian G Sellers,

University of South Florida, Tampa, FL, United States

Dr. Jodi L. Viljoen,

Simon Fraser University, Burnaby, British Columbia, Canada

Dr. Keith R. Cruise,

Fordham University, Bronx, NY United States

Dr. Tonia L. Nicholls, and

University of BC / BC Mental Health and Addiction Services, Coquitlam, British Columbia, Canada

Dr. Joel A. Dvoskin

University of Arizona College of Medicine, Psychiatry, Tucson, United States

Sarah L. Desmarais: sdesmarais@ncsu.edu

Abstract

The Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV) is a new structured professional judgment guide for assessing short-term risks in adolescents. The scheme may be distinguished from other youth risk assessment and treatment planning instruments by its inclusion of 23 dynamic factors that are each rated for both vulnerability and strength. In addition, START:AV is also unique in that it focuses on multiple adverse outcomes—namely, violence, self-harm, suicide, unauthorized leave, substance abuse, self-neglect, victimization, and general offending—over the short-term (i.e., weeks to months) rather than long-term (i.e., years). This paper describes a pilot implementation and preliminary evaluation of START:AV in three secure juvenile correctional facilities in the southern United States. Specifically, we examined the descriptive characteristics and psychometric properties of START:AV assessments completed by 21 case managers on 291 adolescent offenders (250 boys and 41 girls) at the time of admission. Results provide preliminary support for the feasibility of completing START:AV assessments as part of routine practice. Findings also highlight differences in the characteristics of START:AV assessments for boys and girls and differential associations between the eight START:AV risk domains. Though results are promising, further research is needed to establish the reliability and validity of START:AV assessments completed in the field.

Keywords

START:AV; structured professional judgment; psychometric properties; juvenile justice; risk assessment

More than 81,000 young offenders reside in more than 2,400 juvenile justice facilities in the United States (Hockenberry, Sickmund, & Sladky, 2011). These youths are typically high risk and high need; they are not only at risk of violence and offending, but also disproportionate service utilization for mental health, substance use and physical health needs, and elevated rates of victimization and future involvement in the adult justice system (Ford, Hartman, Hawke, & Chapman, 2008; Reza, Krug, & Mercy, 2001; Slowikowski, 2009; Snyder, 2004; Wasserman & McReynolds, 2006). To effectively match justice-involved youths to appropriate interventions, case managers and allied professionals working in juvenile justice settings assess adolescents' risks and needs via the use of risk assessment instruments (Borum & Verhaagen, 2006). Consequently, courts in the United States increasingly require the use of risk assessment tools in secure facilities serving justice-involved youths to ensure evidence-informed practice (e.g., Annie E. Casey Foundation, 2000; Austin, Johnson, & Weitzer, 2005; Juvenile Justice and Delinquency Prevention Act, 2002; Memorandum of Agreement, 2008; U.S. v. Georgia, 1998). As a result, the use of structured risk assessment instruments in state juvenile correctional facilities has nearly doubled in recent years (Griffin & Bozynski, 2003).

Many instruments are available for assessing risks in justice-involved youths. Findings support their reliability and validity in predicting recidivism and violence (e.g., Olver, Stockdale, & Wormith, 2009; Schwalbe, 2007, 2008; Singh, Grann, & Fazel, 2011; Welsh, Schmidt, McKinnon, Chattha, & Meyers, 2008); however, their use in secure juvenile correctional facilities has been examined infrequently. Recent meta-analyses conducted by Schwalbe (2007, 2008) found that more than two-thirds of youth risk assessment studies focused on probation or community-based samples. For instance, the reliability and validity of assessments completed using the Youth Level of Service/Case Management Inventory (YLS/CMI; Hoge & Andrews, 2002), Psychopathy Checklist: Youth Version (PCL:YV; Forth, Kosson, & Hare, 2003), and Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006) have been shown to predict both violent and general recidivism, even when completed retrospectively through review of clinical-legal records (Catchpole & Gretton, 2003). Risk assessments completed using the SAVRY also have exhibited good validity in predicting violent recidivism among samples drawn from mental health settings (Meyers & Schmidt, 2008) and secondary educational settings (McGowan, Horn, & Mellott, 2011). To our knowledge, only one study has examined the field reliability of SAVRY assessments: Vincent, Guy, Fusco, and Gershenson (2012) found substantial inter-rater agreement (81%) between SAVRY summary risk ratings completed by juvenile probation officers.

Most research examining risk assessments in secure juvenile correctional facilities has focused on accuracy vis-à-vis the prediction of *sexual* offending (Viljoen, Mordell, & Beneteau, 2012). In one exception, a recent study investigated risk assessments completed using the SAVRY of 66 boys receiving treatment in a residential, correctional facility in the Netherlands (Lodewijks, Doreleijers, de Ruiter, & Borum, 2008). In this first prospective study of SAVRY assessments in institutional settings, analyses revealed excellent validity in predicting incidents of physical violence against others during residential treatment. Another study recently investigated how risk assessments might predict future violence (sexual and nonsexual) among 169 boys in a residential facility for adolescent sex offenders (Viljoen et al., 2008). In this study, research assistants completed the Juvenile Sex Offender Assessment Protocol-II (J-SOAP-II; Prentky & Righthand, 2003), Juvenile Sexual Offense Recidivism Risk Assessment Tool-II (J-SORRAT-II; Epperson, Ralston, Fowers, & DeWitt, 2006), and SAVRY through comprehensive file reviews. Results showed that SAVRY and J-SOAP-II assessments predicted nonsexual aggression during treatment and serious nonsexual violence after release.

Despite these promising findings, some important limitations remain. In particular, there is a relative lack of consideration of short-term outcomes or protective factors in existing instruments, though the SAVRY does include a short protective factors subscale. Because adolescence is a period of transition and formative change, a focus on short-term outcomes may be particularly relevant to the assessment and management of adolescents (Borum & Verhaagen, 2006; Scott & Steinberg, 2008). Furthermore, as reviewed earlier, youths at risk for violence and offending often are at increased risk for other adverse outcomes (Ford et al., 2008; Snyder, 2004; Wasserman & McReynolds, 2006) that are not addressed by existing risk assessment measures. The Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV; Nicholls, Viljoen, Cruise, Desmarais, & Webster, 2010) was developed to address these limitations.

Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV)

The START:AV represents an adaptation of the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009). Briefly, the START is a structured professional judgment (SPJ) scheme that guides the assessment of a variety of short-term risks among adults with behavioral health disorders. The START may be distinguished from other risk assessment tools in several important ways. First, the START guides the assessment of vulnerabilities *and* strengths for every item. Second, many instruments emphasize static factors, such as gender or history of abuse (e.g., the Violence Risk Appraisal Guide or VRAG, Quinsey, Harris, Rice, & Cormier, 2006). In contrast, although historical information is used as the foundation of a START assessment, *all* START items are potentially dynamic in nature. Third, although instruments typically focus on identifying factors associated with risk for violence or recidivism, the START guides a comprehensive assessment of risk for violence and related adverse outcomes, including suicide, self-harm, victimization, substance use, unauthorized leave, and self-neglect. Fourth and finally, the START assesses risk over a short-term (i.e., weeks to months) rather than long-term (i.e., months to years).

Through review of the empirical and practice literature, as well as consultation with experts, the START scheme was modified according to three principles to increase its applicability to adolescents (see Viljoen, Cruise, Nicholls, Desmarais, & Webster, 2012, for further details regarding the adaptation process). First, though adolescents and adults may share common risk and protective factors, there also may be important differences in the relevance of certain factors across the lifespan. For example, peers appear to be particularly relevant to risk behaviors among adolescents (e.g., Lipsey & Derzon, 1998). Thus, START:AV items and outcome domains were modified, as necessary, to ensure inclusion of developmentally-relevant factors, such as relationships with and social support from peers. To demonstrate, we adjusted item anchors and disaggregated two START items – Relationships and Social Support – into sub-items for caretakers/other adults and peers. Second, adolescents are embedded within multiple systems (e.g., peer networks, families, schools) (Bronfenbrenner, 1979) and are less able to change their environments than are adults (Fagan, 2000). As such, caregivers and home environments may play an especially important role in determinations of risk and figure prominently in the START:AV assessment process. For this reason, we added a new item specifically designed to assess the quality of the parenting and home environment. Third, adolescents experience diverse developmental changes (biological, physical, psychosocial), and expectations and acceptability of various behaviors will change over the course of adolescence (Scott & Steinberg, 2008). Accordingly, START:AV assessments are anchored on age-appropriate norms such that what constitutes “0”, “1” or “2” ratings for strengths and vulnerabilities may change over adolescence. Another substantive change was the addition of an eighth risk domain: risk for general offending,

defined as illegal behavior not included in the other risk domains (e.g., trespassing, theft) and explicitly excluding violent offending (which would be captured under the violence risk estimate). The addition of this eighth risk domain reflects the need to attend to risk for both violent and non-violent offending in justice-involved youths, recognition that non-violent offending appears to share predictors with the other risk domains, and parallels a similar addition to the START under consideration (see Desmarais, Van Dorn, Telford, Petrila, & Coffey, 2012).

The Present Study

Practically speaking, the START:AV is a 1-page assessment tool comprised of 23 dynamic items. Each item is rated for both strength and vulnerability regarding the youth's functioning in the past two to three months, based upon all available information. Strength and vulnerability ratings are scored independent of one another, and a youth may be scored high (or low) on both strength and vulnerability for any particular item. For example, a youth may receive a high vulnerability rating for Relationships with Peers (item 2a) if s/he is involved in an abusive dating relationship, but also may receive a high strength rating if s/he has a warm, supportive, and reciprocal relationship with prosocial peers. As another example, and consistent with a harm reduction perspective, a youth may receive a high vulnerability rating for Substance Use (item 8) if s/he is currently abusing drugs, but also may receive a moderate strength rating if s/he recognizes the negative effects of the drug abuse and the need for treatment. Assessors use these item ratings, as well as key and critical items (i.e., strengths and vulnerabilities, respectively, that are particularly relevant, either recently or historically) and historical factors, to estimate the youth's risk for each outcome domain over the next three months.

Though the START:AV closely resembles the adult version of the START, research is essential to ensure START:AV assessments are similarly reliable and valid. Also, little is known about the characteristics and psychometric properties of assessments completed in the context of 'real world' implementations of structured risk tools, START:AV or otherwise, rather than for research purposes (e.g., Crocker et al., 2011; Douglas, Otto, Desmarais, & Borum, in press; Nicholls, Petersen, Brink, & Webster, 2011; Vincent et al., 2012). Moreover, there are few published papers documenting the process through which risk assessments instruments were implemented, though the process is critical to implementation success (Nonstad & Webster, 2011). Finally, few studies have examined the characteristics of risk assessments completed on female samples using structured risk instruments (Garcia-Mansilla, Rosenfeld, & Nicholls, 2009). Though a recent study showed that predictive validity of assessments completed using the SAVRY were comparable for boys and girls (Penney, Lee, & Moretti, 2010), another study found better predictive validity for SAVRY, PCL:YV, and YLS/CMI assessments of boys than girls (Schmidt, Campbell, & Houlding, 2011; see Olver et al., 2009 and Schwalbe, 2008 for meta-analytic reviews). Indeed, understanding the risk profiles of youths in secure correctional facilities is critical to risk management and treatment efforts (Gammelgård, Weizmann-Henelius, Koivisto, Eronen, & Kaltiala-Heino, 2012; Nicholls et al., 2011). Thus, research examining characteristics of START:AV assessments of boys and girls will inform the continued development and use of the START:AV, as well as the field more broadly.

To address these empirical gaps, this paper describes the pilot implementation and preliminary evaluation of START:AV assessments completed in three secure juvenile correctional facilities in the southern United States. We present data on the descriptive characteristics (e.g., number of missing ratings, distribution of item ratings, total scores and risk estimates, and number of key and critical items endorsed) and psychometric properties (e.g., structural reliability, and associations between item ratings, total scores and risk

estimates) of START:AV assessments completed at the time of admission. Our study objectives were twofold: first, examine the feasibility, reliability and integrity of START:AV assessments completed in the field; and second, inform the ongoing development and refinement of the START:AV. We sought to answer three specific research questions:

1. What are the descriptive characteristics of START:AV assessments completed for youths at the time of admission to one of three secure juvenile correctional facilities?
2. Do START:AV assessments have adequate psychometric properties when completed by case managers in the field?
3. Do the characteristics of START:AV assessments differ significantly for boys and girls?

Methods

Implementation of the START:AV

The START:AV was implemented in November 2010 in three secure juvenile correctional facilities under one state agency in the southern United States. Youths are placed at these facilities as part of disposition for adjudicated juvenile court offenses. Two of the facilities house adjudicated boys and one houses adjudicated girls. The latter has a capacity for 60 adjudicated girls between the ages of 12 and 18 years, and in 2011, admitted 226 girls (including repeat admissions). The two other facilities have a combined capacity for over 300 adjudicated boys, and house boys between the ages of 12 and 15 years, and 16 years and older, respectively. Together, these two facilities admitted 1,515 boys in 2011 (including repeat admissions). In addition to educational services and vocational training, services offered at all three sites include medical, mental health and addictions services, as well as specialized programs (e.g., sex offender treatment programming).

The agency's decision to implement the START:AV in their secure juvenile correctional facilities reflected the confluence of several factors. First, as introduced earlier, juvenile justice facilities in the United States increasingly are required to assess youths using structured risk assessment tools. Second, the oversight agency for these three facilities was acting upon the recommendations of an external evaluation to increase uniformity and accuracy of the assessment processes across sites. Before implementation of the START:AV, each facility followed their own intake assessment protocols, none of which included a structured risk assessment instruments. Third, for several years these facilities had struggled to make an organizational shift from a focus on *risk* assessment and *risk* management to *strength*-based case management. With its inclusion of strength ratings for each factor, the START:AV was seen as a tool that might assist in this process. Fourth, there was recognition that youths admitted to these facilities, as seen in other jurisdictions, increasingly were presenting with complex behavioral health problems associated with more than just risk for violence and recidivism, but assessments primarily focused on risks to others. Fifth, case managers represent diverse professional backgrounds, including but not limited to training in social work, criminology, psychology and rehabilitation, and often lacked a common language with which to communicate regarding their cases. Finally, agency officials and staff saw the formative state of the START:AV and the pilot nature of the implementation as a unique and exciting opportunity to assist in the development and refinement of a new risk assessment approach.

Following several discussions with the agency's administrators, counsel, treatment coordinator and campus administrators, and in consultation with a third-party expert (JD),

the START:AV was implemented across the three facilities to inform the development and amendment of individualized service plans (ISPs). The implementation plan was designed to complement existing case management practices, such that START:AV assessments were to be completed on new admissions and current residents, as appropriate. In other words, case managers were not required to complete START:AV assessments for current residents until the time at which the case would normally be reviewed. Specifically, START:AV assessments were to be completed based upon all available information, including interviews with the youths and collaterals and clinical-legal records, at four different times: 1) within 21 days of admission (with the exception of new admissions to an intensive treatment unit) to inform the development of the ISP; 2) every three months during each youth's quarterly case management review to monitor their progress and to inform amendments to the ISP, if necessary; 3) when there was a request for or anticipated change in restriction level (e.g., in preparation for release or transfer to a different unit); and 4) should there be an important change in a youth's well-being or circumstances (e.g., death in the family). Implementation followed an initial, one-day training workshop and included regularly scheduled booster training sessions (described in further detail below).

Sample

START:AV assessments were completed for 291 adolescent offenders (250 boys and 41 girls) at the time of their admission to one of three secure juvenile correctional facilities ($n = 115$, $n = 135$, and $n = 41$, respectively). At the time of their START:AV assessments, youths were 16 years old on average ($M = 16.16$, $SD = 1.25$, Range = 13 to 20). Average age did not differ between boys and girls ($M = 16.16$, $SD = 1.23$; $M = 16.21$, $SD = 1.38$), $t(255) = 0.25$, $p = .80$. Almost half of the sample had a history of substance abuse (44.5%) and more than one quarter had a history of perpetrating violence (28.6%). Boys and girls did not differ significantly with respect to the prevalence of substance use (45.8% vs. 36.6%), $\chi^2(1, N = 290) = 1.21$, $p = .27$, though there was a trend for more boys than girls to have histories of violence (30.5% vs. 17.1%), $\chi^2(1, N = 290) = 3.12$, $p = .08$. Mental health diagnoses were rarely available at the time of admission, but, when noted based upon available file information (17.9% of the sample, $n = 52$), commonly included anxiety disorder (9.6%, $n = 5$), depression (23.1% $n = 12$), attention deficit hyperactivity disorder (38.5%; $n = 20$), conduct disorder (28.8%, $n = 20$), oppositional defiance disorder (21.2%, $n = 11$), and/or substance abuse disorder (40.4% $n = 21$). Co-occurring mental and substance use disorders were noted for 38.5% ($n = 20$) of these youths.

Measure

The START:AV is a SPJ guide for the assessment of risk of violence to others, self-harm, suicide, unauthorized leave, substance abuse, self-neglect, being victimized, and general offending in youths between 13 and 18 years of age. Its 23 dynamic factors are coded for both strengths and vulnerabilities evidenced in the past two to three months on a 3-point ordinal scale from 0 (*minimally present*) to 2 (*maximally present*). For each item, strength and vulnerability are rated independently of one another; that is, a youth may be scored high (or low) on strength and vulnerability for any particular item. Assessors estimate risk over the next three months as *low*, *moderate*, or *high* for each of the eight START:AV domains. These specific risk estimates reflect structured professional judgments based on the strength and vulnerability ratings, as well as the presence of key items (i.e., particularly relevant strengths) and critical items (i.e., particularly relevant vulnerabilities), physical health problems, and other historical factors (e.g., history of perpetrating violence or of being victimized by others).

Strength and vulnerability total scores were calculated for research purposes by summing the item ratings (possible range = 0 to 46).¹ When there were five or fewer missing item

ratings, we followed the instructions for prorating the strength and vulnerability total scores provided in the START manual, version 1.1 (see Webster et al., 2009, p. 34). Specifically, total scores were prorated by summing the item ratings, dividing by the total possible scale score (i.e., 46), multiplying that amount by the number of omitted items and adding that to the original total: prorated total score = [(raw total score / 46) × number of missing items] + raw total score. Total scores were not prorated for assessments missing more than five item ratings, which led to the exclusion of two and three START:AV assessments from analyses examining strength and vulnerability total scores, respectively.

Study Procedures

Training—Prior to the START:AV implementation, case managers ($n = 28$), as well as the treatment coordinator (the case managers' direct supervisor) and the clinical services liaison, attended a one-day training workshop presented by one of the START:AV authors (SLD). The vast majority of case managers had very limited experience in risk assessment and all were unfamiliar with the SPJ approach. The workshop introduced the general principles of the SPJ approach to risk assessment (approximately one hour), operationalized the START:AV assessment framework (approximately two hours), and included completion of a practice case in small groups that was reviewed with the trainer to ensure coding accuracy and adequate agreement (approximately three hours). The training workshop concluded with an hour-long discussion of the specifics of the START:AV implementation within this agency (e.g., implementation plan, resources, expectations, etc.). Three half-day booster training workshops were delivered by the same trainer every three to six months ($M = 128.33$ days between training workshops, $SD = 40.61$, Range = 101 to 175) during the first year of the implementation. During the first two hours of these sessions, the trainer reviewed the purpose, goals, and status of the START:AV implementation; provided feedback regarding the completion of START:AV assessments (e.g., frequently missing ratings and other common coding errors), as well as summaries of the START:AV assessment profiles for boys and girls; and reviewed definitions and anchors. During the third hour, case managers from each facility completed a START:AV assessment for a youth from that site and then shared the results of the assessment with the larger group. Group discussion focused on how the ISPs could be best tailored to the identified strengths, vulnerabilities, and specific risk levels, as well as on coding accuracy. All training was conducted using the START:AV abbreviated manual (Nicholls et al., 2010).

Assessments—The START:AV implementation was rolled out within two weeks of the initial training. Between November 2010 and November 2011, 391 START:AV assessments were completed for 291 adolescent offenders. Herein we focus on 291 unduplicated admission assessments; that is, START:AV assessments completed by case managers ($n = 21$) for new, unduplicated admissions based upon all available information, typically interviews with the youths and their legal records, to inform development of the individualized service plans. (At the time of admission, the opportunity to conduct collateral interviews was limited and few clinical records were available beyond a brief referral form detailing the index offense.) On average, each case manager completed approximately eight assessments ($M = 8.24$, $SD = 0.58$; Range = 1 to 35) included in the present sample.

Data Analysis

We first examined the mean (standard deviation), median and modal number, as well as the range, of missing ratings for START:AV items and risk estimates. We then calculated descriptive statistics for the strength and vulnerability item ratings and total scores, and risk

¹Though calculation of total scores may be informative for research purposes, total scores are not calculated for use in practice, consistent with the SPJ approach.

estimates, and examined the mean (standard deviation), median and modal number of key items and critical items. We additionally calculated correlations between risk estimates to explore the co-occurrence of diverse short-term risks and treatment needs (e.g., self-harm and general offending) among adolescent offenders. Finally, we conducted *t*-test comparisons and chi-square analyses, as appropriate, to examine whether missing data, item ratings, key and critical items, total scores, or risk estimates differed between boys and girls.

To examine structural reliability, we computed Cronbach's coefficient alpha to measure internal consistency of strength and vulnerability total scores. Values greater than .80 indicated good internal consistency (Nunnally & Bernstein, 1994). We also calculated mean inter-item correlations (MICs) for the strength and vulnerability total scores. Values between .20 and .50 reflected unidimensionality (Nunnally & Bernstein, 1994). Mean corrected item-total correlations (CITCs) were calculated to examine strength of association between item and total scores. Values greater than .30 were considered acceptable (Nunnally & Bernstein, 1994). We calculated correlations between: 1) strength and vulnerability item ratings; 2) strength and vulnerability total scores; 3) item ratings and risk estimates, and 4) total scores and risk estimates. Coefficients of .10 reflected a small association, .30 moderate, and .50 large (Cohen, 1988).

Human Subjects

All study procedures were approved by the University of South Florida's Institutional Review Board. Because we only accessed de-identified secondary data and did not have direct contact with the sample, participant informed consent was not required.

Results

Preliminary analyses showed no systematic differences in the descriptive characteristics or psychometric properties of assessments as a function of assessor or facility (for boys). Thus, in the subsequent sections we present findings collapsed across assessors and facilities.

Descriptive Characteristics

Missing item and risk estimates—Table 1 presents the descriptive characteristics for missing strength and vulnerability ratings, as well as risk estimates, overall and for boys and girls. Across sites, assessors left, on average, one strength rating ($M = 0.91$, $SD = 1.84$, $Mdn = 0.00$, $Mode = 0$) and one vulnerability rating ($M = 0.87$, $SD = 1.67$, $Mdn = 0.00$, $Mode = 0$) blank per assessment, though the number of missing item ratings ranged widely, from 0 to 23 and 0 to 23, respectively. Assessors were not significantly more likely to leave strength or vulnerability ratings blank, $t(290) = 0.67$, $p = .51$. On average, three risk estimates ($M = 2.92$, $SD = 2.40$, $Mdn = 0.00$, $Mode = 0$) were left blank per assessment, with the number of missing risk estimates ranging from 0 to 8. With respect to the specific risk domains, 42.6% of START:AV assessments were missing risk estimates for violence; 60.5% for self-harm; 62.5% for suicide; 25.4% for unauthorized leave; 11.0% for substance abuse; 32.5% for self-neglect; 31.6% for victimization; and 25.8% for general offending. Further analyses revealed that the number of missing risk estimates decreased over the course of the implementation, $r = -.14$, $p = .03$. No such changes were seen in the number of missing strength or vulnerability ratings, $r^2s < .05$, $p^2s < .36$.

Analyses of missing data by gender revealed some differences. Specifically, assessors were more likely to leave vulnerability than strength ratings blank for girls, $t(40) = 2.48$, $p = .02$, but there were no significant differences in the number of vulnerability or strength ratings left blank for boys, $t(249) = 1.55$, $p = .12$. Assessors were more likely to leave items blank – both strength and vulnerability ratings – for boys than girls, $t(289) = 3.09$, $p = .002$ and

$t(289) = 1.89, p = .06$, respectively. The mean number of risk estimates left blank for girls and boys did not differ significantly, $t(289) = 1.07, p = .29$.

Item ratings—Table 1 displays mean item ratings, endorsement frequencies, and correlations between strength and vulnerability ratings. For all items, assessors made use of the full range of scores (from 0 to 2) for both strength and vulnerability ratings. With the exception of Material Resources, associations between strength and vulnerability ratings were moderate to strong and in the expected direction, ranging in size from small for Self-Care to large for School/Work, for instance (see Table 1). Comparison of item ratings between boys and girls revealed several items upon which girls received higher vulnerability ratings than boys: Relationships with Adults, Recreation, Emotional State, Impulse Control, Social Support from Adults, Treatability, and Parenting/Home Environment, $t's(275 \text{ to } 286) = 2.41, p's \leq .02$. There was one item, Coping, for which boys received higher vulnerability ratings than girls, $t(282) = 3.57, p < .001$. Mean strength ratings differed between boys and girls for only one item: boys received higher strength ratings on Recreation than did girls, $t(283) = 3.39, p < .001$.

Overall, 95.6% ($n = 281$) of the sample was identified as having at least one strength, rated as *moderately* or *maximally present*, and just over half (57.4%, $n = 167$) as having at least one strength rated as *maximally present*. A similar number of youths (95.5%, $n = 278$) were identified as having at least one vulnerability *moderately* or *maximally present*, but more than three-quarters (75.6%, $n = 220$) were identified as having at least one vulnerability *maximally present*. The mean number of strengths and vulnerabilities identified as *moderately* or *maximally present* were 15.61 ($SD = 6.51$) and 15.80 ($SD = 5.75$), respectively, and did not differ significantly, $t(290) = 0.36, p = .72$. In contrast, significantly more vulnerabilities than strengths were identified as *maximally present*, on average ($M = 3.14, SD = 3.40$ vs. $M = 2.46, SD = 3.49$), $t(290) = 2.26, p = .03$. The number of vulnerabilities, but not strengths, differed significantly between boys and girls: Boys were identified as having significantly more vulnerabilities both when we examined ratings of *moderately* or *maximally present* ($M = 17.44, SD = 4.77$ vs. $M = 15.53, SD = 5.86$), $t(289) = 1.98, p = .05$, and also when we evaluated only ratings of *maximally present* ($M = 5.22, SD = 3.77$ vs. $M = 2.80, SD = 3.21$), $t(289) = 4.36, p < .001$.

Total scores—Strength total scores ranged from 4 to 42 ($M = 18.53, SD = 8.77$) and vulnerability total scores ranged from 0 to 36 ($M = 19.36, SD = 7.78$), out of the possible range of 0 to 46. As observed at the item level, strength and vulnerability total scores were correlated in the expected direction, $r = -.22, p < .001$. Pairwise comparison showed that although strength total scores did not differ significantly, $t(287) = 0.21, p = .75$, girls had higher vulnerability total scores than did boys ($M = 22.85, SD = 7.10$ vs. $M = 18.78, SD = 7.75$, respectively), $t(286) = 3.15, p = .002$.

Key and critical items—Key and critical items afford assessors the opportunity to identify those items that are particularly relevant to the individual youth. Specifically, key items may be particularly protective against adverse outcomes or serve as therapeutic levers to motivate youth in their treatment and rehabilitation. Critical items, in contrast, may be particularly risky and represent ‘red flag’ issues that require close monitoring and attention in risk management and intervention.

On average, case managers endorsed one key ($M = 0.96, SD = 1.73, \text{Mdn} = 0, \text{Mode} = 0$) and three critical items per assessment ($M = 2.56, SD = 2.46, \text{Mdn} = 2, \text{Mode} = 0$). The number of critical and key items identified per assessment ranged widely, from 0 to 15 and 0 to 13, respectively. However, the vast majority of assessments identified five or fewer critical items (97.6%) and key items (88.7%), as recommended during training. Across

assessments, case managers identified significantly more critical than key items, $t(290) = 9.87, p < .001$. They also identified significantly more critical ($M = 2.62, SD = 2.51$) than key items ($M = 0.84, SD = 1.16$) for boys, $t(249) = 10.92, p < .001$, but not for girls ($M = 2.17, SD = 1.05$ vs. $M = 1.66, SD = 3.55$), $t(40) = 0.82, p = .36$. Though the number of critical items did not differ significantly between boys and girls, $t(289) = 1.10, p = .27$, case managers identified significantly more key items for girls than for boys, $t(289) = 2.84, p = .005$.

Table 2 presents the frequency with which each START:AV item was endorsed as a key or critical item overall, and separately for boys and girls. In order, the five most frequently endorsed key items were School/Work (9.3%), Social Skills (9.0%), Plans (9.0%), Treatability (9.0%), and Recreation (6.6%). At 43.1%, Substance Use was the most frequently endorsed critical item by far, followed by Impulse Control (25.9%), External Triggers (22.5%), School/Work (16.3%), and Rule Adherence (16.2%). Though some items were endorsed as both key and critical items at similar rates (e.g., School/Work, Plans), most items were more commonly identified as either a key item (e.g., Recreation, Treatability) or a critical item (e.g., Relationships with Peers, Substance Use, Impulse Control, External Triggers). A handful of items were never or very rarely identified as key or critical items (e.g., Self-Care, Material Resources). Review of Table 2 also shows that some key and critical items were more frequently endorsed for girls than boys and vice versa. Chi-square analyses revealed that these differences were statistically significant for endorsement of Self-Care, Mental State, Substance Use, Material Resources, Medication Adherence, Coping, and Parenting/Home Environment as key items, $\chi^2(1, Ns = 290) = 3.70, p's < .05$, and endorsement of Recreation, External Triggers, and Parenting/Home Environment as critical items, $\chi^2(1, Ns = 290) = 4.29, p's < .04$.

Specific risk estimates—With the exception of substance abuse, estimates of *low risk* were most common. For example, more than three-quarters of youths were rated as *low risk* for engaging in self-harm (84.3%), suicidal behavior (91.7%), unauthorized leave (79.3%), self-neglect (95.4%), or for being victimized by others (91.5%) in the next three months. Few were rated as *high risk* across these domains: self-harm = 3.5%; suicidal behavior = 1.8%; unauthorized leave = 7.4%; self-neglect = 0.0%; and victimization = 1.0%. More than half of youths were rated as *low risk* for violence (55.7%) and almost half were rated as *low risk* for general offending (49.1%). That said, a considerable number of youths were rated as *moderate risk* for violence (32.3%) and for general offending (41.7%) in the next three months, indicating a need to implement risk management strategies targeting these domains. In contrast, and consistent with their histories, more than one-quarter of youths were identified as *high risk* for substance abuse (27.0%) and another third were identified as *moderate risk* (36.7%). Chi-square tests revealed that, with the exception of substance abuse, $\chi^2(2, N = 259) = 4.64, p = .098$, the distribution of *low*, *moderate*, and *high* estimates of risk differed significantly for all risk domains, $\chi^2(2, N = 109 \text{ to } 216) = 47.94, p's < .001$.

We additionally conducted chi-square analyses to compare the distribution of risk estimates for boys and girls. When significant, and consistent with our comparisons of total scores, results showed higher estimates of risk for girls than for boys. Specifically, girls were more likely to be rated as *moderate* (37.5%) or *high* (18.8%) risk of self-harm than were boys (8.1% and 1.0%, respectively), $\chi^2(2, N = 115) = 25.89, p < .001$. The same pattern of results was seen for suicide, $\chi^2(2, N = 109) = 25.05, p < .001$, and for being victimized by others, $\chi^2(2, N = 199) = 7.35, p = .02$. Girls also were more likely than boys to be rated as *high risk* for unauthorized leave (29.4% vs. 3.3%), $\chi^2(2, N = 109) = 28.76, p < .001$.

Table 3 provides evidence of differential associations between the risk estimates. In general, findings highlight the co-occurrence of risks for externalizing behaviors and the co-

occurrence of risks for internalizing behaviors. For example, estimates of risk for general offending correlated significantly with estimates of risk for violence, unauthorized leave, and substance abuse. In comparison, estimates of risk for self-harming behavior were highly correlated with estimates of risk for suicide, as well as self-neglect, but not unauthorized leave, substance abuse or general offending. There was, however, a significant association between estimates of risk for self-harm and violence. Finally, there also was evidence of co-occurrence of both externalizing and internalizing behaviors with victimization: Estimates of risk for victimization were significantly correlated with estimates of risk for self-harm, suicide, unauthorized leave, self-neglect, and general offending, but not violence and general offending.

Structural Reliability

Internal consistency of the strength and vulnerability total scores was good, as indicated by Cronbach's coefficient α of .93 and .89, respectively. Examination of the MICs revealed good item homogeneity for the strength and vulnerability total scores with values of .37 and .26, respectively. Mean CITCs indicated good associations between strength item ratings and strength total scores (.58) and between vulnerability item ratings and vulnerability total scores (.48).

Review of Table 4 reveals that START:AV item ratings were differentially associated with the risk estimates. When significant correlations were observed, they were in the expected direction; that is, as vulnerability ratings increased, estimated level of risk increased and as strength ratings increased, estimated level of risk decreased. Table 4 also shows that the magnitude of association between strength and vulnerability total scores and risk estimates differed across the eight risk domains, suggesting that the presence of current strengths and vulnerabilities were influencing the final estimates of *low*, *moderate*, or *high* risk to differing degrees. As observed at the item level, when correlations were significant, they were in the expected directions: As strength total scores increased, estimates of risk for violence and unauthorized leave decreased, and as vulnerability total scores increased, estimates of risk for violence, self-harm, suicide, substance use, self-neglect, and general offending increased.

Discussion

Despite advances in the assessment and management of risks among justice-involved youths, important limitations remain, including the relative lack of consideration of short-term outcomes or protective factors in existing instruments, as well as a focus on risk for violence and offending to the exclusion of other adverse outcomes. The START:AV was developed to address these limitations. To inform the ongoing development and refinement of this new risk assessment instrument, the current study explored the feasibility, reliability and integrity of START:AV assessments completed in the field. Specifically, we examined descriptive characteristics and psychometric properties of START:AV assessments completed by case managers for 291 adolescent offenders (250 boys and 41 girls) at the time of their admission to one of three secure juvenile correctional facilities in a southern state in the United States.

Summary and Interpretation of Findings

Analysis of missing data provides evidence of the feasibility of completing START:AV assessments during routine practice. Even though assessments described herein were completed within 21 days of admission, when there was typically scant information available, youths were relatively unknown to their case managers, and there was limited time to gather additional information, the median and modal number of missing item ratings and missing risk estimates was zero. The mean number of missing item ratings for both

strengths and vulnerabilities also was low at one. In contrast, the mean number of missing risk estimates was quite high: On average, START:AV assessments were missing more than one-third (three out of a possible eight) specific risk estimates. Possible reasons for the high frequency of missing risk estimates was a prominent topic of discussion at the first booster training session, during which it became clear that there was confusion regarding when the risk estimates should be completed. Specifically, many case managers mistakenly believed that estimates only were to be completed if there was a documented history of that behavior. It was clarified during this first and subsequent booster training sessions that *all* risk estimates should be completed for *all* youths, even in the absence of a history. Our analysis suggests these efforts were not in vain; the number of missing risk estimates decreased over time, emphasizing the value of conducting ongoing training to support the implementation of evidence-based practices (Bond, Drake, & Becker, 2010; Powell et al., 2012).

These results suggest that the information required to complete the START:AV assessments is readily available in secure juvenile correctional facilities in the United States. However, they also underscore the value of implementing a structured risk assessment approach, such as the START:AV, to support the intake *process* itself. Evidence of gaps in knowledge regarding a particular youth's history, diagnoses, behavior and needs can be revealed during the START:AV assessment process, as evidenced by an inability to code some items. In this way, the START:AV framework may provide a structured approach to identifying the need for further information gathering (e.g., interviews with collaterals or requesting additional files), rather than the prevalence of missing ratings reflecting poorly on the assessor or instrument *per se*.

A high proportion of youth in our sample of incarcerated adolescent offenders were identified as having current strengths. In fact, the vast majority of START:AV assessments (95.6%) included at least one moderately or maximally present strength, a rate almost identical to those with at least one area of vulnerability (95.5%) and much higher than seen in prior research. Specifically, a recent examination of SAVRY assessments found that only about half of youths in a high-risk sample of adolescent offenders were identified as having any protective factors (Penney et al., 2010). Thus, the SAVRY's subscale of only six protective factors scored as present or absent may not be sensitive enough to tap into strengths in such high-risk samples. Indeed, in the present study, approximately half of youths (57.6%) were identified as having at least one maximally present strength, a rating perhaps more akin to the coding of "present" on the SAVRY protective factors subscale. These findings support the broader scope of variables and more nuanced coding of strengths in the START:AV. They also highlight areas of strength that might be useful building blocks for the development of risk management and treatment plans. Future research should examine whether these features can improve our understanding of the role of protective factors in the prediction of violence, offending, and other adverse outcomes (*cf.* Penney et al., 2010).

With the exception of violence, substance abuse and general offending, the prevalence of ratings of low risk was very high. As outcome data are not available at this time, it is unclear whether this is an accurate representation of risks in this population. However, the predominance of low risk estimates is consistent with studies of START implementations in forensic settings (e.g., Nicholls et al., 2011). The higher estimates of risk for substance abuse also is consistent with other START:AV research (e.g., Viljoen, Beneteau, et al., 2012). That said, the high rate of low risk estimates may reflect the challenge of estimating what the risk would be outside the institution and in the absence of intervention. Or, higher estimates of risk for violence, substance abuse and general offending may reflect case managers' focus on threats to others and public safety and, perhaps, a tendency to overlook or underestimate the risks to the youths themselves. Indeed, such a focus is consistent with

the ethos of the field. Moreover, all adolescents are placed in these specific facilities following adjudication and more than 85% undergo substance use treatment (Desmarais, Singh, & Sellers, 2012). Additionally, a very high proportion — approximately 75% — go on to serve time in adult prisons. Together, these factors probably heighten the salience of risk for violence, substance abuse, and general offending. Even so, many youths were rated as low risk across these three domains, questioning the need for residential as compared to community-based dispositions.

Our comparisons of START:AV assessments between boys and girls highlighted both important differences, as well as similarities. For instance, the number of missing ratings generally were very similar, as were item ratings, risk estimates, and the endorsement of key and critical items, suggesting the relevance of START:AV items and feasibility of completing assessments for both male and female adolescent offenders. That said, girls were frequently identified as having more vulnerabilities and being at higher risk across a range of outcomes compared to boys. Whether these represent ‘true’ differences in the risk profiles and treatment needs of adjudicated boys and girls is unknown. It is possible they reflect assessor biases or biases inherent in the instrument. Nonetheless, this pattern of results is consistent with prior research showing greater mental health needs (e.g., Cruise, Marsee, Dandreaux, DePrato, 2007) and studies showing higher risk ratings for self-injurious behavior (e.g., Gammelgård et al., 2012) among female compared to male justice-involved youths. Importantly, the present data do not speak to the predictive validity of START:AV assessments of boys and girls, a topic to be examined in future research. Additionally, continued evaluation of the potential impact of assessors and/or facilities would be of value, particularly in implementation studies.

There has been much discussion in the field regarding whether strengths and vulnerabilities, or risk and protective factors, represent distinct constructs that should be attended to separately in the risk assessment process. We have delineated herein and elsewhere the reasons we believe so (e.g., Desmarais, Nicholls, Wilson, & Brink, 2012; Desmarais, Van Dorn et al., 2012; Viljoen, Cruise, et al., 2012; Webster et al., 2006; Wilson, Desmarais, Nicholls, & Brink, 2010). In the current study, associations between strength and vulnerability ratings, as well as between the strength and vulnerability total scores, generally were significant and small to moderate in size. In fact, the observed associations were considerably smaller compared to those found in another study of START:AV assessments (Viljoen, Beneteau, et al., 2012) and research conducted with the adult version of START (e.g., Braithwaite et al., 2010; Desmarais, Nicholls, et al., 2012; Gray et al., 2011; Wilson et al., 2010; but see Desmarais, Van Dorn, et al., 2012 for an exception). We also observed differential associations of strengths and vulnerabilities with the various risk estimates. Collectively these results suggest that strength and vulnerability ratings capture different but related information, and underscore the need for continued efforts focused on understanding the interplay between risk and protective factors and its association with adverse outcomes.

Not only do the various risk domains appear to share risk and protective factors, as demonstrated by the associations between strength and vulnerability ratings and risk estimates, but they also seem to be related to each other: We found significant associations between many of the risk estimates. This finding is consistent with prior research findings suggesting that juvenile correctional facilities are serving multi-problem youths (e.g., Ford et al., 2008). Research investigating START assessments of diverse adult populations also has shown differential associations between the risk estimates (e.g., mental health jail diversion clients, Desmarais, Van Dorn, et al., 2012; forensic psychiatric inpatients and outpatients, Nicholls et al., 2011). Closer examination of the associations between the specific risk estimates revealed clustering of internalizing and externalizing risks, and associations of risk for both internalizing and externalizing behaviors with risk for

victimization. Importantly, these reflect associations among case managers' *estimates* of risk, and not necessarily associations between the outcomes themselves. However, findings of prior research bolster our confidence in the veracity of such clustering. A large body of work has shown linkages between victimization and internalizing problems, as well as between victimization and externalizing problems (e.g., Reijntjes, Kamphuis, Prinzie, & Telch, 2010; Reijntjes et al., 2011), and also has documented the co-occurrence of externalizing and internalizing behaviors in adolescents (Cosgrove et al., 2010; Pesenti-Gritti et al., 2008). Together, findings of the current and past research suggest that the multiple START:AV risk domains are interrelated in incarcerated adolescent offenders and support their joint consideration in one comprehensive assessment.

Finally, our analysis of structural reliability supports the integrity of START:AV assessments completed in the field; alpha coefficients, MICs and MCICs all were in the acceptable range and very similar to values found in other START:AV and START research (e.g., Nicholls et al., 2006; Nonstad et al., 2011; Viljoen, Beneteau, et al., 2012). Some experts have questioned the relevance of structural reliability coefficients to risk assessment instruments (e.g., Douglas, Skeem, & Nicholson, 2011; Kropp & Hart, 2000). Specifically, two arguments have been made: 1) internal consistency is irrelevant because these instruments are not designed to measure an underlying construct (Douglas et al., 2011); and 2) lower levels of item homogeneity would produce greater predictive validity because there would be less redundancy among items (Kropp & Hart, 2000). These are compelling arguments, yet we chose to report structural reliability statistics: a) to afford comparison with previously reported psychometric data on well-established risk assessment instruments, and b) to provide information on how the START:AV items relate to one another. Indeed, in contrast with the viewpoints reviewed above, other experts assert that internal consistency between items is a criterion for determining if a tool has sufficient evidence to be considered an evidence-based risk assessment tool (Vincent, Terry, & Maney, 2009). We leave this issue open for continued debate.

Limitations

Several limitations to the current study reflect our focus on START:AV assessments conducted in the field. Specifically, some features of an optimal research design could not be implemented. For many reasons, for instance, it was not possible to have case managers (or research assistants) complete interrater assessments, nor was it possible to obtain measures of concurrent validity to verify the accuracy of assessments. Because of the limited information available at intake and because youths are not randomly assigned to case managers, the establishment of interrater reliability of START:AV assessments completed in the field is a research priority. Additionally, START:AV assessments were completed over the course of an implementation and our analyses of missing data suggest that these assessments, or completion rates more specifically, changed over time. From an implementation point of view, this represents a very positive outcome, but complicates our interpretation of results. Other limitations relate to our sample. For example, the sample size did not afford systematic comparison of characteristics and psychometric properties as a function of assessor characteristics (e.g., demographics, training, professional background, level of experience, etc.). Similarly, due to nature of the population and facilities, our sample of girls is very small, and power to detect differences between assessments of boys and girls is limited. We also are limited to assessments completed in one jurisdiction and generalizability of our findings to other jurisdictions and populations will need to be tested. Finally, no outcome data is available at this time to afford analyses of predictive validity, though data collection is underway.

Conclusions

Notwithstanding these limitations, our findings provide preliminary support for the feasibility of completing START:AV assessments as part of routine practice in secure juvenile correctional facilities. Findings also highlight differences in the characteristics of START:AV assessments for boys and girls, as well as differential associations between the eight START:AV risk domains. Though results are promising, further research is needed to establish the reliability and validity of START:AV assessments completed in the field. Further research also is needed to establish the feasibility, acceptability, sustainability, and, importantly, impact of START:AV assessments in practice. To date, there has been limited examination of whether implementation of risk assessment tools—START:AV or otherwise—actually improves risk management and intervention and, ultimately, reduces adverse outcomes. This remains a critical knowledge gap and a high priority area for future research.

Acknowledgments

Funding

This study was supported by Award Number P30DA028807 from the National Institute on Drug Abuse. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health.

References

- Annie, E. Casey Foundation Pathways to juvenile detention reform. Baltimore, MD: Author; 2000.
- Austin, J.; Johnson, KD.; Weitzer, R. Alternatives to the secure detention and confinement of juvenile offenders. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention; 2005.
- Bond GR, Drake RE, Becker DR. Beyond evidence-based practice: Nine ideal features of a mental health intervention. *Research on Social Work Practice*. 2010; 20:493–501.
- Borum, R.; Bartel, P.; Forth, A. Manual for the Structured Assessment of Violence Risk in Youth (SAVRY). Odessa, FL: Psychological Assessment Resources; 2006.
- Borum, R.; Verhaagen, D. Assessing and managing violent risk in juveniles. New York: The Guilford Press; 2006.
- Bronfenbrenner, U. The ecology of human development: Experiments by nature and design. Cambridge, MA: Harvard University Press; 1979.
- Catchpole REH, Gretton HM. The predictive validity of risk assessment with violent young offenders. *Criminal Justice and Behavior*. 2003; 30:688–708.
- Cohen, J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum; 1988.
- Cosgrove VE, Rhee SH, Gelhorn HL, Boeldt D, Corley RC, Ehringer MA, et al. Structure and etiology of co-occurring internalizing and externalizing disorders in adolescents. *Journal of Abnormal Child Psychology*. 2010; 39:109–123. [PubMed: 20683651]
- Crocker AG, Braithwaite E, Laferrière D, Gagnon D, Venegas C, Jenkins T. START changing practice: Implementing a risk assessment and management tool in a civil psychiatric setting. *International Journal of Forensic Mental Health*. 2011; 10:13–28.
- Desmarais SL, Nicholls TL, Wilson CM, Brink J. Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment*. 2012; 24:685–700. [PubMed: 22250595]
- Desmarais, SL.; Singh, JP.; Sellers, BG. Assessing and managing short-term risks in justice-involved youth; Symposium presented at the University of South Florida Mini-Conference on Co-Occurring Disorders and the Justice System: Innovations in Research and Practice; Tampa, FL. 2012 Apr.
- Desmarais SL, Van Dorn RA, Telford RP, Petrila J, Coffey T. Characteristics of START assessments completed in mental health jail diversion programs. *Behavioral Sciences & the Law*. 2012; 30:448–469. [PubMed: 22807034]

- Douglas, KS.; Otto, R.; Desmarais, SL.; Borum, R. Clinical forensic psychology. In: Weiner, IB.; Schinka, JA.; Velicer, WF., editors. Handbook of psychology, volume 2: Research methods in psychology. Hoboken, NJ: John Wiley & Sons; (in press)
- Douglas, KS.; Skeem, JL.; Nicholson, E. Research methods in violence risk assessment. In: Rosenfeld, B.; Penrod, SD., editors. Research methods in forensic psychology. Hoboken, NJ: John Wiley & Sons, Inc.; 2011. p. 325-346.
- Epperson, D.; Ralston, C.; Fowers, D.; Dewitt, J.; Gore, K. Juvenile sexual offense recidivism rate assessment tool-II (JSORRAT-II). In: Prescott, D., editor. Risk assessment of youth who have sexually abused. Oklahoma City, OK: Wood 'N' Barnes Publishing and Distribution; 2006. p. 222-236.
- Fagan, J. Contexts of choice by adolescents in criminal events. In: Grisso, T.; Swartz, RG., editors. Youth on trial: A developmental perspective on juvenile justice. Chicago: University of Chicago Press; 2000. p. 371-401.
- Ford JD, Hartman JK, Hawke J, Chapman JF. Traumatic victimization, posttraumatic stress disorder, suicidal ideation, and substance abuse risk among juvenile justice-involved youth. *Journal of Child and Adolescent Trauma*. 2008; 1:75–92.
- Forth, AE.; Kosson, D.; Hare, RD. The Hare Psychopathy Checklist: Youth Version. Toronto, ON: Multi-Health Systems; 2003.
- Gammelgård M, Weizmann-Henelius G, Koivisto A-M, Eronen m, Kaltiala-Heino R. Gender differences in violence risk profiles. *Journal of Forensic Psychiatry & Psychology*. 2012; 23:76–94.
- Gray NS, Benson R, Craig R, Davies H, Fitzgerald S, Huckle P, et al. The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behavior. *International Journal of Forensic Mental Health*. 2011; 10:305–313.
- Griffin, P.; Bozynski, M. National overviews: State juvenile justice profiles. 2003. Retrieved from <http://www.ncjj.org/stateprofiles/>
- Hockenberry, S.; Sickmund, M.; Sladky, A. Juvenile residential facility census, 2008. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention; 2011.
- Hoge, RD.; Andrews, DA. The Youth Level of Service/Case Management Inventory manual and scoring key. North Tonawanda, NY: Multi-Health Systems; 2002.
- Juvenile Justice and Delinquency Prevention Act of 1974, 42 U.S.C. §5601 et seq. 2002
- Kropp PR, Hart SD. The Spousal Assault Risk Assessments (SARA) guide: Reliability and validity in adult male offenders. *Law and Human Behavior*. 2000; 24:101–118. [PubMed: 10693321]
- Lipsey, MW.; Derzon, JH. Predictors of violent or serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In: Loeber, R.; Farrington, DP.; Loeber, R.; Farrington, DP., editors. Serious & violent juvenile offenders: Risk factors and successful interventions. Thousand Oaks, CA U.S.: Sage Publications, Inc.; 1998. p. 86-105.
- Lodewijks HPB, Doreleijers TAH, de Ruiter C, Borum R. Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry*. 2008; 31:263–271. [PubMed: 18508122]
- McGowan MR, Horn RA, Mellott RN. The predictive validity of the Structured Assessment of Violence Risk in Youth in secondary educational settings. *Psychological Assessment*. 2011; 23:478–486. [PubMed: 21381837]
- Memorandum of Agreement Between the United States and the State of Georgia Concerning Georgia Juvenile Justice Facilities, No. 1:98-CV--836-JEC (N.D. Ga. 1998) Amended 2008.
- Meyers JR, Schmidt F. Predictive validity of the Structured Assessment for Violence Risk in Youth (SAVRY) with juvenile offenders. *Criminal Justice and Behavior*. 2008; 35:344–355.
- Nicholls TL, Petersen K, Brink J, Webster C. A clinical risk profile of forensic psychiatric patients: Treatment team STARTs in a Canadian service. *International Journal of Forensic Mental Health*. 2011; 10:187–199.
- Nicholls, TL.; Viljoen, JL.; Cruise, KR.; Desmarais, SL.; Webster, CD. Short-Term Assessment of Risk and Treatability: Adolescent Version (START: AV) (Abbreviated Manual). Coquitlam, Canada: BC Mental Health and Addiction Services; 2010.

- Nonstad K, Nasset MB, Kroppan E, Pedersen TW, Nøttestad JA, Almvik R, Palmstierna T. Predictive validity and other psychometric properties of the Short-Term Assessment of Risk and Treatability (START) in a Norwegian high secure hospital. *International Journal of Forensic Mental Health Services*. 2010; 9:294–299.
- Nunnally, J.; Bernstein, I. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.
- Olver ME, Stockdale KC, Wormith J. Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior*. 2009; 36:329–353.
- Penney SR, Lee Z, Moretti MM. Gender differences in risk factors for violence: An examination of the predictive validity of the Structured Assessment of Violence Risk in Youth. *Aggressive Behavior*. 2010; 36:390–404. [PubMed: 20623508]
- Pesenti-Gritti P, Spatola CAM, Fagnani C, Ogliari A, Patriarca V, Stazi MA, Battaglia M. The co-occurrence between internalizing and externalizing behaviors: A general population twin study. *European Child & Adolescent Psychiatry*. 2008; 17:82–92. [PubMed: 17846816]
- Powell BJ, McMillen C, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, et al. A compilation of strategies for implementing clinical innovations in health and mental health. *Medical Care Research and Review*. 2012; 29:123–157. [PubMed: 22203646]
- Prentky, R.; Righthand, S. *Juvenile Sex Offender Assessment Protocol II (J-SOAP-II) manual*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention; 2003. Retrieved from <https://www.ncjrs.gov/pdffiles1/ojjdp/202316.pdf>
- Quinsey, VL.; Harris, GT.; Rice, ME.; Cormier, CA. *Violent offenders: Appraising and managing risk*. 2nd ed.. Washington, DC: American Psychological Association; 2006.
- Reijntes A, Kamphuis JH, Prinzie P, Boelen PA, van der Schoot M, Telch MJ. Prospective linkages between peer victimization and externalizing problems in children: A meta-analysis. *Aggressive Behavior*. 2011; 37:215–222. [PubMed: 21433031]
- Reijntes A, Kamphuis JH, Prinzie P, Telch MJ. Peer victimization and internalizing problems in children: A meta-analysis of longitudinal studies. *Child Abuse & Neglect*. 2010; 34:244–252. [PubMed: 20304490]
- Reza A, Krug EG, Mercy JA. Epidemiology of violent deaths in the world. *Injury Prevention*. 2001; 7:104–111. [PubMed: 11428556]
- Schmidt F, Campbell MA, Houlding C. Comparative analyses of the YLS/CMI, SAVRY, and PCL:YV in adolescent offenders: A 10-year follow-up into adulthood. *Youth Violence and Juvenile Justice*. 2011; 9:23–42.
- Schwalbe CS. Risk assessment for juvenile offenders: A meta-analysis. *Law and Human Behavior*. 2007; 31:449–462. [PubMed: 17211688]
- Schwalbe CS. A meta-analysis of juvenile justice risk assessment instruments: Predictive validity by gender. *Criminal Justice and Behavior*. 2008; 35:1367–1381.
- Scott, ES.; Steinberg, L. *Rethinking juvenile justice*. Cambridge, MA: Harvard University Press; 2008.
- Singh JP, Grann M, Fazel S. A comparative study of violence risk assessment tools: A systematic review and meta-regression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*. 2011; 31:499–513. [PubMed: 21255891]
- Slowikowski, J. *OJJDP News at a Glance*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention; 2009 Nov-Dec. Department of Justice sets high priority on reentry programs for ex-offenders.
- Snyder HN. An empirical portrait of the youth reentry population. *Youth Violence and Juvenile Justice*. 2004; 2:39–55.
- United States of America v. State of Georgia, et al., No. 1:98-CV--836-JEC (N.D. Ga. 1998).
- Viljoen JL, Cruise KR, Nicholls TL, Desmarais SL, Webster CD. Taking stock and taking steps: The case for an adolescent version of the Short-Term Assessment of Risk and Treatability. 2012; 11:135–149.
- Viljoen JL, Mordell S, Beneteau JL. Prediction of adolescent sexual reoffending: A meta-analysis of the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99. *Law and Human Behavior*. 2012; 36:423–438. [PubMed: 22353046]

- Viljoen JL, Beneteau J, Gulbransen E, Brodersen E, Desmarais SL, Nicholls TL, Cruise KR. Assessment of multiple risk outcomes, strengths, and change with START:AV assessments: A short-term prospective study with adolescent offenders. 2012; 11:165–180.
- Viljoen JL, Scalora M, Cuadra L, Bader S, Chávez V, Ullman D, Lawrence L. Assessing risk for violent in adolescents who have sexually offended: A comparison of the J-SOAP-II, J-SORRAT-II, and SAVRY. *Criminal Justice and Behavior*. 2008; 35:5–23.
- Vincent GM, Guy LS, Fusco SL, Gershenson BG. Field reliability of the SAVRY with juvenile probation officers: Implications for training. *Law and Human Behavior*. 2012; 36:225–236. [PubMed: 22667812]
- Vincent, GM.; Terry, A.; Maney, S. Risk/needs tools for antisocial behavior and violence among youthful populations. In: Andrade, J., editor. *Handbook of violence risk assessment and treatment: New approaches for forensic mental health practitioners*. New York: Springer; 2009. p. 377-423.
- Wasserman GA, McReynolds LS. Suicide risk at juvenile justice intake. *Suicide and Life-Threatening Behavior*. 2006; 36:239–249. [PubMed: 16704327]
- Webster, CD.; Martin, M-L.; Brink, J.; Nicholls, TL.; Desmarais, SL. *Manual for the Short-Term Assessment of Risk and Treatability (START) Version 1.1*. Coquitlam, Canada: Forensic Psychiatric Services Commission; 2009.
- Welsh JL, Schmidt F, McKinnon L, Chattha HK, Meyers JR. A comparative study of adolescent risk assessment instruments: Predictive and incremental validity. *Assessment*. 2008; 15:105–115.
- Wilson CM, Desmarais SL, Nicholls TL, Brink J. The role of client strengths in assessments of risk using the Short-Term Assessment of Risk and Treatability (START). *International Journal of Forensic Mental Health Services*. 2010; 9:282–293.

Table 1

START:AV item ratings

Item		Descriptives		Frequency of Endorsement			<i>r</i>
		<i>M</i>	<i>SD</i>	Minimally Present (%)	Moderately Present (%)	Maximally Present (%)	
1. Social Skills	Strength	1.05	0.59	15.3%	64.8%	19.9%	-.41**
	Vulnerability	0.76	0.60	33.2%	58.0%	8.7%	
2a. Adult Relationships	Strength	0.83	0.60	28.1%	60.9%	11.0%	-.35**
	Vulnerability	0.80	0.62	31.0%	57.7%	11.4%	
2b. Peer Relationships	Strength	0.75	0.52	28.9%	67.1%	3.9%	-.29**
	Vulnerability	0.99	0.58	17.7%	66.1%	16.3%	
3. School/Work	Strength	0.83	0.68	33.1%	51.1%	15.8%	-.58**
	Vulnerability	0.96	0.67	24.8%	54.6%	20.6%	
4. Recreation	Strength	0.91	0.62	24.2%	60.7%	15.1%	-.25**
	Vulnerability	0.65	0.59	40.8%	53.5%	5.6%	
5. Self-Care	Strength	1.15	0.60	11.1%	62.4%	26.5%	-.12*
	Vulnerability	0.39	0.51	62.0%	36.9%	1.0%	
6. Mental State	Strength	0.91	0.62	23.8%	61.4%	14.8%	-.24**
	Vulnerability	0.57	0.62	49.7%	43.8%	6.6%	
7. Emotional State	Strength	0.87	0.60	24.7%	63.1%	12.2%	-.37**
	Vulnerability	0.74	0.61	35.1%	55.9%	9.0%	
8. Substance Use	Strength	0.52	0.64	55.8%	36.4%	7.8%	-.47**
	Vulnerability	1.18	0.75	20.8%	40.8%	38.4%	
9. Impulse Control	Strength	0.60	0.56	43.5%	52.6%	3.9%	-.35**
	Vulnerability	1.16	0.60	10.9%	62.0%	27.1%	
10. External Triggers	Strength	0.59	0.53	42.8%	55.1%	2.1%	-.33**
	Vulnerability	1.18	0.57	8.5%	64.9%	26.6%	
11a. Adult Social Support	Strength	0.93	0.59	21.0%	65.0%	14.0%	-.22**
	Vulnerability	0.77	0.60	32.1%	58.5%	9.4%	
11b. Peer Social Support	Strength	0.55	0.54	47.0%	50.9%	2.1%	-.34**
	Vulnerability	1.11	0.56	11.1%	67.1%	21.8%	

Item	Descriptives		Frequency of Endorsement			<i>r</i>
	<i>M</i>	<i>SD</i>	Minimally Present (%)	Moderately Present (%)	Maximally Present (%)	
12. Material Resources	Strength	0.91	0.54	19.4%	70.1%	10.6%
	Vulnerability	0.59	0.59	46.5%	48.2%	5.3%
13. Attitudes	Strength	0.89	0.84	25.4%	63.6%	10.6%
	Vulnerability	0.82	0.64	30.9%	56.1%	13.0%
14. Medication Adherence	Strength	1.01	0.64	19.7%	59.6%	20.8%
	Vulnerability	0.51	0.58	53.6%	42.0%	4.4%
15. Rule Adherence	Strength	0.79	0.60	30.7%	60.0%	9.3%
	Vulnerability	1.01	0.63	18.9%	61.1%	20.0%
16. Conduct	Strength	0.76	0.61	33.3%	57.4%	9.2%
	Vulnerability	1.00	0.61	18.6%	62.8%	18.6%
17. Insight	Strength	0.69	0.55	35.5%	59.9%	4.6%
	Vulnerability	0.96	0.55	17.2%	69.5%	13.3%
18. Plans	Strength	0.71	0.63	38.7%	51.9%	9.4%
	Vulnerability	0.91	0.56	20.6%	67.8%	11.5%
19. Coping	Strength	0.78	0.54	27.9%	66.1%	6.0%
	Vulnerability	0.92	0.59	21.5%	64.8%	13.7%
20. Treatability	Strength	1.03	0.61	17.4%	62.5%	20.1%
	Vulnerability	0.75	0.60	33.3%	58.2%	8.4%
21. Parenting / Home Environment	Strength	0.85	0.56	24.2%	66.9%	8.9%
	Vulnerability	0.84	0.62	28.2%	59.6%	12.3%

Notes. *N* = 178–290 START:AV assessments. % = valid percent. *r* = correlation between strength and vulnerability ratings.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 2

START:AV items identified as key or critical

START:AV Items	Key Item (%)			Critical Item (%)		
	Overall	Girls	Boys	Overall	Girls	Boys
1. Social Skills	9.0%	4.9%	9.7%	5.2%	4.9%	5.2%
2a. Adult Relationships	4.8%	7.3%	4.4%	6.3%	9.8%	5.7%
2b. Peer Relationships	0.7%	0.0%	0.8%	13.9%	15.4%	4.9%
3. School/Work	9.3%	8.8%	9.3%	16.3%	10.0%	17.3%
4. Recreation	6.6%	4.9%	6.8%	0.7%	4.9%	0.0%
5. Self-Care	0.0%	14.6%	2.0%	1.7%	0.0%	0.0%
6. Mental State	1.7%	9.8%	0.4%	7.9%	7.3%	8.0%
7. Emotional State	3.8%	7.3%	3.2%	14.5%	17.1%	14.1%
8. Substance Use	3.8%	9.8%	2.8%	43.1%	39.0%	43.8%
9. Impulse Control	2.4%	4.9%	2.0%	25.9%	34.1%	24.5%
10. External Triggers	0.0%	0.0%	0.0%	22.5%	4.9%	25.4%
11a. Adult Social Support	5.2%	9.8%	4.4%	4.5%	4.9%	4.4%
11b. Peer Social Support	0.3%	0.0%	0.4%	13.8%	4.9%	15.3%
12. Material Resources	1.7%	12.2%	0.0%	1.7%	2.4%	1.6%
13. Attitudes	5.9%	9.8%	5.2%	10.0%	9.8%	10.0%
14. Medication Adherence	3.5%	12.2%	2.0%	3.8%	0.0%	4.5%
15. Rule Adherence	3.8%	2.4%	4.0%	16.2%	9.8%	17.3%
16. Conduct	3.1%	4.9%	2.8%	13.8%	9.8%	14.5%
17. Insight	4.1%	7.3%	3.6%	7.6%	9.8%	7.2%
18. Plans	9.0%	9.8%	8.8%	10.0%	2.4%	11.2%
19. Coping	2.8%	7.3%	2.0%	8.6%	7.3%	8.8%
20. Treatability	9.0%	8.8%	9.8%	2.4%	2.4%	2.4%
21. Parenting/Home Environment	2.1%	7.3%	1.2%	8.7%	17.1%	7.3%

Notes. $N = 287-291$ START:AV assessments. Key Item = item that is particularly relevant, either recently or historically, to the youth's treatment and intervention. Critical Item = item that is particularly relevant, either recently or historically, to the youth's risk management. % = valid percent.

Table 3

Associations between START:AV risk estimates

START Risk Estimates	Correlations							
	1	2	3	4	5	6	7	8
1. Violence	--							
2. Self-Harm	.22 [*]	--						
3. Suicide	.08	.86 ^{***}	--					
4. Unauthorized Leave	.16 [†]	-.06	-.06	--				
5. Substance Abuse	.08	-.11	-.18 [*]	.25 ^{***}	--			
6. Self-Neglect	.31 ^{***}	.35 ^{**}	.29 ^{**}	.14 [*]	.17	--		
7. Victimization	.05	.23 [*]	.25 [*]	.22 ^{**}	.06	.22 ^{**}	--	
8. General Offending	.31 ^{***}	-.05	-.07	.25 ^{**}	.39 ^{***}	.02	.19 [*]	--

Note. $N = 93\text{--}216$ START assessments. Values are Kendall's tau-b correlation coefficients.

^{*}
 $p < .05$.

^{**}
 $p < .01$.

^{***}
 $p < .001$.

[†]
 $p = .06$.

Table 4

Correlations between START:AV item ratings and risk estimates

Item Ratings	Risk Estimates									
	VIO	SH	SUI	UL	SA	SN	VIC	GO		
1. Social Skills	S	-.16*	-.23*	-.19*	.02	.10	-.03	.004	-.02	
	V	.15*	.21*	.18	-.02	.05	.11	.11	.16*	
2a. Adult Relationships	S	-.32**	-.06	.04	-.13*	.01	.002	-.09	-.10	
	V	.23**	.13	.01	.05	.16**	.06	.15*	.28**	
2b. Peer Relationships	S	-.09	.02	-.02	-.01	-.06	.001	.01	-.06	
	V	.17*	.04	-.02	.06	.26**	.11	.11	.32**	
3. School/Work	S	-.19**	-.08	-.05	-.20**	-.15**	-.07	-.09	-.18**	
	V	.12	.02	-.03	.10	.28**	.09	.03	.22**	
4. Recreation	S	-.11	-.11	-.03	-.20**	.002	.12	.15*	-.03	
	V	-.02	.26**	.18	.07	.23**	.03	-.10	.06	
5. Self-Care	S	-.15*	-.18*	-.13	-.08	-.06	-.11	.08	-.01	
	V	-.03	.10	.16	-.16*	.18**	.11	-.16*	.06	
6. Mental State	S	-.26**	-.15	-.14	-.05	.11	-.09	-.03	-.05	
	V	.14*	.26**	.26**	-.13*	-.01	.16*	.17*	.13*	
7. Emotional State	S	-.20**	-.27**	-.19*	.07	-.20**	.01	.002	.05	
	V	.11	.41**	.35**	-.10	-.02	.17*	.07	.09	
8. Substance Use	S	-.01	.07	.16	-.18**	-.28**	-.04	-.02	-.12	
	V	.02	.05	-.07	.16*	.59**	.17*	.03	.16*	
9. Impulse Control	S	-.26**	-.15	-.13	-.22**	-.02	-.03	-.15*	-.15*	
	V	.24**	.21*	.16	.17*	.15**	.14*	.08	.22**	
10. External Triggers	S	-.30**	-.04	-.08	-.26	-.07	-.08	-.11	-.21**	
	V	.33**	.05	.10	.18**	.20**	.14*	.08	.33**	
11a. Adult Social Support	S	-.16*	-.05	.004	-.10	-.003	.01	-.07	-.10	
	V	.03	.22*	.14	.04	.16**	.08	.11	.15*	
11b. Peer Social Support	S	-.12	-.03	.14	-.26**	-.06	-.06	-.11	-.11	
	V	.11	-.01	-.01	.15*	.29**	.08	.15*	.34**	
12. Material Resources	S	-.18*	.16	.20*	-.02	.07	.20**	-.12	-.05	

Item Ratings	Risk Estimates									
	VIO	SH	SUI	UL	SA	SN	VIC	GO		
13. Attitudes	V .03	.06	.00	-.06	.18**	.04	.04	.05**		
	S -.29**	-.07	-.09	-.13*	.03	-.04	-.04	-.07		
14. Medication Adherence	V .35**	.08	.08	.08	.13*	.14*	.02	.27**		
	S -.11	.00	.12	-.07	.00	.16	.21*	.10		
15. Rule Adherence	V .22*	.04	-.09	.02	.01	.01	-.20*	-.03		
	S -.39**	-.10	.02	-.32**	-.02	-.06	-.12	-.29**		
16. Conduct	V .21**	.04	.02	.16*	.15**	.07	-.03	.33**		
	S -.38**	-.17	-.07	-.14*	.04	-.05	-.17*	-.24**		
17. Insight	V .36**	.27**	.18	.07	.14*	.28**	.10	.32**		
	S -.28**	-.12	-.10	-.09	.15**	-.004	-.10	-.10		
18. Plans	V .16*	.19*	.20*	.07	.10	.06	.06	.23**		
	S -.27**	-.13	-.04	-.04	.05	.04	.07	-.05		
19. Coping	V .25**	.14	.08	.10	.13*	.05	-.08	.27**		
	S -.32**	-.23*	-.22*	-.09	.05	.04	-.18*	-.09		
20. Treatability	V .22**	.22**	.18	.14*	.14*	.11	.15*	.11		
	S -.34**	-.25**	-.16	-.20**	.07	-.07	-.09	-.05		
21. Parenting / Home Environment	V .35**	.23**	.22*	.10	.02	.16*	-.05	.21**		
	S -.16*	-.09	-.03	-.17**	.07	.05	-.06	-.03		
START Total Scores	V .05	.21*	.16	.04	.17*	.09	.12	.08		
	S -.25**	-.14	-.05	-.16**	.01	-.01	-.05	-.10		
Victimization; GO = General Offending.	V .21**	.23**	.18*	.05	.23**	.18**	.05	.24**		

Notes. N = 93–288 START assessments. S = Strength; V = Vulnerability; VIO = Violence; SH = Self-Harm; SUI = Suicide; UL = Unauthorized Leave; SA = Substance Abuse; SN = Self-Neglect; VIC = Victimization; GO = General Offending.